Ontologías para el modelado y el análisis de microRNAs

Sergi Albert Ballestar
Master de Bioinformática & Bioestadística
M0.168 TFM

Tutor: Albert Pla Planas
Professor: Jose Antonio Morán Moreno

24/05/2017

"[...] each area of ontology activity builds ontologies in different ways, and so the only way you can take advantage of the Linked Open Data (as the experts in Linked Open Data will tell you) is through considerable manual effort, usually involving not sleeping."

Barry Smith, en el *2015 International Conference on Biomedical Ontology*, Lisboa, Portugal, 28 de Julio de 2015

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Ontologías para el modelado y el análisis de microRNAs*
Nombre del autor:	*Sergi Albert Ballestar*
Nombre del consultor/a:	*Albert Pla Planas*
Nombre del PRA:	*Jose Antonio Morán Moreno*
Fecha de entrega (mm/aaaa):	05/2017
Titulación::	*Master de Bioinformática & Bioestadística*
Área del Trabajo Final:	*M0.168 - TFM*
Idioma del trabajo:	*Español*
Palabras clave	*ontología, microRNA, modelaje*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.*

Los miRNA son secuencias de RNA no codificante de unos 19-25nt, en muchos casos conservadas entre especies, que tienen un importante papel en la regulación génica. En contraste con el gran número de herramientas computacionales, no existe todavía una ontología que sea capaz de representar todos los conceptos relacionados con los miRNA. El objetivo de este proyecto es la extensión de una ontología de miRNA (y sus dianas) para el modelaje y estudio de los miARN; además de conseguir una plataforma computacional que permita realizar consultas a dicha ontología. Para ello, se han analizado las ontologías existentes acerca de miRNA, identificando sus virtudes y limitaciones. A partir de miRNA Ontology (MIRNAO) y Ontology of microRNA Target (OMIT) se ha desarrollado una nueva ontología llamada miRNA Target Ontology (MIRNATO), que tiene una mayor capacidad de representación en comparación a sus antecesoras. Para la extensión de la ontología se ha usado *Protégé 5.2.0* y para la plataforma un servidor *Fuseki Jena Apache* (basado en Java). El producto final ha sido una ontología llamada MIRNATO capaz de modelar muchos más conceptos que sus antecesoras, y más adaptada al uso de BBDD y algoritmos de predicción de dianas.

Abstract (in English, 250 words or less):

miRNAs are non-coding RNA sequences of 19-25nt, in many cases conserved across species, which have an important role in gene regulation. In contrast to the large number of computational tools, there is not yet an ontology that is able to represent all concepts related to miRNA. The objective of this project is to extend an ontology about miRNA (and its targets) for the modeling and study of concepts about miRNAs. In addition, another goal is to obtain a computational platform that allows making queries to the extended ontology. To achieve this goals, the existing ontologies about miRNA were analyzed, identifying their virtues and limitations. A new ontology called miRNA Target Ontology (MIRNATO) has been developed from miRNA Ontology (MIRNAO) and Ontology of microRNA Target (OMIT), which has a greater capacity of representation compared to its predecessors. Protégé 5.2.0 has been used to extend the ontology and for the computational platform was provided by a server Fuseki Jena Apache (based on Java). The final product has been a new ontology called miRNA Target Ontology (MIRNATO), capable of modeling many more concepts than its predecessors, and more adapted to the use of databases and target prediction algorithms.

Índice

1. Introducción **10**
1.1 Contexto y justificación del Trabajo 10
1.1.1 Descripción general 10
1.1.2. Justificación del TFM 11
1.2 Objetivos del Trabajo 11
1.2.1 Objetivos generales 11
1.2.2 Objetivos específicos 11
1.3 Enfoque y método seguido 12
1.4 Planificación del Trabajo 12
1.5 Breve sumario de productos obtenidos 15
1.6 Breve descripción de los otros capítulos de la memoria 15

2. Introducción a microRNA y Ontologías **17**
2.1 Descripción y contexto de los miRNA 17
2.1.1 Origen y definición 17
2.1.2 Biogénesis canónica en animales 19
2.1.3 Herramientas computacionales 19
2.2 Ontologías 22
2.2.1 Necesidad de análisis ontológico 23
2.3 Partes de una ontología 25

3. Ontologías para el modelaje de dianas de microRNA **27**
3.1 Enfoque del trabajo: Extensión o creación de una nueva ontología 27
3.2 Extensión de la Ontología de miRNA 28
3.2.1 Conceptos a modelar 28
3.2.2 Retos y dificultades 35
3.2.3 Extensión 35

4. Plataforma de consulta **49**
4.1. Almacenaje de miRNA 51

5. Resultados **53**

6. Conclusiones **55**
6.1 Lecciones aprendidas 56
6.2 Reflexión & Análisis crítico 57
6.3 Líneas de trabajo futuro 58

7. Bibliografía **59**

8. Anexos **64**

8.2 Glosario de términos ontológicos 65

Lista de figuras

TABLA 1 Descripción cronólogica de las tareas. Pág 13
TABLA 2 Descripción cronólogica de los hitos. Pág. 14
TABLA 3 Características comunes en distintas herramientas de predicción de miRNA targets (Shirdell et al., 2011) (Peterson et al., 2014) Pag 35
TABLA 4 Pag 65
FIGURA 1 FIGURA 1. Diagrama de Grantt con tareas e hitos Pag 16
FIGURA 2 Biogénesis canónica en animales (Arenz C., 2014) Pag 20
FIGURA 3 Pequeño esquema con algunas de las herramientas y bases de datos disponibles relacionadas con miRNA (Akhtar et al., 2015) Pag 22
FIGURA 4 Comparación de algunas clases de MIRNAO y NCRO. A la izquierda, una captura de pantalla de MIRNAO en Online Look up Service (OLS)[1] de EMBL. A la derecha, NCRO en OLS[2] de EMBL Pag 26
FIGURA 5 La ruta canónica de procesamiento de miARN animal (panel derecho) y los términos GO que se espera estén asociados con los componentes proteícos de esta vía (panel izquierdo). La asociación de estos términos GO dependerá de la evidencia experimental disponible; otros términos pueden estar asociados a estos productos génicos si la evidencia adecuada está disponible. "Contributes_to" es una propiedad utilizada en la anotación GO para indicar que la entidad anotada no realiza la función molecular de forma aislada, sino como miembro de un complejo. Nombres de proteínas: (DROSHA) Ribonucleasa 3; (DGCR8) Subunidad del complejo de microprocesador DGCR8; (XPO-5) Exportin-5; (RAN - GTP) GTPasa cargada con GTP; (DICER1) Endoribonucleasa Dicer; (TARBP2) subunidad compleja de carga RISC TARBP2; (AGO) Argonaute. (Huntley et al., 2016) Pag 32
FIGURA 6 El color de las líneas en los gráficos de OntoGraf indica sus propiedades o relacionantes pag 37
FIGURA 7 Literatura relacionada con el miRNA. Todas las flechas son de color azul, és decir, has_subclass. Ver leyenda en (Figura 6) pag38
FIGURA 8 Información de las BBDD. Ver leyenda en (Figura 6) pag 39
FIGURA 9 Tipo de validación empírica. Ver leyenda en (Figura 6) pag 39
FIGURA 10 Predicción computacional del miARN y su target. Ver leyenda en Figura 6 pag 40
FIGURA 11 Organismo al que pertenece. Ver leyenda en Figura 6 pag 40
FIGURA 11.2 Tipos de seed region. Ver leyenda en Figura 6 pag 42
FIGURA 12 IsomiRs. Ver leyenda en Figura 6 pag 42
FIGURA 13 Transcripción. Ver leyenda en Figura 6 Pag 43
FIGURA 14 Escisión (de pri-miRNA a pre-miRNA). Ver leyenda en Figura 6

[1] http://www.ebi.ac.uk/ols/ontologies/mirnao (2017)
[2] http://www.ebi.ac.uk/ols/ontologies/ncro (2017)

Pag 44
FIGURA 15 Transporte del núcleo al citoplasma. Ver leyenda en Figura 6 Pag 44
FIGURA 16 Escisión del loop de pre-miRNA en el medio citoplasmático. Ver leyenda en Figura 6 pag 45
FIGURA 17 Formación de RISC & Efecto regulador. Consultar leyenda en Figura 6 pag 46
FIGURA 18 Propiedades termodinámicas. Consultar leyenda en Figura 6 Pag 47
FIGURA 19 Propiedades estructurales del ácido nucleico. Consultar leyenda en Figura 6 Pag 47
FIGURA 20 Regiones flanquentes. Consultar leyenda en Figura 6 Pag 48
FIGURA 21 Posiciones de inicio y final. Consultar leyenda en Figura 6 Pag 48
FIGURA 23 Gaps, bulges y wobbles. Consultar leyenda en Figura 6 Pag 48
FIGURA 24 Archivos dentro de apache-jena-fuseki-2.6.0.zip pag 51
FIGURA 25 Inicio del servidor Fuseki Jena Apache Pag 51
FIGURA 26 mirnato.owl ha sido cargado con éxito Pag 52
FIGURA 27Ventana para realizar consultas SPARQL Pag 52
Fesuki Jena Apache devuelve información de la consulta “SELECT * {mir:hsa-let-7a-5p mir:has_predicted_target ?miRNA_target_gene} Pag 53
TABLA 5. Comparación de MIRNATO respecto a sus ontologías antecesoras. Los ontologías están formadas por conjuntos de axiomas, y estos son afirmaciones que determinan lo que es cierto en el dominio (ya sea sobre una clase o culaquier otra característica) (Motik et al., 2009). Para entender el resto de elementos de la tabla, consultar Apartado 2.3. Pag 54
FIGURA 28Gráfico con información acerca del miRNA de ejemplo hsa-let-7a-5p. Consultar leyenda de las propiedades en Figura 28.2
FIGURA 28.2Leyenda de la figura 28

1. Introducción

1.1 Contexto y justificación del Trabajo

1.1.1 Descripción general

Los miroRNAs (o miRNAs) son una clase de RNA monocatenário no codificante de unos 19 a 25 nucleótidos de longitud, que son mayoritariamente conocidos por inhibir la transcripción de mRNAs en proteínas o promover la represión de la expresión de mRNAs (Yue et al., 2009). El genoma humano codifica más de 1800 microRNAs, que son pequeñas moléculas de RNA no codificantes que regulan la expresión génica post-transcripcional. Debido a que potencialmente un miRNA puede tener múltiples tránscritos de genes como targets, los miRNAs son considerados como un importante mecanismo para regular la expresión génica y la transcripción de mRNA. (Peterson et al, 2014). Entendemos por target un gen diana cuya expresión se ve regulada por un miRNA.

Una ontología cataloga las variables requeridas para algún conjunto computacional y establece las relaciones entre ellos (Gruber, 1993). Las ontologías son formales, modelos de representación del conocimiento y juegan un rol clave en la en la definición semántica formal en la tradicional ingeniería del conocimiento. El ejemplo más exitoso de aplicar técnicas ontológicas en la investigación científica es el proyecto Gene Ontology (GO) (Huang et al, 2010).

Una gran cantidad de datos de miRNA y sus targets detectados experimentalmente son guardados en distintas bases de datos: ejemplos populares son microRNA.org[3] (Betel et al., 2008), miRBase[4] (Griths-Jones et al., 2006) y TarBase[5](Vergoulis et al., 2010). Actualmente existen algunas ontologías de miRNA/miRNA targets, como MicroRNA Target Prediction, pero ninguna de ellas representa toda la información necesaria acerca de miRNA (Dritsou et al, 2012).

El objetivo principal de este proyecto es cubrir la necesidad de una ontología que represente de manera satisfactoria toda la información acerca de miRNA. Para ello, se ha desarrollado una nueva ontología de miRNA con una mayor capacidad de representación que sus antecesoras. Se han analizado las limitaciones de las ontologías existentes, para posteriormente diseñar la nueva ontología con nuevas clases y propiedades. También se ha desarrollado una plataforma de

[3] http://www.microrna.org/microrna/home.do (2017)
[4] http://www.mirbase.org/ (2017)
[5] http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index (2017)

consulta, para poder extraer información acerca de cuestiones concretas relacionadas con los miRNA.

1.1.2. Justificación del TFM

Los microRNAs juegan un papel clave en la regulación de la expresión génica, y su estudio en el paradigma actual requiere el entendimiento y el dominio de distintas herramientas bioinformáticas. Las ontologías permiten el intercambio de información, determinando una terminología común y una base de conocimiento sólida y coherente.

Dado que no existe ninguna ontología de miRNA que represente de manera satisfactoria todas las particularidades de cada microRNAs, este proyecto responde dicha necesidad mediante el desarrollo de una nueva ontología. Se han identificado las limitaciones de las ontologías existentes sobre miRNA (los conceptos que no se podían modelar en ellas) para después extender una ontología con nuevos conceptos y propiedades.

No solo es importante el diseño de la ontología en sí misma, sino también poder realizar consultas que otorguen un enfoque más práctico y de mayor utilidad. Para potenciar su utilidad se ha desarrollado una plataforma de consulta que permite realizar consultas de tipo SPARQL.

El uso de herramientas bioinformáticas no se ha limitado al del estudio de los miRNAs sino también a aquellas necesarias para el diseño de su ontología, que ha requerido además el aprendizaje y puesta en práctica de muchas de las mismas.

.

1.2 Objetivos del Trabajo

1.2.1 Objetivos generales

- Extender una ontología para la representación de microRNAs.
- Implementar una plataforma y una base de datos para gestionar la información en la ontología creada mediante sparQL.

1.2.2 Objetivos específicos

- Identificar las limitaciones existentes en la representación de microRNAs.
- Identificación de las características que serán necesarias dentro de una ontología de miRNAS
- Compilar toda la información sobre miRNA targets relevante para

la ontología en una base de datos
- Diseñar una plataforma de consulta que permita el acceso a la información de la ontología
- Implementar en la web la posibilidad descarga de información de la ontología en distintos formatos

1.3 Enfoque y método seguido

En todos los enfoques es necesario un estudio inicial de las limitaciones para la representación de microRNAs en fuentes bibliográficas. No obstante, este proyecto tiene **dos enfoques posibles**: la creación desde cero de una ontología de microRNAs, o por el contrario, la ampliación de una ontología ya existente de microRNAs.

En el **primer enfoque**, inicialmente se indaga sobre el funcionamiento de otras ontologías para intentar desarrollar una ontología con un funcionamiento análogo adaptado a miRNA, comparando distintas ontologías y analizando sus virtudes y limitaciones. A continuación se realiza la selección de las características más relevantes de los microRNAs. El siguiente paso es el diseño de la ontología, que más adelante se podrá aplicar sobre las instancias de una base de datos que compilará toda la información disponible sobre microRNAs. Finalmente se desarrolla una web de consulta complementaria al proyecto que implemente la base de datos de microRNAs desarrollada basada en la ontología.

En el **segundo enfoque**, se realiza una exhaustiva búsqueda bibliográfica para hallar ontologías de microRNAs ya desarrolladas, para posteriormente ampliarlas o extenderlas. No obstante, esta opción es descartada en caso de no tener acceso a dicha la ontología. En caso de tener acceso y/o permiso para ampliarla, se lleva a cabo un proceso parecido al anterior, con la diferencia de que no se parte desde 0 en su diseño sino que está basado en el ya establecido y se centra en su ampliación y mejora. Su aplicación en una base de datos se lleva a cabo igual que en el enfoque anterior, de la misma manera que la plataforma web.

1.4 Planificación del Trabajo

Para este trabajo, son necesarios recursos básicos como un ordenador con conexión a internet, que contenga un sistema operativo capaz de soportar algún software para editar ontologías en algún formato que soporte una ontología (por ejemplo .owl). En este caso el proyecto se lleva a cabo en un sistema operativo Windows 7 Home Premium (64 bits) en un HP Compaq Presario CQ70. Para la extensión se usa el

software de edición de ontologías Protégé 5.2.0[6] (software gratuito y libre de la Universidad de Stanford).

La planificación temporal de las tareas se puede observar en la Tabla 1, y la de los hitos en la Tabla 2. De un modo más gráfico, también se pueden observar en el Diagrama de Grantt de la Figura 1

Tarea	Inico	Final
Inspeccionar y determinar el funcionamiento de otras webs de ontologías (Plant Ontology, Gene Ontology, etc).	21/03	24/03
Indagar sobre la existencia de otras ontologías de microRNAs y analizar su funcionamiento (sus ventajas y limitaciones)	24/03	29/03
Realizar una búsqueda bibliográfica para determinar qué es relevante en la representación de miRNAs y sus limitaciones	29/03	03/04
Establecer la necesidad de extender una ontología existente o de crear una nueva	03/04	05/04
Entrega de la PEC2	05/04/17	
Definición e implementación de la ontología	05/04	13/04
Desarrollo de una base de datos sobre miRNA targets basada en la ontología desarrollada	13/04	21/04
Desarrollo de una plataforma de consulta basada en la ontología	21/04	01/05
Evaluación de su funcionamiento	01/05	09/05
Entrega de la PEC3	**10/05/17**	
Memoria y presentación del trabajo final	**11/05/17 - 24/05/17**	
Defensa publica	**07/06/17 – 21/06/17**	

TABLA 1 Descripción cronólogica de las tareas

HITO	DESCRIPCION	FECHA
Hito 1	Breve resumen del state of the art de las ontologías actuales de miRNAs	29/03/17
Hito 2	Determinar si se diseñará una ontología des de 0 o se partirá de una pre-establecida	04/04/17
Hito 3	Entrega de la PEC2	**05/04/17**
Hito 4	Creación de la base de datos sobre miRNA targets	24/04/17
Hito 5	Creación de una plataforma web	02/05/17

[6] https://protege.stanford.edu (2017)

Hito 6	Entrega de la PEC3	**10/05/17**
Hito 7	**Memoria y presentación del trabajo final**	**11/05/17 - 24/05/17**
Hito 8	**Defensa publica**	**07/06/17 – 21/06/17**

TABLA 2 Descripción cronólogica de los hitos.

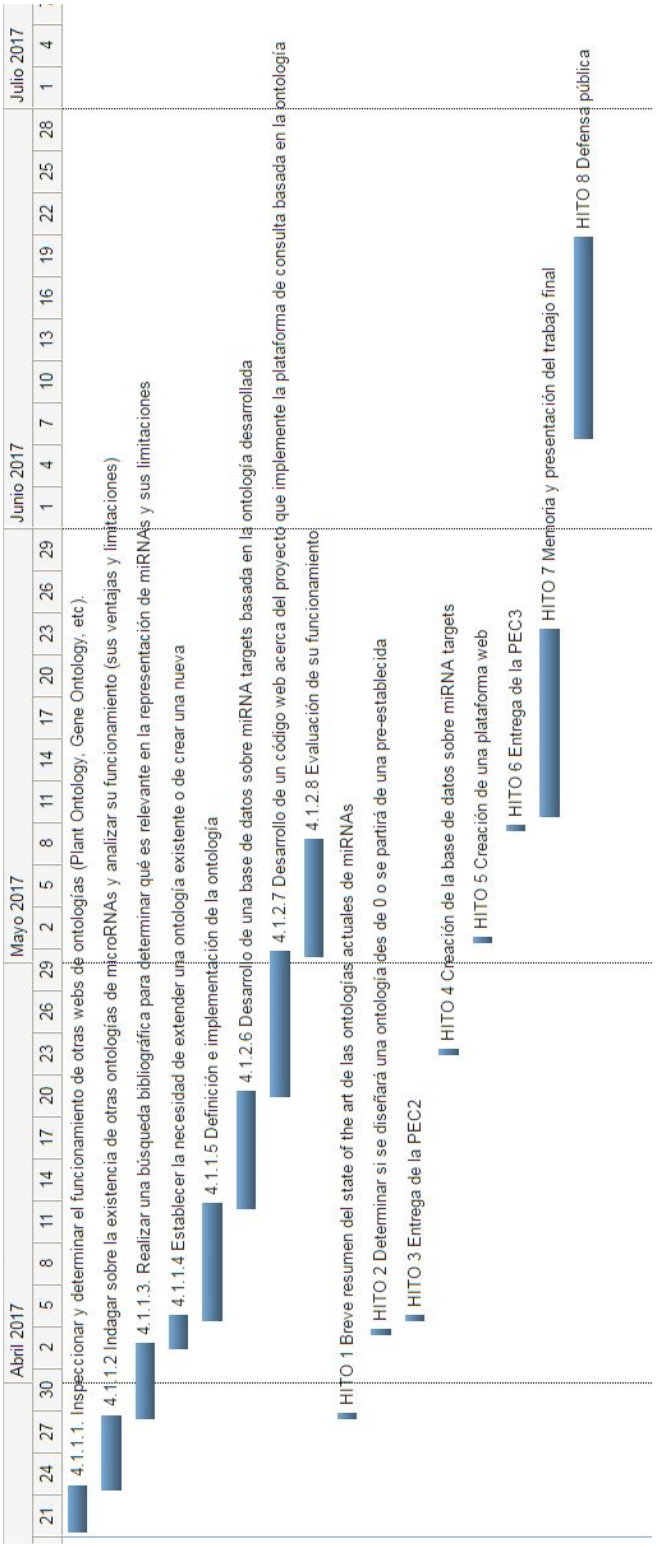


FIGURA 1. Diagrama de Grantt con tareas e hitos

1.5 Breve sumario de productos obtenidos

- Memoria del Trabajo
- Plataforma de consulta
- miRNA Target Ontology (MIRNATO)
    - Debido a derechos de autor, la distribución posterior a la modificación de MIRNAO no puede realizarse bajo el mismo nombre de la ontología original. Es por ello que se le ha dado un nuevo nombre a la nueva ontología fruto de la extensión de MIRNAO: “MIRNATO”.

1.6 Breve descripción de los otros capítulos de la memoria

En el **Capítulo 2** se aporta información acerca de miRNA que no tiene un impacto directo en el diseño de la ontología, pero sí es importante para entender los conceptos que se quieren modelar. Trata sobre su historia y definición, las vías biológicas canónicas de biogénesis y herramientas computacionales. También se aporta información acerca de las ontologías, sobre qué son y su importancia en el ámbito de la investigación.

- Capítulo 2. Introducción a miRNA y Ontologías
    - Capítulo 2, Capítulo 2.1 Descripción y contexto de los miRNA.
        - Capítulo 2, Capítulo 2.1, Capítulo 2.1.1 Origen y definición
        - Capítulo 2, Capítulo 2.1, Capítulo 2.1.2 Biogénesis canónica en animales
        - Capítulo 2, Capítulo 2.1, Capítulo 2.1.3 Herramientas computacionales
    - Capítulo 2, Capítulo 2.2 Ontologías
        - Capítulo 2, Capítulo 2.2 Ontologías, Capítulo 2.2.1 Necesidad de análisis ontológico
        - Capítulo 2, Capítulo 2.2 Ontologías, Capítulo 2.2.2 Enfoque del trabajo: Extensión o diseño des de 0
    - Capítulo 2, Capítulo 2.3 Partes de una Ontología

Una vez los conceptos a modelar han sido determinados en el Capítulo 2, es necesario comprobar cuáles se pueden representar con MIRNAO, cuáles se pueden representar con términos de otras ontologías y cuales deben crearse nuevos. Este **Capítulo 3** trata sobre este estudio de la capacidad de representación de miRNAO y su extensión, que consiste en la creación de las clases determinadas en el punto anterior, junto con la creación de relaciones (o “*Object Propierties*”) que hay entre ellas con

el software *Protégé 5.2.0* en la el archivo de la ontología MIRNAO *mirnao.owl*.

- Capítulo 3. Ontologías para el modelaje de dianas de microRNA
    - Capítulo 3, Capítulo 3.1 Enfoque del trabajo: Extensión o creación de una nueva ontología
    - Capítulo 3, Capítulo 3.2 Extensión de la Ontologia de miRNAs
        - Capítulo 3, Capítulo 3.2, Capítulo 3.2.1 Conceptos a modelar
        - Capítulo 3, Capítulo 3.2, Capítulo 3.2.2 Retos y dificultades
        - Capítulo 3, Capítulo 3.2, Capítulo 3.2.3 Extensión

En este **Capítulo 4** se ha usado un servidor Fuseki Jena Apache que permite realizar consultas mediante SPARQL de manera local. Una ontología puede contener Individuos que instancian una clase concreta. Esto significa que podemos almacenar información de uno o varios casos concretos de miRNA.

Además, en este Capítulo 4 también se almacena en forma de individuos que instancian clases (donde las clases representan los conceptos generales que se pretende modelar) la información obtenida en el punto anterior acerca del miRNA hsa-let-7a-5p.

- Capítulo 4, La plataforma de consulta
    - Capítulo 4, Capítulo 4.1, Almacenaje de miRNA
- Capítulo 5. Resultados
- Capítulo 6. Discusión y Conclusiones
- Capítulo 7. Bibliografía
- Capítulo 8. Anexos
    - Capítulo 8, Capítulo 8.2 Glosario de términos ontológicos

En muchas imágenes de OntoGraf aparecen términos que pueden ser de difícil comprensión. En este glosario (Capítulo 8, Capítulo 8.2) están todas las definiciones.

2. Introducción a microRNA y Ontologías

2.1 Descripción y contexto de los miRNA

En este capítulo se intenta explicar de forma clara y sencilla qué es un miRNA, cuando se descubrieron y porqué son importantes.

2.1.1 Origen y definición

El primer miARN fue descubierto hace más de 30 años en el nematodo *Caenorhabditis elegans* con la identificación del regulador de desarrollo lin-4. Originalmente se creía que era un gen convencional de codificación de proteínas, pero los laboratorios de Ruvkun y Ambros hicieron el descubrieron que lin-4 no codificaba una proteína sino que codificaba un ARN regulador de 22 nucleótidos. Ellos demostraron que el ARN lin-4 podía unirse con el ARNm de otro gen en la red de desarrollo de *C. elegans*, lin-14, y controlar la producción de la proteína LIN-14. (Hammond S., 2015)

Este descubrimiento habría tenido un significado limitado fuera de la comunidad científica, si en *C. elegans* no se hubiera descubierto un segundo miRNA, let-7. Más importante aún, let-7 se conserva a través de muchos organismos, incluidos los seres humanos, lo que sugiere que esta clase de pequeños RNA reguladores tienen un papel más general en la biología. El otro impresionante hallazgo, ocurriendo casi simultáneamente, fue el descubrimiento de la vía de RNA de interferencia (o RNAi); específicamente, la vía de ~ 21 nucleótidos de ARN que desencadena una maquinaria de silenciación. (Hammond S., 2015)

Se ha demostrado que estas dos vías son diferentes brazos de la misma vía de silenciamiento de genes. Posteriormente, muchos miles de miRNAs se han descubierto en muchos organismos, y actualmente hay más de 2588 miRNAs anotados en el genoma humano. (Hammond S., 2015). Debido a que potencialmente un miRNA puede tener múltiples transcritos de genes como diana, los miRNAs son considerados como un importante mecanismo para regular la expresión génica y la transcripción de mRNA. (Peterson et al, 2014).

Podemos definir miroRNAs (o miRNAs) como una clase de RNA monocatenario no codificante de unos 19 a 25 nucleótidos de longitud

hallados en animales, plantas y algunos virus. Son mayoritariamente conocidos por inhibir la transcripción de mRNAs codificante o promover la represión de la expresión de mRNAs (Yue et al., 2009). Aunque la mayoría suelen hallarse dentro de la célula, algunos miRNA llamados “miRNA extracelulares” suelen encontrarse en ambientes extracelulares incluyendo fluidos biológicos. (Sohel M., 2016)

Más allá de su importancia, los miRNA no solo están conservados en distintas especies cumpliendo funciones parecidas sino que además la gran mayoría tienen un mecanismo de biogénesis y actuación muy parecido, al que se ha llamado “vía de biogénesis canónica” (existe una para animales y otra para plantas).

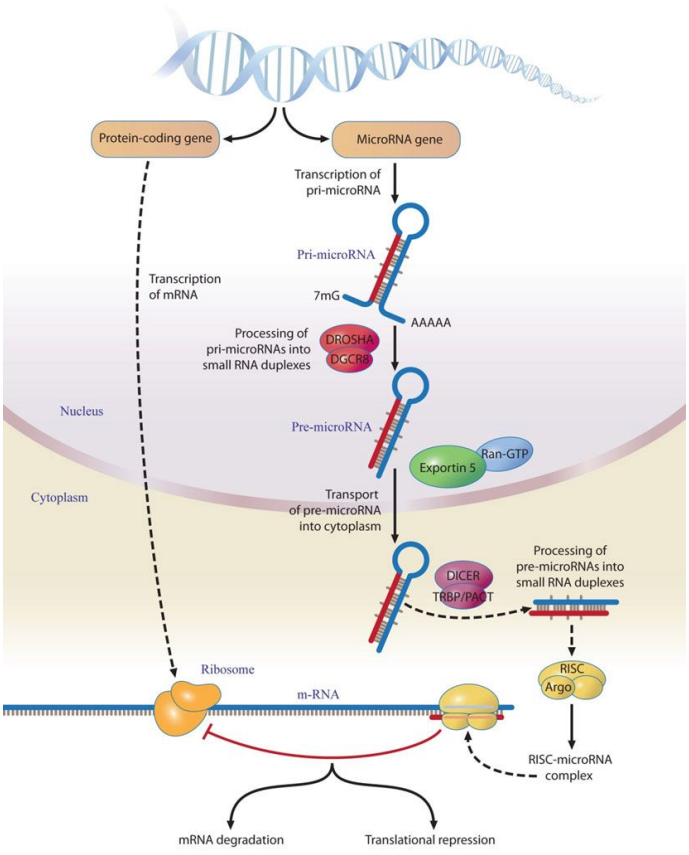
Los miRNA también se encuentran en plantas (incluyendo algas), donde tiene un comportamiento bastante diferente al que se observa en animales. Tan diferente es su comportamiento que no está claro si se trata de un mecanismo de silenciamiento de genes conservado en plantas y animales, o por el contrario los miRNA aparecieron en plantas de manera independiente. En plantas, la complementariedad del miRNA con su target es casi perfecta en una región codificante del mRNA, a diferencia de los animales donde los miRNA se caracterizan por reconocer a sus targets (entre otras características) por la complementariedad en una pequeña región de 6-8 nucleótidos en la zona 3’UTR. En el proceso de biogénesis (ver Capitulo 2.1.2) las plantas no sufren una escisión dentro del núcleo y luego otra fuera, sino que sufren ambas en el nucleoplasma con una enzima homóloga a DICER llamada DLC1. (Axtel et al., 2011) (Wang et al., 2004)

En su estudio, las herramientas computacionales han jugado un papel clave, principalmente en la identificación de nuevos miRNA y en la predicción de sus targets. Sin embargo, siempre es necesario realizar una validación experimental de la predicción de sus targets, ya sea con métodos directos o indirectos.

En el siguiente capítulo 2.1.2 se hará una breve explicación de la vía canónica en animales.

2.1.2 Biogénesis canónica en animales

FIGURA 2 Biogénesis canónica en animales (Arenz C., 2014)



El conocimiento de las rutas de biogénesis es muy importante, especialmente si se trata de una ontología que aspira a poder representar todos sus conceptos. En especial porque muchas terapias médicas que están en proceso de desarrollo para tratar enfermedades relacionadas con miRNA tienen una relación directa con estos pasos.

La biogénesis de miRNA (Figura 2) se inicia en el núcleo, donde se transcribe en miRNA primario (o pri-miRNA) por la RNA polimerasa II o III. La transcripción de genes de miRNA produce largos transcritos de pri-miRNA con una pliegue local. El siguiente pri-miRNA es cortado dando lugar al precursor del miRNA (pre-miRNA) en el núcleo por el complejo procesador formado por Drosha y su cofactor Pasha. (Akhtar et al., 2015)

El pre-miRNA resultante en forma de horquilla es exportado del núcleo al citoplasma por el complejo formado por Exportina 5 y Ran-GTP. En el citoplasma, la enzima Dicer de la RNasa III unida a TAR RNA binding protein (o TRBP) recorta la horquilla de pre-miRNA hasta su longitud madura (~21-nt), dando lugar al duplex miRNA:miRNA. (Akhtar et al., 2015)

El duplex es separado y el miRNA maduro se carga en las proteínas de Argonaute (Ago 2) dentro del *RNA-induced silencing complex (RISC)*. Una vez el miRISC ha sido constituido, el miRNA lo conduce a silenciar la diana mRNA via corte del mRNA, represión transcripcional o deadenilación. (Akhtar et al., 2015)

2.1.3 Herramientas computacionales

Es importante tener claro el proceso de biogénesis, porque muchas de las herramientas computacionales disponibles están especializadas o son relevantes en alguna paso concreto del proceso. Un ejemplo son las bases de datos que almacenan información acerca de los Factores de

Transcripción validados experimentalmente que interactúan con cada miRNA (Figura 3).

Dada la implicación de los miRNA en la regulación génica y en distintos procesos relacionados con enfermedades, los experimentos se han multiplicado generando a su tiempo un aumento exponencial de información aportada por miles de artículos. Un gran número de herramientas bioinformáticas están disponibles para gestionar esta enorme corriente de datos. Tanto la investigación básica como la aplicada están siendo mejoradas por **herramientas computacionales y bases de datos** (Akhtar et al, 2015).

Para diseñar una ontología para el modelado y el estudio de miRNA, es necesario tener en cuenta sus características en los procesos de **biogénesis y actuación** pero también tener en cuenta cuales son los aspectos más importantes en las bases de datos y herramientas actuales, de manera que pueda ser una **ontología útil y práctica** para determinar y compartir información.

Dado que existe una gran cantidad de herramientas bioinformáticas relacionadas con miRNA y cada una de ellas toma en consideración unos aspectos u otros, es necesario que en la ontología se representen por lo menos las **características más comunes**.

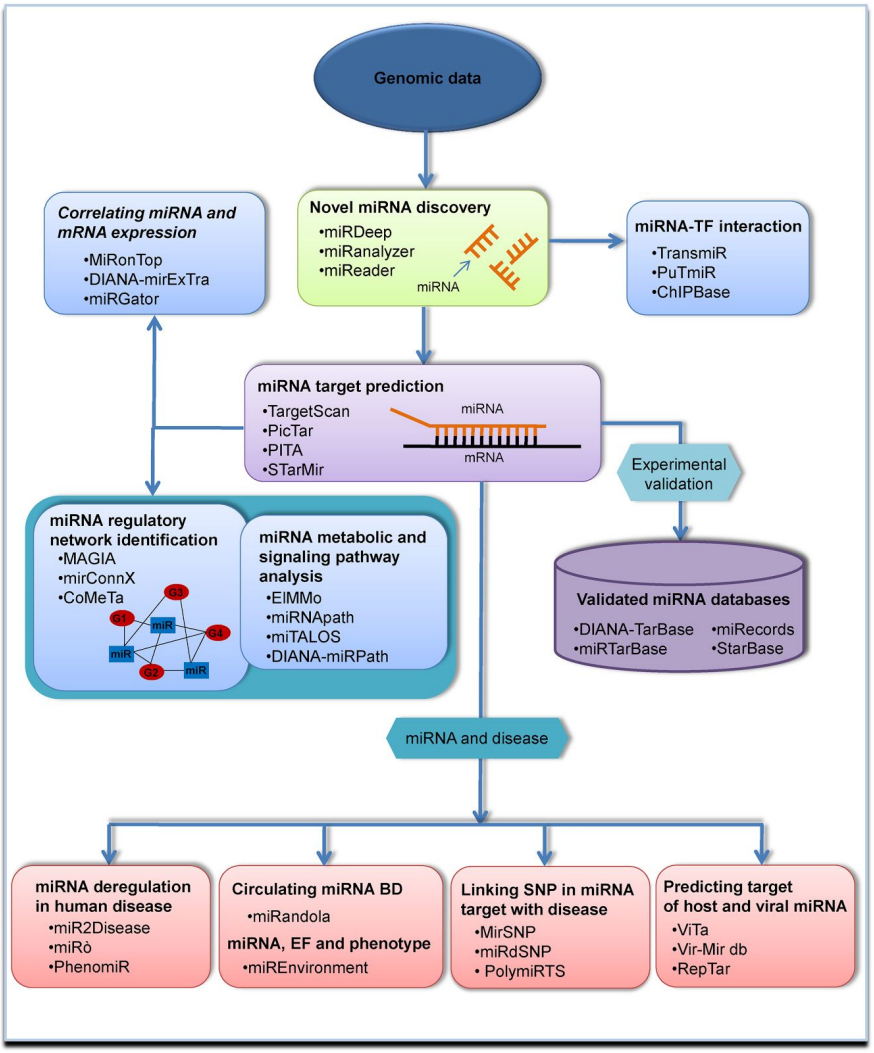


FIGURA 3. Pequeño esquema con algunas de las herramientas y bases de datos disponibles relacionadas con miRNA (Akhtar et al., 2015)

Algunas bases de datos (BBDD) como MiRBase[7] (Griths-Jones et al., 2006) o miRDB[8] (Want X., 2008) son simplemente anotaciones de texto llano que almacenan información de muchos miRNA (no tienen una estructura ontológica). Otras como DIANA-TarBase[9] (Vergoulis et al., 2010) se centran en la recopilar información acerca de la validacion experimental de genes diana de miRNA.

Más allá del simple almacenamiento de información, existen también algoritmos computacionales con distintas funciones. Aquí destacaremos dos tipos de herramientas clave: Primero, las que permiten la identificación de nuevos miRNA como miRDeep (Friedländer et al., 2008). Segundo, las que permiten predecir miRNA targets *de novo*, como por ejemplo PITA[10] (Kertesz et al., 2007) o TargetScan[11] (Agarwal

[7] http://www.mirbase.org/ (2017)
[8] http://mirdb.org/ (2017)
[9] http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index (2017)
[10] https://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html (2017)
[11] http://www.targetscan.org/vert_71/ (2017)

et al., 2015).

Sin embargo, como veremos más adelante en el apartado 3, no todas almacenan la misma información ni usan las mismas características miARN para llevar a cabo la predicción de genes diana (Tabla 3).

2.2 Ontologías

Las herramientas computacionales son necesarias, pero no sustituyen la función que desarrolla una ontología. En cualquier caso, antes debe quedar claro qué es una ontología, ya que si bien el uso del término lleva años usándose en el mundo de la investigación, se trata de un concepto bastante abstracto que puede dar lugar a malentendidos.

Una ontología (del griego ντος 'del ente', genitivo del participio del verbo εἰμί 'ser, estar'; y λόγος 'ciencia, estudio, teoría')[12] es una representación de vocabulario, normalmente especializada en algún dominio o temática. Más concretamente, no es sólo el vocabulario lo que la convierte en una ontología, sino las conceptualizaciones que los términos del vocabulario intentan capturar. (Chandrasekaran, 1999). Una ontología cataloga las variables requeridas para algún conjunto de computación y establece las relaciones entre ellos (Gruber, 1993).

Las ontologías son formales, modelos de representación del conocimiento y juegan un rol clave en la en la definición semántica formal en la tradicional ingeniería del conocimiento. El ejemplo más exitoso de aplicar técnicas ontológicas en la investigación científica es el proyecto Gene Ontology (GO) (Huang et al, 2010).

Sin duda, una de las más grandes e importantes utilidades de una ontología es que facilita el intercambio de información entre diferentes herramientas, bases de datos, etc. (independientemente de la fuente de origen o destino de dicha información). Si en su lugar se usaran esquemas locales de codificación, el intercambio de información sería mucho más complicado.

Una ontología proporciona una conceptualización explícita (es decir, una meta-información) que describe la semántica de los datos. Tienen una función similar a un esquema de base de datos. Las diferencias son:

- El lenguaje para definir ontologías es sintácticamente y semánticamente más rico que los enfoques comunes para las bases de datos.
- La información descrita por una ontología consiste en textos semi-estructurados en lenguaje natural y no en información tabular.
- Una ontología debe ser una terminología común y consensuada

[12] https://es.wikipedia.org/wiki/Ontolog%C3%ADa (2017)

porque se utiliza para el intercambio de información.
- Una ontología proporciona una teoría de dominios y no la estructura de un contenedor de datos (Fensel et al., 2000)

Existen actualmente ontologías altamente relacionadas con miRNA:
- RNA Ontology / RNAO (Hoehndorf et al., 2011)
- Non Coding RNA Ontology / NCRO (Huang et al., 2016)
- Ontology of MicroRNA Target / OMIT (Huang et al., 2012)
- miRNA Ontology / MIRNAO (Dritsou et al., 2012)

En el siguiente apartado se explica más detenidamente su relación con el trabajo.

2.2.1 Necesidad de análisis ontológico

Una vez ha quedado claro el concepto de “Ontología” en el ámbito de la investigación y entender qué las distingue de una base de datos, se puede responder a la pregunta ¿porqué es necesario?

En primer lugar, existe la necesidad *per sé* en el ámbito de la investigación biomédica de desarrollar las ontologías pertinentes en cada ámbito del conocimiento:

- El análisis ontológico aclara la estructura del conocimiento. Dado un dominio del conocimiento, su ontología forma la base de cualquier otro sistema de representación del conocimiento dentro de ese mismo dominio. Sin ontologías no puede haber un vocabulario para representar el conocimiento. **Débiles análisis ontológicos conducen a bases de conocimiento incoherentes.** (Chandrasekaran et al., 1999)

- En la búsqueda de lo que es biológica y clínicamente significativo en los conjuntos de datos generados por las tecnologías de alto rendimiento actuales, **una estrategia común es la creación y el análisis de 'anotaciones' que enlazan datos primarios con expresiones en vocabularios estructurados y controlados**. Datos disponibles para la búsqueda y procesamiento algorítmico (Smith et al., 2007). Sin ontologías bien definidas, esta estrategia no sería posible.

- El dominio de la biología molecular está caracterizado por la disponibilidad de grandes cantidades de datos bien definidos que pueden utilizarse sin restricción para el procesamiento algorítmico. Por el contrario, en el ámbito clínico sólo se dispone de cantidades limitadas de datos para fines de investigación, y éstos siguen siendo mayoritariamente de texto en lenguaje natural. Incluso cuando se dispone de datos clínicos más

sistemáticos, **el uso de esquemas locales de codificación significa que estos datos no se acumulan de formas útiles para la investigación** (Smith et al., 2007). **Las ontologías permiten el intercambio de conocimiento** (Chandrasekaran et al., 1999)

Por otro lado, es necesaria una ontología de miRNA por razones más concretas:

- Dada la implicación de miRNAs en la regulación de genes así como de procesos vinculados a distintas enfermedades, los experimentos han aumentado a una gran velocidad, generando un **flujo exponencial de datos** dispersos en miles de artículos (Akhtar et al., 2015). Una ontología de miRNA es necesaria para poder compartir información de manera que los datos estén estructurados en formas útiles para la investigación.

- A pesar de que existen ontologías altamente relacionadas con miRNA, como por ejemplo: RNA Ontology / RNAO (Hoehndorf et al., 2011), Non Coding RNA Ontology / NCRO (Huang et al., 2016) o Ontology of MicroRNA Target / OMIT (Huang et al., 2012); no todas cubren exactamente el mismo dominio del conocimiento. OMIT concentra su potencial en los roles de miRNAs en el cáncer humano. RNAO y NCRO son muy generales y tampoco satisfacen la necesidad de una ontología de miRNAs. En el primer caso, RNAO, porque su objetivo es capturar el conocimiento general de RNA en todos los aspectos, centrándose también en la estructura primaria, secundaria y terciaria de los alineamientos (Dritsou et al, 2012).
    - **La información específica de miARN, por lo tanto, no se encuentra en el dominio público**, lo que contrasta con el número de bases de datos específicas de miARN (incluyendo las de genes diana de miRNA o miRNA targets) que cubren información del miARN y miARN target. Algunos ejemplos son miRBase (Griths-Jones et al., 2006), miRGEN 2.0 (Alexiou et al., 2010) y TarBase 6.0 (Vergoulis et al., 2010), o que cubren organismos específicos como PMRD para las plantas (Zhang et al., 2012) y MirFANs (Liu et al., 2012), una base de datos específica para *Arabidopsis thaliana* miRNAs. (Dritsou et al, 2014)

Una prueba de que cubren dominios del conocimiento ligeramente distintos es la *Tabla 4 (ver anexo)*, que muestra el origen de los términos de tres de las ontologías relacionadas con miRNA. De forma más resumida, se puede apreciar ciertas diferencias entre algunas clases en la *Figura 4*.

FIGURA 4 Comparación de algunas clases de MIRNAO y NCRO. A la izquierda, una captura de pantalla de MIRNAO en Online Look up Service (OLS)[13] de EMBL. A la derecha, NCRO en OLS[14] de EMBL.

La ontología de miRNA que más se adecua al objetivo principal de este proyecto es miRNA Ontology / MIRNAO (Dritsou et al, 2014). Sin embargo, en estos últimos años ha habido distintos hallazgos muy relevantes en el ámbito de miRNA que han cambiado la forma de entender estos oligonucleotidos no codificantes, **dejando MIRNAO obsoleta**.

Este proyecto responde la necesidad actual de una ontología que sea capaz de cubrir de forma satisfactoria la representación de las particularidades de cada microRNA.

2.3 Partes de una ontología

El desarrollo más reciente en lenguaje standard de ontologías es OWL de World Wide Web Consortium (W3C). Protege y OWL permiten describir conceptos pero también otras facilidades.

Los **individuos** representan objetos en el dominio en el que estamos interesados. Los individuos también se llaman instancias, o se puede hacer referencia a los individuos como “instancias de clases”. (Horridge et al., 2011)

Las **propiedades** son relaciones binarias de individuos (por ejemplo, las propiedades relacionan dos individuos conjuntamente). Las propiedades

[13] http://www.ebi.ac.uk/ols/ontologies/mirnao (2017)
[14] http://www.ebi.ac.uk/ols/ontologies/ncro (2017)

pueden establecerse entre objetos (clases) o individuos. (Horridge et al., 2011)

Las propiedades pueden ser de varios tipos, como por ejemplo inversas (*por ejemplo, tiene_padre* seria inversa de *tiene_hijo*), transitorias o simétricas. En este trabajo no se entrará en detalle a describir todos los tipos de propiedades, para más información puede consultarse el document (Horridge et al., 2011)

Las **clases** en OWL se interpretan como conjuntos que contienen individuos. Se determinan usanto descripciones formales (matemáticas) que indican con precisión las condiciones para formar parte de esa clase. (Horridge M., 2011)

3. Ontologías para el modelaje de dianas de microRNA

3.1 Enfoque del trabajo: Extensión o creación de una nueva ontología

Ahora ya se ha aportado información acerca de qué es un miRNA, qué es una ontología y por qué es necesario el análisis ontológico. Antes de empezar hay que tomar la decisión sobre qué enfoque se adapta mejor a los objetivos y las condiciones del proyecto, y este capítulo trata de abarcar las virtudes y defectos principales de extender una ontología existente o crear una nueva.

El diseño de una ontología de ámbito biomédico puede conllevar meses o incluso años de trabajo. Es una labor que requiere el **conocimiento minucioso de las características** del objeto de estudio y de las relaciones existentes (tanto de los elementos en los que se divide como de otros superiores de los que forma parte y/o interacciona). Dado que este proyecto tuvo tan solo unos meses para desarrollarse, es prácticamente imposible poder desarrollar una ontología de manera adecuada con tan poco tiempo.

Además, no es conveniente que si se intenta determinar una ontología para así mejorar el entendimiento y la compartición de un ámbito del conocimiento concreto, en este caso miRNA, se desarrollen ontologías paralelas con criterios distintos. **Toma prioridad desarrollar ontologías ya diseñadas** para después extenderlas. (Noy et al., 2001) (Uschold et al., 1996)

Es cierto que esta decisión conlleva ciertas limitaciones, principalmente tener que **aceptar una gran estructura pre-establecida** de relaciones e interconexiones entre los elementos de mi-RNA que difícilmente podrá estar sujeta a cambios en sus fundamentos o elementos base. En caso de surgir discrepancias, partir de una ontología pre-diseñada condiciona al autor que la extiende a aplicar dichos cambios debido a que puede alterar la estructura general u otras relaciones entre elementos.

A pesar de estas limitaciones, se considera que las **ventajas sobrepasan con creces a las limitaciones** y por ello se determina que el mejor enfoque es partir de una o varias ontologías pre-establecidas para su futura extensión.

Finalmente, debido a que miRNAO es la ontología ya existente que mejor se adapta al objetivo de este proyecto, es la ontología que se ha

extendido.

3.2 Extensión de la Ontología de miRNA

3.2.1 Conceptos a modelar

Una vez realizada la fase introductoria y decidido el enfoque, el primer paso es decidir qué conceptos se quiere modelar. Este capítulo trata sobre este primer paso, justificando brevemente el porqué de cada concepto.

No existe ningún documento que indique de forma concreta la lista de conceptos que deben modelarse en una ontología de miARN. Sin embargo, como el objetivo de MIRNATO es representar toda la información posible acerca de un miARN, los conceptos a modelar se pueden obtener intentando representar toda la información disponible de algún/algunos miARN. En otras palabras, el mero hecho de intentar representar la información disponible en la literatura de algunos miARN concretos ya aporta información sobre las limitaciones de la ontología y cómo extenderla.

Para determinar qué conceptos se quieren modelar es imprescindible realizar una búsqueda bibliográfica que permita entender las vías canónicas de biogénesis de miRNA (*Figura 2*) y las características relevantes del miRNA en relación al gen diana sobre el que tendrá un efecto regulador.

A continuación se enumera, por temáticas y junto con una breve explicación, la lista de conceptos que se ha decidido representar en la ontología:

Identificación / Referencias
Cualquier investigador que desee obtener información acerca de un miRNA (y su target) tendrá que recurrir, en primer lugar, a la literatura. Es por ello que los primeros conceptos que se necesitan modelar hacen referencia a las referències bibliográficas que puedan tener el miRNA y su target:

- Literatura relacionada con el miRNA
- Información de las BBDD de referencia de donde se extrae el miRNA y el mRNA
    - ID
    - Versión de la BBDD
- Tipo de validación empírica
    - Las predicciones computacionales deben comprobarse posteriormente de forma experimental (Kuhn et al., 2007)
- Predicción computacional del miARN y su target (Akhtar et al.,

2015)
- Métodos de predicción computacional del miARN y su target
    - Cada día son más y más variadas las herramientas computacionales que realizan predicciones *in silico* de targets de miARN. (Akhtar et al., 2015)

Propiedades de localización & Clasificación del miRNA y su target
Todos los miARN tienen algún origen. La mera condición de miRNA implica que este pequeño fragmento de ARN pertenece a algún gen, de algún cromosoma, de algun tejido, de algún organismo. Además, cada miARN se puede clasificar en una familia de miARN concreta o puede tener una secuencia que sea una isoforma/variación de un miRNA:

- Organismo al que pertenece
- Tejido al que pertenece
- Conservación entre especiès
    - Los miRNA que estan conservados entre especies suelen tener funciones parecidas y ser altamente importantes en la regulación génica (Lewis et al., 2005)
- La familia del miRNA
    - Los miRNA de una misma familia tienen antecesores comunes que han sido conservados en un conjunto de especies concreto. (Zou et al., 2014)
- Región del genoma donde se encuentra
- Gen del que forma parte el miRNA
- Gen diana del miRNA
- Tipos de *seed region*
    - Hay un número limitado de tipos de *seed region*, y es importante tenerlo en cuenta porque puede alterar el resultado de una predicción algorítmica computacional si no está correctamente configurado (Ellwanger et al., 2011)
- IsomiRs, isoformas de miARN
    - A pesar de que no es un ámbito que haya sido tan estudiado como otros aspectos de los miARN, se sabe que algunos pueden tener isoformas en 3' ó 5', además de polimorfismos. (Neilsen et al., 2012)

Propiedades de la vía biológica
Todas las formas de miARN maduros han pasado antes por un proceso de biogénesis y modificación, aunque los pasos concretos pueden ser diferentes según el miARN. Dado que existe una gran variedad de vías biológicas distintas, los conceptos que se enumeran a continuación están basados principalmente en las vías canónicas de maduración en animales (Hammond et al., 2015) y plantas (Xie et al., 2015):

- Actividad enzimática desde pri-miRNA hasta llegar a pre-miRNA
- Proteïnas que lo transportan al citoplasma
- Actividad enzimática que procesa el pre-miRNA hasta miRNA maduro

- Actividad enzimática después de formarse el dúplex miRNA:miRNA
- Elementos que regulan la desaparición de miRNAs
    - Dado que no todos los genes que regulan los miARN van a sufrir una represión permanente, son necesarios procesos que eliminen los miARN. Generalmente se eliminan mediante actividades enzimáticas que generan inestabilidad en las estructuras primarias y secundarias (Kai et al., 2010)

Transcription factor (TF) regulators:
GO:1902893 regulation of pri-miRNA transcription from RNA polymerase II promoter (or +/- regulation)
GO:0001085 RNA polymerase II transcription factor binding

Transcription factors (TFs):
GO:1902893 regulation of pri-miRNA transcription from RNA polymerase II promoter
GO:0000981 sequence-specific DNA binding RNA polymerase II transcription factor activity (or +/- regulation)

Microprocessor complex proteins (DROSHA, DGCR8):
GO:0070878 primary miRNA binding
GO:0031053 primary miRNA processing
GO:0070877 microprocessor complex
DROSHA: GO:0004525 ribonuclease III activity
GO:0090502 RNA phosphodiester bond hydrolysis, endonucleolytic

Exportin-5 complex (XPO-5, RAN):
GO:0035281 pre-miRNA export from nucleus
GO:0070883 pre-miRNA binding
GO:0042565 RNA nuclear export complex
XPO-5:
GO:0008536 Ran GTPase binding
GO:0090631 pre-miRNA transporter activity
RAN:
contributes_to GO:0090631 pre-miRNA transporter activity
contributes_to GO:0070883 pre-miRNA binding

RISC-loading complex proteins (DICER1, TARBP2/PRKRA, AGO1-4):
GO:0098851 double-stranded miRNA binding
GO:0010501 RNA secondary structure unwinding
GO:0035280 miRNA loading onto RISC involved in gene silencing by miRNA
GO:0035196 production of miRNAs involved in gene silencing by miRNA
GO:0070578 RISC-loading complex

DICER1:
GO:0004525 ribonuclease III activity
GO:0090502 RNA phosphodiester bond hydrolysis, endonucleolytic

Argonautes (AGO1-4):
GO:0035198 miRNA binding
GO:0010501 RNA secondary structure unwinding
GO:0016442 RISC complex
GO:0035278 miRNA mediated inhibition of translation
GO:0098806 deadenylation involved in gene silencing by miRNA

AGO2:
GO:0090624 endoribonuclease activity, cleaving miRNA-paired mRNA
GO:0035279 mRNA cleavage involved in gene silencing by miRNA

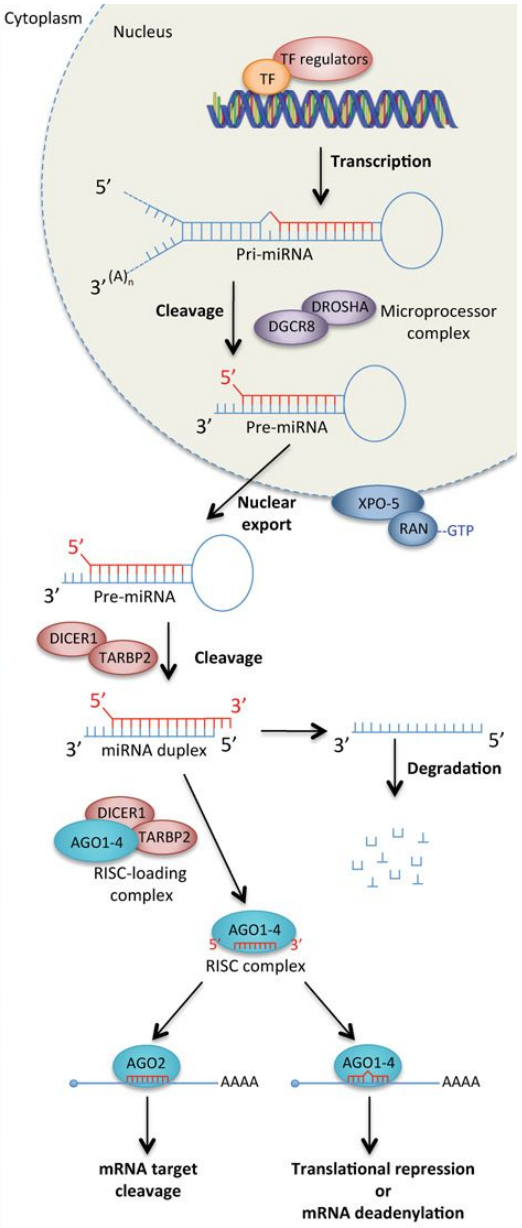


FIGURA 5. La ruta canónica de procesamiento de miARN animal (panel derecho) y los términos GO que se espera estén asociados con los componentes proteícos de esta vía (panel izquierdo). La asociación de estos términos GO dependerá de la evidencia experimental disponible; otros términos pueden estar asociados a estos productos génicos si la evidencia

adecuada está disponible. "Contributes_to" es una propiedad utilizada en la anotación GO para indicar que la entidad anotada no realiza la función molecular de forma aislada, sino como miembro de un complejo. Nombres de proteínas: (DROSHA) Ribonucleasa 3; (DGCR8) Subunidad del complejo de microprocesador DGCR8; (XPO-5) Exportin-5; (RAN - GTP) GTPasa cargada con GTP; (DICER1) Endoribonucleasa Dicer; (TARBP2) subunidad compleja de carga RISC TARBP2; (AGO) Argonaute. (Huntley et al., 2016)

Propiedades termodinámicas

La accesibilidad del lugar de unión del gen diana, o *target site*, es donde el miARN se une impidiendo la transcripción. Se ha demostrado que las propiedades termodinámicas relacionadas con esta unión son un factor a tener en cuenta en la predicción de genes diana de miARN, de hecho, se usa en herramientas de predicción de *targets* como PITA[15] (Kertesz et al., 2007)

- Energía de apareamiento del dúplex miRNA:mRNA (o energia libre o energia libre de Gibbs)
    - La energía libre (o energía libre de Gibbs) se puede usar como una medida de la estabilidad a nivel biológico. Si se pronostica que la unión de un miARN a un ARNm diana es estable, se considera más probable que sea un verdadero miARN target. Dada la dificultad para medir la energía libre directamente, generalmente se considera como tal el intercambio de energía libre durante una reacción de unión (G). Dado que las reacciones con un G negativo tienen menos energía disponible para reaccionar en el futuro, resultan en sistemas con mayor estabilidad. Al predecir cómo el miRNA y su posible target hibridan, se pueden inferir regiones de alta y baja energía libre y el G global puede ser usado como un indicador de cuán fuertemente están enlazados.
    - A pesar de esto, este factor no se toma en consideración en muchos algoritmos predictivos. se mide en DDG en PITA: como más negativo sea DDG más favorable será la energia del pliegue a formar el dúplex miRNA:mRNA.(Shirdel et al., 2011) (Peterson et al., 2014)
- Mínima energía libre (o *Minimum free energy*) (Yue et al., 2009)
    - *Minimum free energy* del duplex miRNA:mRNA
    - *Minimum free energy* de mRNA libre

Propiedades estructurales del ácido nucleico

Como los miARN son un tipo de ácido ribonucleico, las características comunes en todos los ARN se encuentran también en los miARN. Estas propiedades hacen referencia a las estructuras primarias y secundarias de los ácidos nucleicos, que en muchos casos pueden tener una influencia en procesos de biogénesis de miARN o en mecanismos de regulación génica de los genes diana:

[15] https://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html (2017)

- Secuencia del pri-miRNA
- Estructura del pliegue de pri-miRNA
- Secuencia de pre-miRNA
    - Pueden existir distintos polimorfismos en un pre-miRNAs (Iwai et al., 2005)
- Estructura de la horquilla del pre-miRNA
    - La estructura en forma de horquilla del pre-miRNA puede ser diferente a la calculada computacionalmente y tiene una especial relevancia en cuanto al proceso de biogénesis. (Castilla-Llorente et al., 2013) (Krol et al., 2004)
- Secuencia miRNA
- Secuencia de la *seed region*
    - Dado que la seed region es lo que principalmente determina la afinidad entre el miARN y el target site, es imprescindible poder representar su secuencia en la ontología.
- Contenido local de AU
    - El contenido local de AU hace referencia a la concentración de nucleótidos A y U en las regiones contiguas a la "seed región" del miRNA. Se ha demostrado que un enriquecimiento en pares de bases A o U en los siguientes 30nt (tanto upstream como downstream) en el lugar de unión del miRNA en 3' UTR tiene a asociarse favorablemente con la represión en la expresión del gen diana (o target). Se toman en consideración en Targetscan (Shirdel et al., 2010) (Peterson et al., 2014).
- Secuencia del miRNA target
    - En general mRNA que formará el lugar de unión del miRNA, pero también puede ser ARN de enzimas reguladoras o DNA.

Propiedades estructurales vinculadas a su papel regulador

Más allá de ser una secuencia de unos 20 ribonucleótidos, los miARN juegan un papel importante en la regulación génica. Este papel regulador viene determinado principalmente por las características concretas que existen en el duplex miARN:mARN que se forma cuando el miARN se une al gen diana. (Bartel D.P., 2009) Estas características son en las que se basan muchos de los algoritmos computacionales de predicción de miRNA targets (ver *Tabla 3*). (Peterson et al., 2014) (Shirdel et al., 2011)

Es especialmente en este aspecto donde miRNAO ha quedado relativamente obsoleta, porque en estos últimos años ha habido un cambio de perspectiva en cuanto a la predicción de miRNA targets. En parte debido a descubrimientos importantes como la complementariedad en 3' más allá de la "seed region" (Broughton et al., 2016)

- Secuencia de mRNA flanqueante al miRNA target binding site (o secuencia local del mRNA) (Lewis et al., 2005)

    - Ocasionalmente se toma en consideración la secuencia alrededor del miRNA *binding site* en 3' UTR. Se puede examinar la accesibilidad del lugar o el contenido de algunos nucleótidos en particular.
- Posición dentro del genoma donde se encuentra el miRNA target
- Posición de miRNA target dentro del gen en concreto (o *position contribution / position effects*)
    - La posición del lugar de miRNA target puede tener contribuir al efecto represivo si cae en el primer cuartil de 3' UTR después del codón stop (+15pb) o en el cuartil final en 3' UTR, cerca de la cola de poli(AAAA). Este efecto posicional es más pronunciado en UTRs largos.
    - Esta región también es importante pues puede provocar lo que se conoce como " ribosomal shadow ". Se toma en consideración en Targetscan: los 15nt después del codón stop en 3' UTR forman un lugar de unión del miRNA que muestra poca capacidad de reprimir la expresión, y se ha postulado que puede ser debido al ribosomal shadow effect. (Shirdel et al., 2011) (Peterson et al., 2014)
- Complementariedad en 3' fuera de la *seed region*
- Complementariedad de bases entre miRNA y mRNA en la "seed region" (o *seed sequence match*)
    - Posición de inicio
    - Posición final
    - Tipos de enlaces
    - Número (y posición) de gaps, bulges y wobbles.
        - Todos los algoritmos de predicción dependen de este criterio. Permitiendo desajustes en los pares de bases y uniones G:U, que han demostrado ser importantes en la vinculación de microRNA, los algoritmos de predicción buscan un alto grado de complementariedad entre el extremo 5' del microRNA y el extremo 3' de la secuencia diana de mRNA. Se presta especial atención a la *seed region* (bases 28 desde el extremo del microARN 5').
- Complementariedad compensatoria en 3'
    - A parte de una fuerte complementariedad en la *seed region*, muchos algoritmos que no están basados en miRanda también requieren uniones nucleotídicas entre miRNA y su target entre las bases 12-17 del final 3' del miRNA (Shirdell et al., 2011)
- Número de lugares de unión de un miRNA y su target (o *multiple binding sites*)
    - Estos lugares de unión múltiple pueden ser de un mismo miRNA o varios, mostrando en este segundo caso cooperatividad entre miRNAs que resulta en la represión de la expresión génica de forma sinérgica. Se mide su abundancia en 3' UTR, hecho que indica la capacidad de regular de forma gradual la expresión de ese gen.

Además, se conoce el efecto de clusters de miRNA (llamada represión sinérgica) pero se ha dejado fuera del alcance del proyecto actual y se considera como parte del trabajo en un futuro.

Feature	Targetscan	RNA22	microRNA.org	mirBase	PITA	picTar	DIANA-microT	miRanda	mirTarget2	targetMiner	SVMicro	RNAhybrid
Conservación	x		x	x	x	x	x	x	x	x	x	
Accesibilidad del target site					x		x		x	x	x	
Contenido local de AU	x											
Multiples lugares de unión	x		x	x	x	x*	x					
***Free energy* del duplex**			x	x	x	x	x	x	x	x	x	x
Examina sequencia de alrededor	x				x							
Considera distintos *seed region*	x		x	x			x					

TABLA 3 Características comunes en distintas herramientas de predicción de miRNA targets (Shirdell et al., 2011) (Peterson et al., 2014)

**Tambien posee esta opción para más de un microRNA.*

Propiedades funcionales vinculadas a su papel regulador

Sin duda alguna, lo que hace más interesantes a los miRNA es su papel en la regulación génica. Sin embargo, no todos los miRNA tienen el mismo efecto ni realizan el silenciamiento con los mismos mecanismos moleculares. (Huntley et al., 2016)

- Efecto regulador a nivel molecular de mRNA (Huntley et al., 2016)
  - Corte del mRNA
  - Deadenilación
- Efecto regulador a nivel de expresión génica (downregulation/upregulation)

3.2.2 Retos y dificultades

La representación de las características de los miRNA supone, en muchas ocasiones, retos y dificultades que se deben tener en cuenta. Más allá de los conceptos a modelar determinados en la sección anterior, algunos factores pueden influir de manera directa o indirecta en su representación. A continuación se indican algunos de ellos:

- Existen muchos targets distintos para un mismo miRNA y muchos miRNA pueden tener un solo target. Es complicado representar todas las relaciones entre elementos reguladores de la expresión génica de la vía de los miRNA porque son muchos y muy variados.
- Un mismo miRNA puede tener un efecto regulador distinto en función del organismo en el que se encuentre o incluso en función del órgano/tejido dentro de ese mismo organismo.
- Existen mecanismos de biogénesis y acción de miRNA diferentes a los de la vía general, como por ejemplo los miRNA extracelulares en biofluidos o los víricos, entre otros recientemente descubiertos. (Cullen et al., 2006) (Ruby et al., 2007) (Sohel et al., 2016)
- Un mismo miRNA puede mostrar polimorfismos (SNP) o variaciones genéticas que pueden conllevar heterogeneïdad a nivel fenotípico del gen diana regulado. (Iwai et al., 2005) (Hu et al., 2008)
- A pesar de que en animales su longitud suele oscilar entre 21-22nt tienen tamaños muy variados y en plantas pueden llegar a medir 60nt. (Mendes et al., 2009)
- Los ID en bases de datos como miRBase pueden variar en función de la versión (Haunsberger et al., 2016)
- La base de datos más grande de miRNA, miRBase, aporta ocasionalmente información incorrecta sobre las estructuras secundarias de miRNA (Saçar et al., 2013)

3.2.3 Extensión

Una vez se ha determinado los conceptos a modelar, es necesario añadir a la ontología las clases y propiedades necesarias. Este capítulo muestra cómo y qué clases y propiedades que han sido añadidas a MIRNAO.

En primer lugar, siguiendo las premisas sobre el diseño de ontologías (Noy et al., 2001), si los conceptos a modelar ya se hallaban presentes en miRNAO. Para ello, se recurrió a Online Lookup Service (OLS)[16] de miRNAO, o directamente se consultó su archivo original *mirnao.owl* en

[16] http://www.ebi.ac.uk/ols/ontologies/mirnao (2017)

Protege. También se consultaron las definiciones de los términos en el archivo original o en Ontobee[17] y en el artículo original de miRNAO (Dritsou et al., 2012)

En caso de no estar representados, se intentó buscar términos en otras ontologías. Los más específicos acerca de miRNA se buscaron en OMIT y NCRO, mientras que otros más generales se buscaron ontologías como Ontology for Biomedical Investigation (OBI) (Peters et al., 2009), Gene Ontology (GO) o Sequence Ontology (SO) (Eilbeck et al., 2005). Los conceptos que se encontraron en dichas ontologías, fueron añadidos usando referencias a dichas ontologías externas.

Finalmente, y como último recurso, en caso que los conceptos que se querían representar no estuvieran presentes en ninguna ontología, se crearon términos (clases o propiedades) nuevos para ellos.

Dado que representar una ontología en un archivo de texto en un formato comprensible para el lector es complicado, se ha optado por mostrar las nuevas clases añadidas mediante OntoGraf (Falconer S., 2010).

En OntoGraf no se muestran los nombres de las propiedades, sino que cada color representa una propiedad distinta (ver Figura 6). *has_individual* indica una relación entre una clase y un individuo u objeto que pertenece a dicha clase. *has_subclass* indica una relación donde una clase pertenece a otra (es decir, el objeto es una subclase del sujeto). *has_participant* (subclase de *has_agent*) indica una relación donde una clase participa en otra (*has_participant* puede usarse para representar la participación de una enzima en un proceso).

En OntoGraf, las casillas marcadas con un punto amarillo representan clases dentro de la ontología, mientras que los marcados con un punto lila representan individuos.

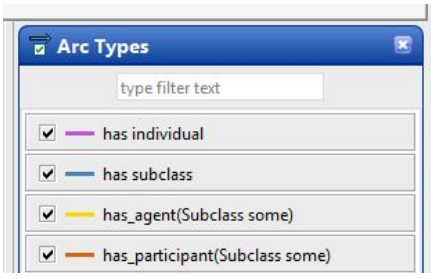


FIGURA 6 El color de las líneas en los gráficos de OntoGraf indica sus propiedades o relacionantes

Identificación / Referencias

No existía ninguna clase para representar el concepto artículo (acerca

[17] http://www.ontobee.org/ (2017)

de miRNA). Se añadieron clases de OMIT, como *PubMed_summary*, y algunas de *Information Artifact Ontology* (Ceusters W., 2012). Además, se añadieron propiedades “has_ID” para poder relacionar cada miRNA (o miRNA target) con el ID con su respectivo artículo

- Literatura relacionada con el miRNA

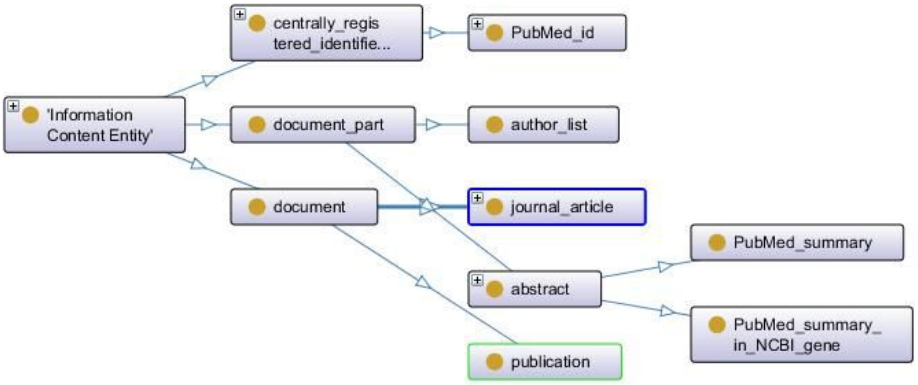


FIGURA 7. Literatura relacionada con el miRNA. Todas las flechas son de color azul, és decir, has_subclass. Ver leyenda en (Figura 6)

En cuanto a las referencias en la literatura, se ha considerado oportuno poder relacionar el microARN con documentos de la literatura científica que hacen referencia a sus propiedades, y dentro del mismo poder modelar conceptos como el abstract o la lista de autores (Figura 7). Todos los términos han sido prestados de otras ontologías, pero ninguno se hallaba previamente en MIRNAO (Information Artifact Ontology y OMIT)

- Información de las BBDD de referencia de donde se extrae el miRNA y el mRNA (ID, Versión de la BBDD)

A medida que los paradigmas científicos evolucionan, las herramientas de estudio también cambian. Es por ello que se debe intentar extender una ontología que tenga clases genéricas, que se puedan mantener en el tiempo, pero al mismo tiempo que puedan ser lo suficientemente concretas para el tipo de información que se quiere representar.

En este caso se han tomado prestadas distintas clases de OMIT que hacen referencia a identificadores, tanto de genes (diana o no) como de miARN, en distintas bases de datos. Además, se ha creado una nueva clase “database_version” que contiene “NCBI_version” (de OMIT) y “miRBase_version” (nuevo término). Esto se debe a que distintas herramientas computacionales trabajan con miRBase como base de datos de referencia, pero no todas trabajan con la última versión (Akhtar et al., 2015)

Si dichas herramientas informáticas quedaran obsoletas, sería necesario realizar un mantenimiento de la ontología para elimninarlas o renovarlas. Si el identificador que se quiere usar no se encuentra en ninguna de las

clases concretas, siempre se puede instanciar la clase "centrally_registered_identifier" directamente (Figura 8).

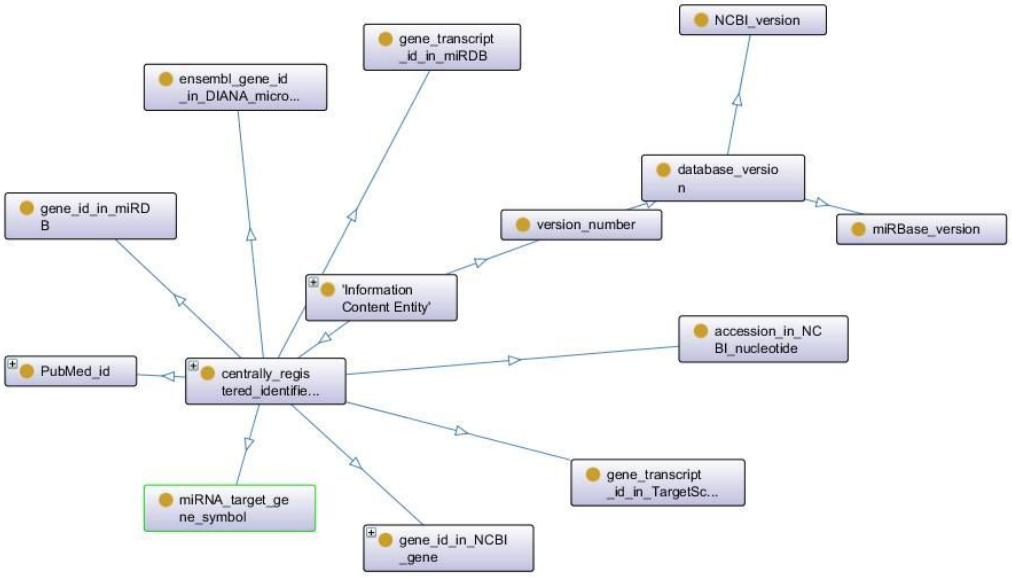


FIGURA 8. Información de las BBDD. Ver leyenda en (Figura 6)

- Tipo de validación empírica (Figura 9)

En cuanto a la validación, se ha dividido la clase prestada de Information Artifact Ontology "information_from_miRNA_target_validation_experiment" en dos subclases: direct_miRNA_target_validation_method y indirect_miRNA_target_validation_method. Esto se debe a que los métodos experimentales de validación de miARN targets se pueden clasificar en estos dos grupos (ambas clases nuevas de MIRNATO). También se han tomado términos prestados de OMIT, como *methods_in_DIANA_TarBase* o *validation_from_miRTarBase*.

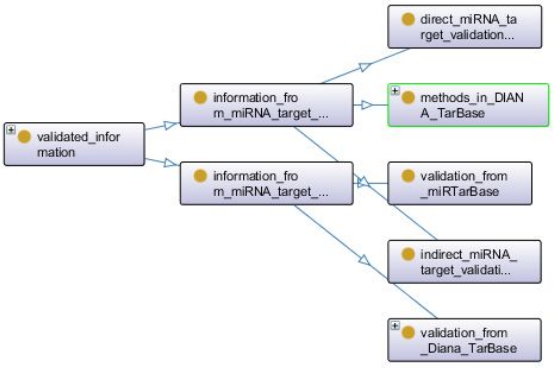


FIGURA 9. Tipo de validación empírica. Ver leyenda en (Figura 6)

- Predicción computacional del miARN y su target (Akhtar et al., 2015)

- Métodos de predicción computacional del miARN y su target

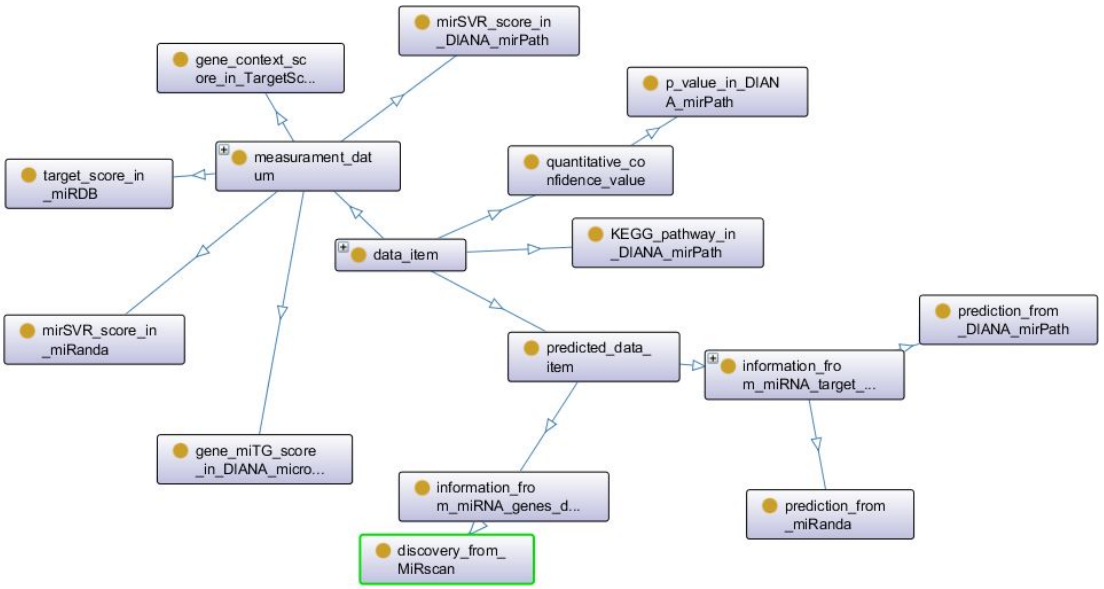


FIGURA 10. Predicción computacional del miARN y su target. Ver leyenda en Figura 6

En este apartado se ha optado por tomar clases prestadas de OMIT en cuanto a las subclases de measurament_datum, que hacen referencia a las puntuaciones de distintas herramientas de predicción de miARN targets. Esto es importante porque no todas las predicciones tienen una base igual de sólida.

Se ha incluido dos de clases generales: una que hace referencia a la predicción de nuevos miARN y otra a la predicción de nuevos miARN targets (information_from_miRNA_genes_discovery_tools y information_from_miRNA_target_prediction_databases respectivamente)

Cada una de estas tiene uno o dos ejemplos de herramientas actualmente populares. (Figura 10)

Propiedades de localización & Clasificación del miRNA y su target
Los miRNA se encuentran en una gran variedad de organismos y tejidos. Es por ello que es importante poder representar conceptos que hacen referencia a dónde se encuentran los miRNA y sus targets.

- Organismo al que pertenece (Figura 11)

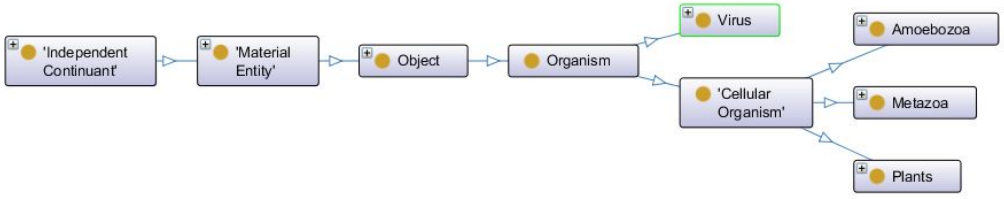


FIGURA 11. Organismo al que pertenece. Ver leyenda en Figura 6

MIRNAO ya posee una larga lista de organismos, términos prestados de NCBI Taxonomy[18]. No se ha añadido ninguno, aunque su revisión será necesaria a medida que se vaya descubriendo nuevos organismos poseedores de miARN.

- Tejido al que pertenece

En este aspecto, tan solo NCRO tenía algunos tejidos humanos en su ontología. En MIRNATO se ha decidido crear el término prestado "tissues,_cell_types_and_enzyme_sources" dentro de la subclase "material_anatomy_entity". Este término está vinculado a la ontología BRENDA[19], que se encarga de la representación de tejidos y enzimas.

- Conservación entre especies

En este aspecto simplemente se ha creado el relacionante "is_conserved_in", dado que la lista de especies ya existe. En este caso, el formato sujeto-relacionante-objeto seria "miRNA, is_conserved_in, organism"

- La familia del miRNA

OMIT y NCRO poseían toda la lista de familias de miRNA humanos como clases. Como MIRNAO no tenía la clase "miRNA gene family" se ha añadido, pero no todas las familias de miARN han sido añadidas. Para poder modelar este concepto, se debe crear la familia de miRNA como un individuo que instancia la clase "miRNA gene family".

- Región del genoma donde se encuentra

Se ha añadido a MIRNAO la clase "chromosome_part" como subclase de "Biological Region", que puede ser instanciada por individuos que sean la región concreta donde se encuentra un gen, un miRNA, un transcrito, etc. Esto también puede servir para determinar la posición dentro del genoma donde se encuentra el miRNA target.

- Gen del que forma parte el miRNA

Se tomó prestado de OMIT el término miRNA_gene

- Gen diana del miRNA

Se tomó prestado de NCRO el término miRNA_target_gene

- Tipos de *seed region*

Ninguna ontología tenía términos para modelar estos conceptos, así que se han añadido nuevos términos para describir seed regions canónicos y no canónicos. (Figura 11.2)

[18] https://www.ncbi.nlm.nih.gov/taxonomy (2017)
[19] http://www.ebi.ac.uk/ols/ontologies/bto (2017)

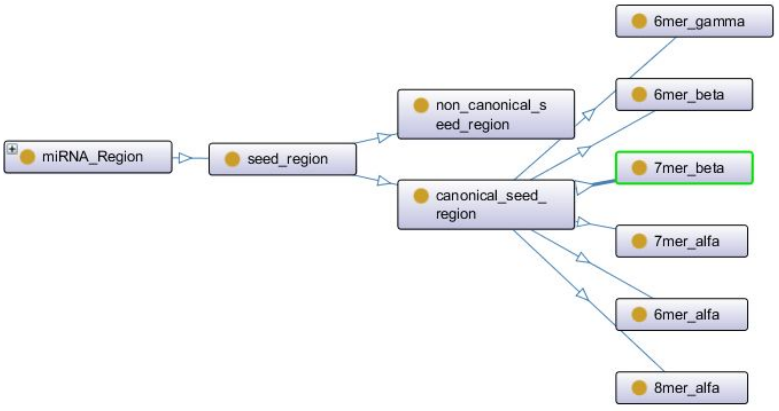


FIGURA 11.2 Tipos de seed region. Ver leyenda en Figura 6

- IsomiRs, isoformas de miARN

Como ninguna ontología tenía clases para representar estos conceptos, se han añadido nuevos para representar las extensiones o deleciones que los miRNa pueden sufrir en 3' o 5', además de polimorfismos. Se han considerado bajo el término prestado de OMIT "RNA Isoforms". A pesar de que no se ha añadido al gráfico, también se han creado la clase "base_pairing" con las subclases "base_pairing_with_miRNA" y "base_pairing_with_mRNA" (Figura 12)

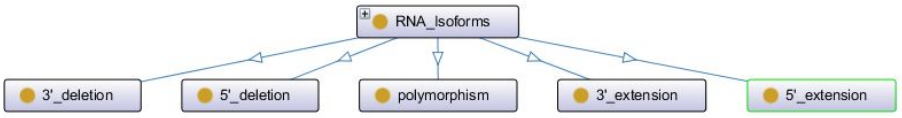


FIGURA 12. IsomiRs. Ver leyenda en Figura 6

Actividad enzimática desde pri-miRNA hasta llegar a pre-miRNA

En este apartado se han cogido principalmente los términos de GO que se recomendaba tomar en el artículo (Huntley et al., 2016). Además, se han tomado muchos términos prestados de NCRO y OMIt que no estaban presentes en MIRNAO. Los siguientes aspectos tiene un gran número de términos y su uso dependerá mucho de qué concepto se quiera modelar.

A continuación se muestra en esquemas generados con OntoGraf las clases añadidas a MIRNAO siguiendo los pasos del proceso canónico de biogénesis de miRNA en animales y plantas:

- Transcripción

Términos que hacen referencia a la actividad RNA polimerasa II y III, a la unión de pri-miRNA, o a factores de transcripción han sido añadidos, todos prestados de GO u OMIT. El objetivo es poder modelar distintos tipos de transcripciones de miRNA, o distintos pequeños elementos y

actividades dentro de la misma. En este sentido es importante representar conceptos como la unión al DNA, transcripción de pri-miRNA, unión a proteínas, etc. (Figura 13)

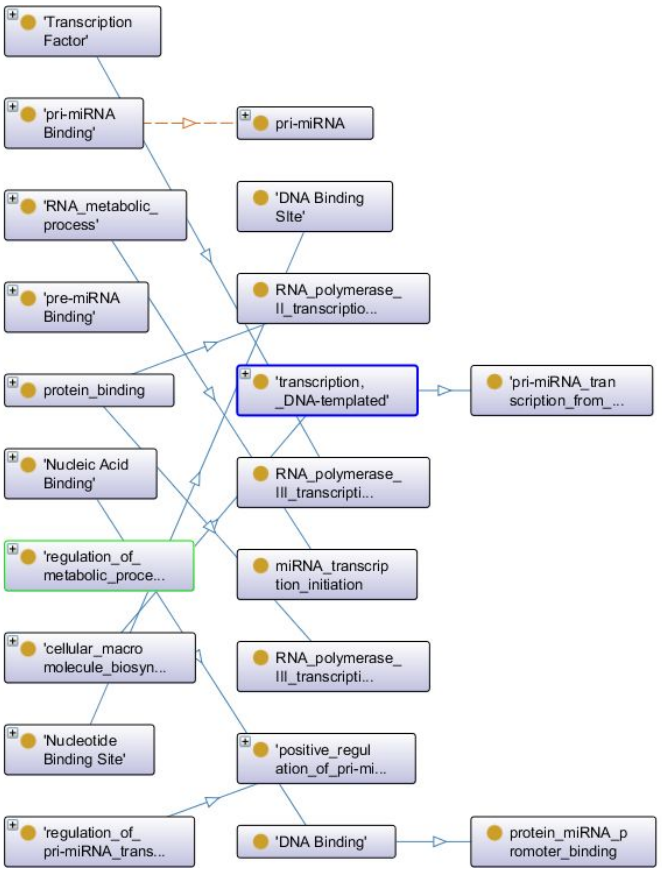


FIGURA 13. Transcripción. Ver leyenda en Figura 6

- Escisión (pri-miRNA a pre-miRNA)

En este apartado se han añadido clases relacionadas con el complejo microprocesador, Drosha y Pasha en animales; y el Dicer en plantas que contiene distintas proteínas y enzimas. Es importante representar sus actividades enzimáticas (ribonucleasa III), procesamiento o modificación del pri-miRNA, 2'-O-metilación, etc.(Figura 14)

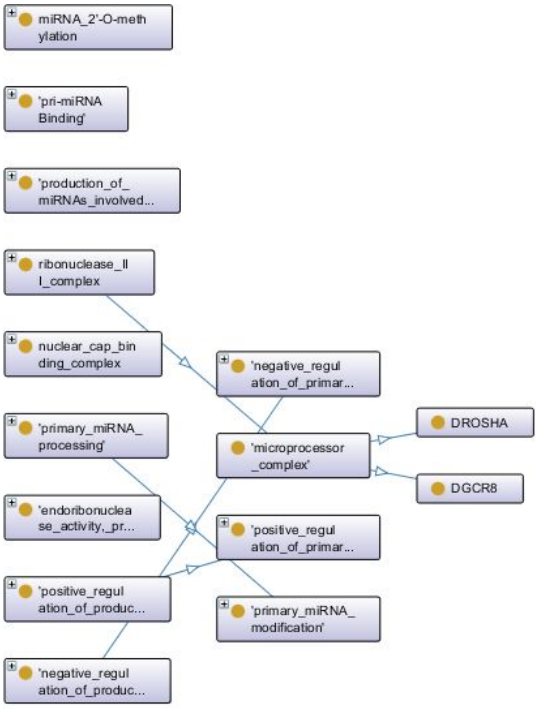


FIGURA 14 Escisión (de pri-miRNA a pre-miRNA). Ver leyenda en Figura 6

- Transporte del núcleo al citoplasma

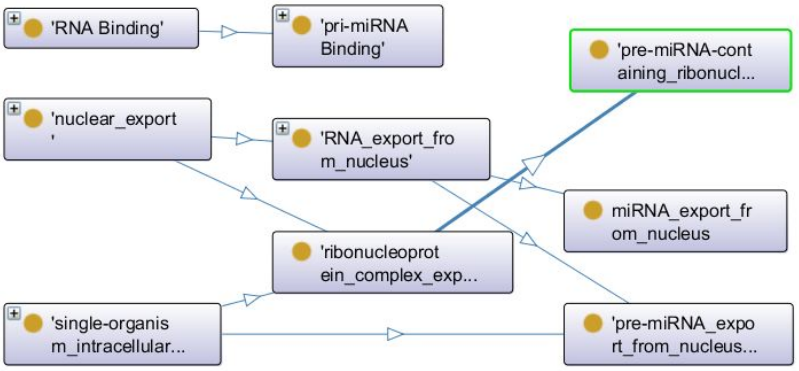


FIGURA 15 Transporte del núcleo al citoplasma. Ver leyenda en Figura 6

En este paso hay que representar conceptos que esten relacionados con la exportación del pre-miRNA del núcleo al citoplasma. En algunos casos el miRNA maduro ya ha madurado en el núcleo, o el pre-miRNA se encuentra formando un complejo ribonucleoproteico. Todos estos conceptos han sido prestados de GO y no estaban presentes en MIRNAO. (Figura 15)

- Escisión del loop de pre-miRNA en el medio citoplasmático

En esta parte de la ontología debe representarse la actividad enzimática de DICER1 y TARBP2, o en otras palabras, la actividad ribonucleasa III. Se ha representado en la ontología distintas actividades enzimáticas relacionadas con el procesado de pre-miRNA, así como el complejo proteico en sí mismo. (Figura 16)

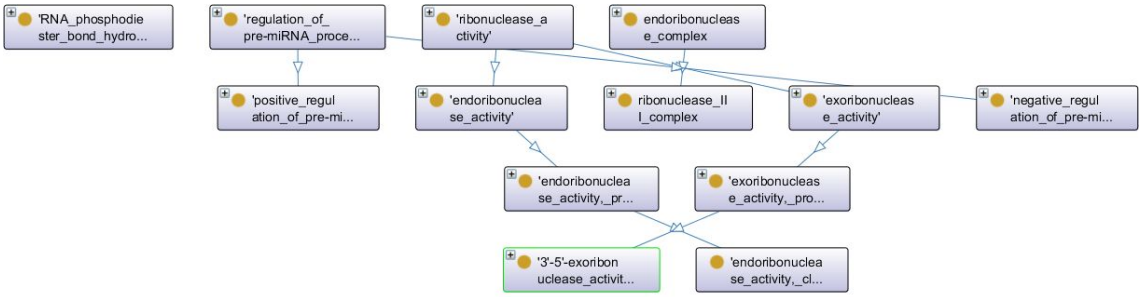


FIGURA 16. Escisión del loop de pre-miRNA en el medio citoplasmático. Ver leyenda en Figura 6

- Formación del complejo RISC
- Efecto regulador

En este apartado se ha representado el desenrollamiento del dúplex miRNA:miRNA que tiene lugar al inicio. También proteinas y actividades enzimáticas sobre la formación del complejo RISC (“RISC_complex” y “dobule_stranded_miRNA_binding”) en colaboración con las proteínas AGO “Argonaute_Proteins”, y finalmente su efecto inhibidor. Ningún término estaba presente anteriormente en MIRNAO, son todos prestados de GO o OMIT. (Figura 17)

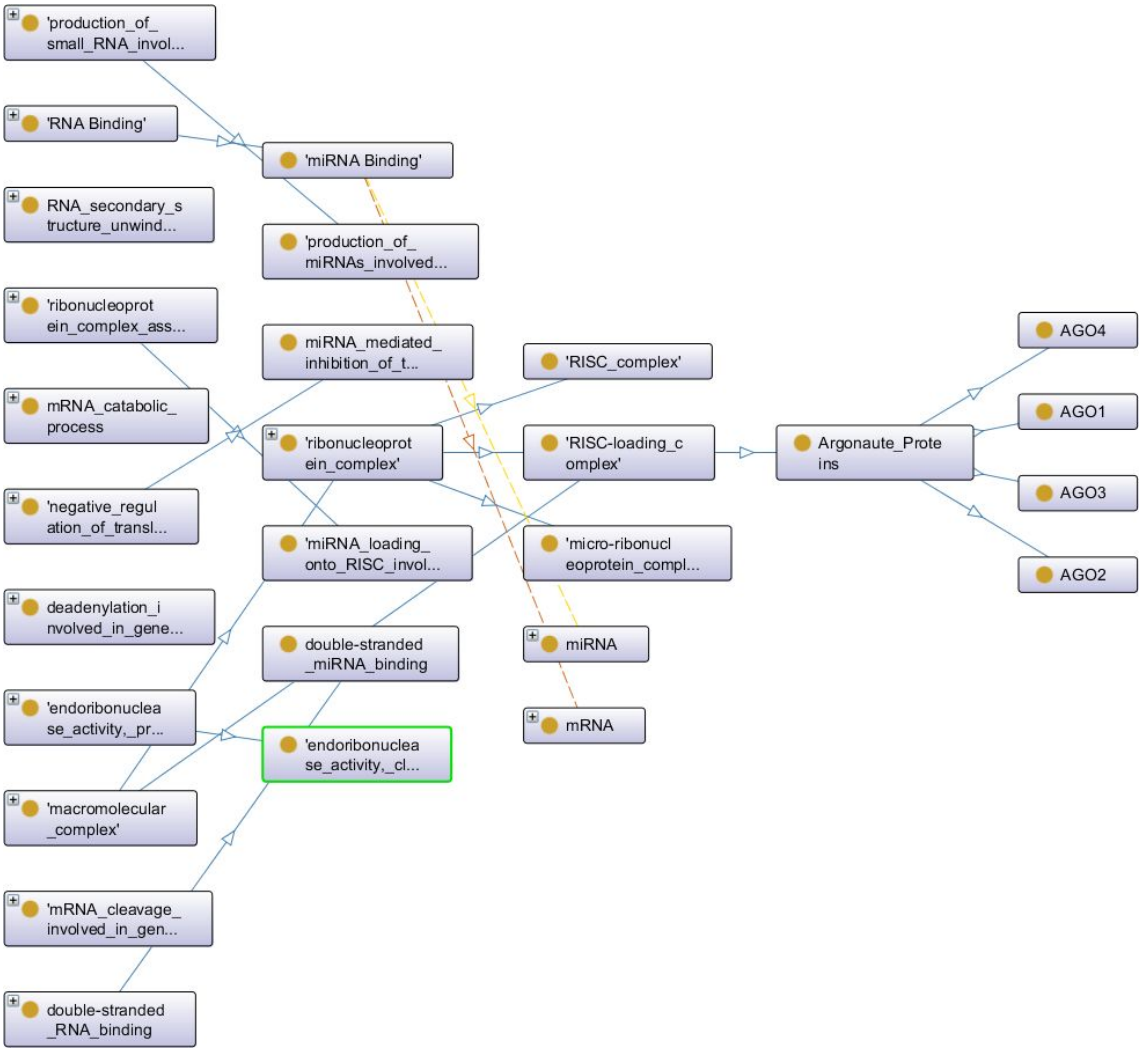


FIGURA 17. Formación de RISC & Efecto regulador. Consultar leyenda en Figura 6

- Elementos que regulan la desaparición de miARN (Figura 18)

Simplemente se ha añadido el término "3'-UTR-mediated mRNA destabilization", prestado de GO (Kai et al., 2010)

Propiedades termodinámicas

En este apartado se han creado nuevas subclases a energía libre de Gibbs, que es un término prestado de Sequence Ontology. Además, también se han creado clases referentes a la mínima energía libre para poder así modelar el concepto de accesibilidad del target.

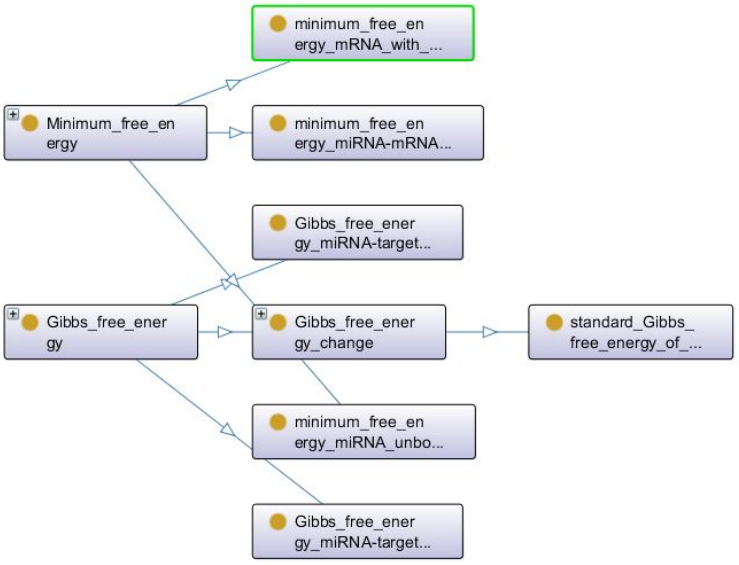


FIGURA 18 Propiedades termodinámicas. Consultar leyenda en Figura 6

Propiedades estructurales del ácido nucleico

En este apartado se han cogido términos que ya existían en MIRNAO para describir la estructura secundaria del RNA, como por ejemplo terminal loop. También existía el término "AU rich element", para poder modelar el concepto de contenido en AU.(Figura 19)

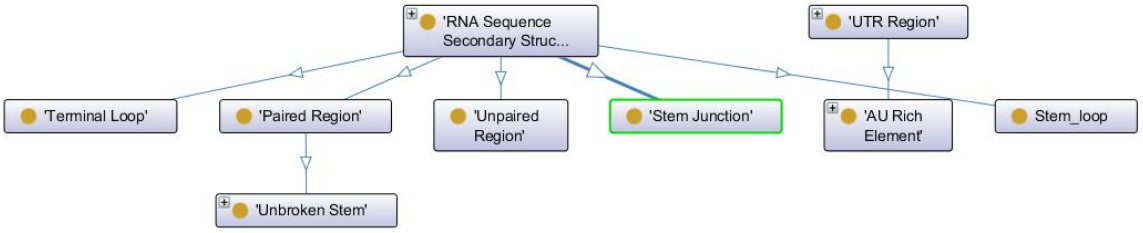


FIGURA 19. Propiedades estructurales del ácido nucleico. Consultar leyenda en Figura 6

Propiedades estructurales vinculadas a su papel regulador

En este apartado se han añadido algunos términos que hacen referencia a las propiedades estructurales de los miRNA, ya sea de su propia secuencia de RNA o de las interacciones con su target.

- Secuencia de mRNA flanqueante al miRNA target binding site (o secuencia local del mRNA) (Lewis et al., 2005) Figura 20

Dentro de la clase "topologically_defined_region" se ha añadido "flanking_region", que tiene subclases para modelar si se encuentra en el extremo 3' ó 5' de la secuencia, o si bien se trata de un transposón.

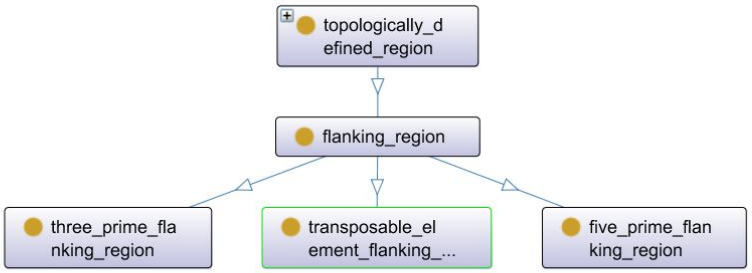


FIGURA 20 Regiones flanquentes. Consultar leyenda en Figura 6

- Posición de miRNA target dentro del gen en concreto (o position contribution / position effects)
- Posición de inicio
- Posición final

En este caso, con las clases “end_position” y “start_position” se pueden modelar conceptos como la posición de inicio y final de un miRNA concreto. También puede ser útil para representar otros modelar conceptos relacionados con la posición de un elemento en una secuencia. Ninguno de estos conceptos se hallaba antes en MIRNAO, han sido todos prestados de SO. (Figura 21)

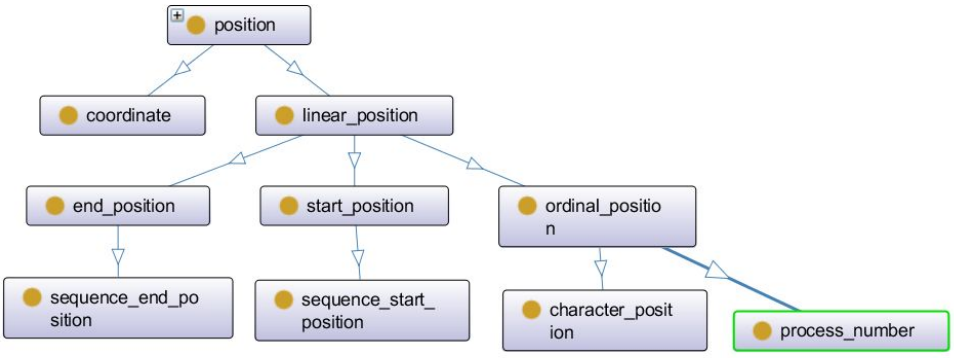


FIGURA 21. Posiciones de inicio y final. Consultar leyenda en Figura 6

- Número (y posición) de gaps, bulges y wobbles.

Las secuencias de distintas fases de la biogénesis y mecanismo de acción de los miRNA pueden tener distintas características en su estructura primária y secundária de alta relevancia como gaps, bulges o wobbles que no han sido representadas en MIRNAO. Es por ello que estos tres conceptos se han añadido a MIRNATO. (Figura 22)

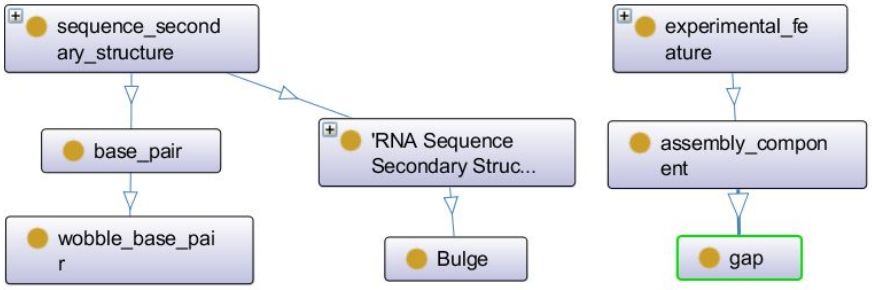


FIGURA 22 Gaps, bulges y wobbles. Consultar leyenda en Figura 6

- Complementariedad compensatoria en 3’

En la clase “sequence_feature” se ha añadido “gene_three_prime_UTR_sequence_in_miRDB”, término prestado de OMIT. A pesar de que es arriesgado poner “in_miRDB” porque es una herramienta que puede estar sujeta a cambios, toma prioridad no crear duplicidad de términos y considerar otros ya diseñados. Simplemente debería mantenerse bajo revisión en futuras versiones de MIRNATO.

Propiedades / relaciones

En toda ontología son necesarias las propiedades para poder relacionar clases e individuos. En esta ontología se han añadido varias para poder relacionar las nuevas clases añadidas. Aquí se muestran algunas propiedades relevantes añadidas a MIRNATO:

- expressed_in_organism → Para poder expresar en qué organismo se expresa un miRNA (o su target)
- expressed_in_tissue → Para poder expresar en qué tejido se expresa un miRNA (o su target)
- has_conserved_miRNA_in → Para poder expresar en qué distintas especies se conserva un miRNA
- has_ID → para relacionar ID de miRNA en bases de datos
- has_location → Para expresar la región del genoma
- has_predicted_target → relacionar un miRNA con la predicción de un target
- has_sequence → expresar al secuencia de un gen, miRNA, seed region, etc
- has_structure → para relacionar loops, wobbles u otras estructuras de ácidos nucleicos
- is_classified_into_gene_family_group → classificar miRNAs en su familia correspondiente
- is_validation_method_of → expresar los métodos con los que se validó un miRNA
- miRNA_target_assumption_based_on → expresar la herramienta computacional de la predicción del target de un miRNA

4. Plataforma de consulta

Si en la introducción se remarcaban las diferencias entre bases de datos y ontologías, en este capítulo se hará especialmente palpable. Este capítulo trata de una plataforma que permite realizar consultas SPARQL a MIRNATO, consultas que no se podrían realizar en una base de datos como miRBase o miRDB. La puesta en práctica de la ontología es fundamental para comprobar si realmente su uso puede ser útil o no.

En el proyecto se ha usado un servidor Fuseki Jena Apache[20] que permite realizar consultas mediante SPARQL de manera local. Este servidor es el motor de la plataforma de consulta. El único subojetivo que no se cumplió fue el que implicaba el desarrollo de una página web, que en este caso corresponde a la Interfaz de Usuario. Los mecanismos lógicos que permiten realizar consultas son los que vienen por defecto en el servidor Fuseki.

En este apartado el objetivo principal consiste en conseguir crear un servidor local que tenga implementado una plataforma que nos permita realizar consultas a nuestra ontología, MIRNATO, o bases de datos basadas en la misma.

El Protocolo SPARQL y RDF Query Language (SPARQL) es un lenguaje de consulta y protocolo para RDF[21]. RDF es un modelo estándar para el intercambio de datos en la Web. Formatos como OWL se basan en RDF y proporcionan un lenguaje útil para definir ontologías estructuradas basadas en Web que permiten una mayor integración e interoperabilidad de datos entre comunidades descriptivas[22]. Por lo tanto, el servidor Fuseki Jena Apache nos permitirá realizar consultas mediante el protocolo SPARQL que es compatible con con *mirnato.owl*

Para llevar a cabo esta parte del proyecto, primero se accedió a la web de Fuseki Jena Apache y se descargaron los paquetes necesarios de alguno de sus repositorios (y se extrajo el contenido del archivo comprimido)[23].(Figura 23)

[20] https://jena.apache.org/documentation/fuseki2/index.html (2017)
[21] https://www.w3.org/TR/sparql11-protocol/ (2017)
[22] https://www.w3.org/RDF/ (2017)
[23] https://jena.apache.org/download/ (2017)

bin	02/05/2017 13:44	Carpeta de archivos	
webapp	02/05/2017 13:38	Carpeta de archivos	
fuseki	02/05/2017 13:38	Archivo	13 KB
fuseki.war	02/05/2017 13:43	Archivo WAR	22.491 KB
fuseki-server	02/05/2017 13:38	Archivo	3 KB
fuseki-server.bat	02/05/2017 13:38	Archivo por lotes ...	2 KB
fuseki-server.jar	02/05/2017 13:43	Executable Jar File	24.923 KB
LICENSE	02/05/2017 13:38	Archivo	30 KB
NOTICE	02/05/2017 13:38	Archivo	10 KB
README	02/05/2017 13:38	Archivo	3 KB

FIGURA 23 Archivos dentro de apache-jena-fuseki-2.6.0.zip

A continuación, se accedió a la línea de comandos (o terminal) se accedió al directorio adecuado y simplemente ejecutamos Fuseki con el comando (usando los comandos de la carpeta System32) (ver Figura 24):

```
java -jar fuseki-server.jar --update
```

```
C:\Windows\system32\cmd.exe
Microsoft Windows [Versión 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Reservados todos los derechos.

C:\Users\Sergi>cd ..

C:\Users>cd ..

C:\>cd FESUKI

C:\FESUKI>dir
 El volumen de la unidad C no tiene etiqueta.
 El número de serie del volumen es: 98CE-C254

 Directorio de C:\FESUKI

09/05/2017  11:55    <DIR>          .
09/05/2017  11:55    <DIR>          ..
02/05/2017  13:44    <DIR>          bin
02/05/2017  13:38            12.342 fuseki
02/05/2017  13:38             2.276 fuseki-server
02/05/2017  13:38             1.264 fuseki-server.bat
02/05/2017  13:43        25.520.456 fuseki-server.jar
02/05/2017  13:43        23.030.378 fuseki.war
02/05/2017  13:38            30.358 LICENSE
02/05/2017  13:38             9.534 NOTICE
02/05/2017  13:38             2.386 README
09/05/2017  11:55    <DIR>          run
02/05/2017  13:38    <DIR>          webapp
               8 archivos     48.608.994 bytes
               5 dirs  17.575.763.968 bytes libres

C:\FESUKI>java -jar fuseki-server.jar --update
[2017-05-23 02:37:13] Server     INFO  Fuseki 2.6.0
[2017-05-23 02:37:13] Config     INFO  FUSEKI_HOME=C:\FESUKI\.
[2017-05-23 02:37:13] Config     INFO  FUSEKI_BASE=C:\FESUKI\run
[2017-05-23 02:37:13] Servlet    INFO  Initializing Shiro environment
[2017-05-23 02:37:13] Config     INFO  Shiro file: file://C:\FESUKI\run\shiro.in
i
```

FIGURA 24 Inicio del servidor Fuseki Jena Apache

Para ejecutarlo, se introdujo en el navegador la dirección http://localhost:3030 para posteriormente cargar *mirnato.owl.*

Se cargó la ontología en la pestaña superior izquierda “manage datasets”. Se siguieron los pasos de carga correctamente y apareció una interfaz gráfica como la de la Figura 25.

Es importante recordar que Jena Apache no funciona bien con formatos .owl, y se debe exportar la ontología desde Protégé como *RDF/XML Syntax*.

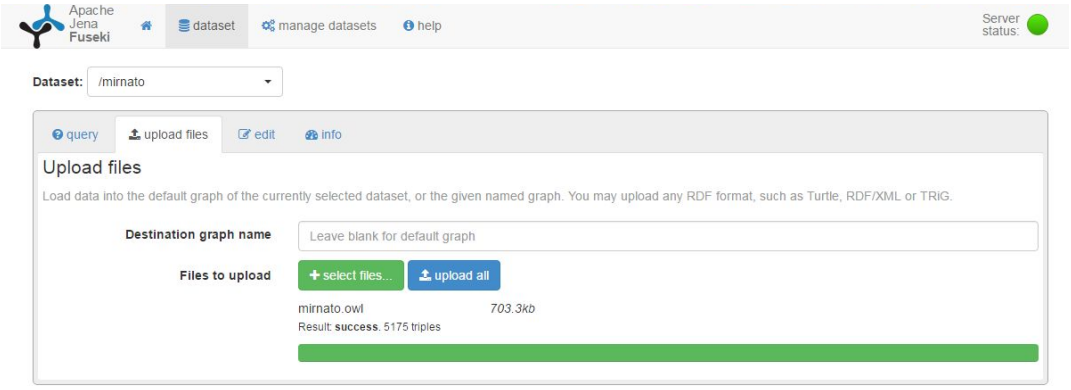


FIGURA 25 mirnato.owl ha sido cargado con éxito.

En este momento al menú de inicio se seleccionó la opción "*query*" en MIRNATO. En el panel principal es posible realizar consultas con el protocolo SPARQL (Figura 26)

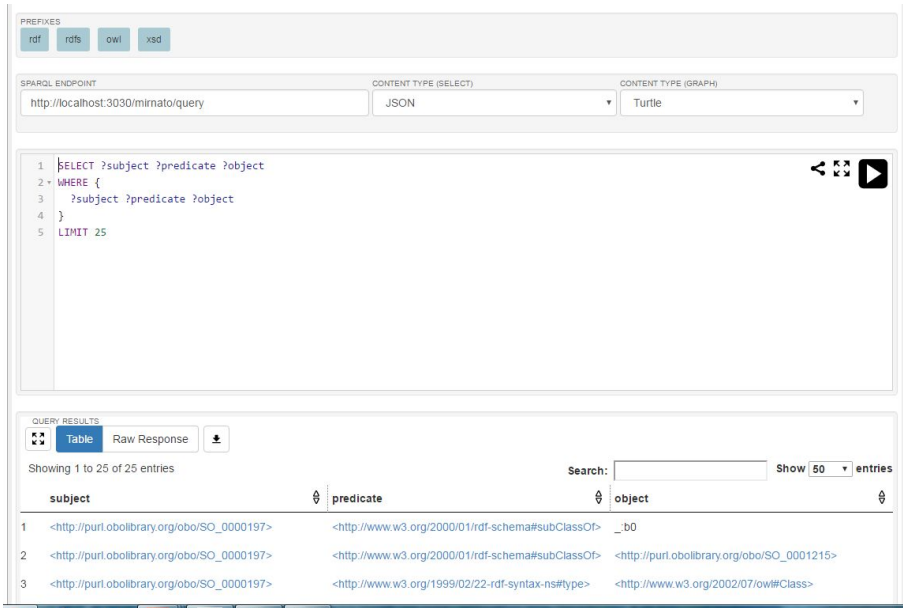


FIGURA 26. Ventana para realizar consultas SPARQL

4.1. Almacenaje de miRNA

En este capítulo se seleccionó un miRNA de ejemplo (en nuestro caso hsa-let-7a-5p) y se intentó obtener tanta información como fue posible. Uno de los objetivos del proyecto es obtener una plataforma que permita extraer "meta-información", es decir, extraer información de consultas concretas de un gran *pool* de información gracias a que se trata de un texto semántico y estructurado.

Una ontología puede contener Individuos que instancian una clase concreta. Esto significa que es posible almacenar información de uno o varios casos concretos de miRNA.

En este paso se almacenó en forma de individuos que instancian clases (donde las clases representan los conceptos generales que se pretende modelar) la información que se obtuvo tenida en el punto anterior acerca del miRNA hsa-let-7a-5p.

Para cargar un nuevo archivo mirnato.owl, con toda la información de interés acerca de algunos miRNA, se subió de nuevo al servidor Jena Apache. Simplemente pulsando en “Update” podemos actualizar estos datos.

En este caso, es posible obtener grandes cantidades de información si se emplea una buena semántica. Este tipo de plataforma permite saber en qué posiciones un miRNA tiene gaps, en qué región del genoma se encuentra o en qué organismos está conservado. No sólo tratándose de un miRNA, sinó que también se puede saber cuántos regulan un gen concreto, o realizar consultas múltiples para luego realizar análisis comparativos.

Se realiza una consulta simple, como por ejemplo a qué genes diana regula hsa-let-7a-5p. Se ejecuta el siguiente código con la opción “Content type = XML” (Figura 27)

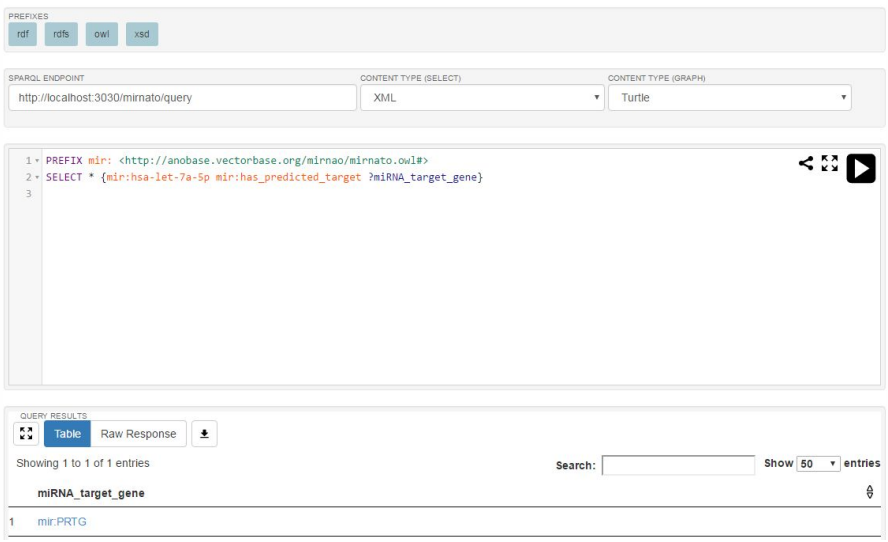


FIGURA 27. Fesuki Jena Apache devuelve información de la consulta “`SELECT * {mir:hsa-let-7a-5p mir:has_predicted_target ?miRNA_target_gene}`

5. Resultados

Los resultados obtenidos son mirnato.owl y el servidor Fuseki Jena Apache. MIRNATO contiene casi el doble número de clases que su ontología original, y casi el triple de propiedades (varias de ellas inversas respecto otras) respecto su ontología original, miRNAO. En cuanto a propiedades supera incluso a NCRO, que tiene proximadamente el triple de clases que MIRNATO. En cuanto a las subclases, MIRNATO tiene casi el doble que miRNAO, es decir, es mucho más específica. Todo esto indica que ha habido un aumento significativo de la capacidad de representación de características y modelaje de conceptos sobre miRNA y sus targets. (ver Tabla 5)

La gran mayoría de las nuevas clases y propiedades son términos adaptados de otras ontologías ya existentes. A parte de estos términos, en MIRNATO se han creado 15 nuevos términos que no existían anteriormente en ninguna otra ontología.

MIRNATO Todavía está lejos de poder competir en número de clases y axiomas con las ontologías del grupo de Huang (NCRO y OMIT). Esta diferencia se debe, entre otras razones, a que NCRO y OMIT tratan a cada miRNA (humano) como una clase, en lugar de como un individuo que instancia una clase (como es el caso de MIRNATO). También ocurre lo mismo con las familias de miRNA. Debido al gran número de miRNA existentes, el número de clases en NCRO y OMIT es mucho mayor que en MIRNATO. En este proyecto se ha decidido aceptar la estructura pre-establecida en miRNAO, donde se considera a los miRNA como individuos.

TABLA 5. Comparación de MIRNATO respecto a sus ontologías antecesoras. Los ontologías están formadas por conjuntos de axiomas, y estos son afirmaciones que determinan lo que es cierto en el dominio (ya sea sobre una clase o culaquier otra característica) (Motik et al., 2009). Para entender el resto de elementos de la tabla, consultar Apartado 2.3.

	MIRNAO	NCRO	MIRNATO
Axiom	3870	17387	4926
Class count	676	3078	1072
Objectproperty count	13	27	**36**
SubClassOf	762	5394	**1365**
InverseObjectProperties	1	2	**9**
Object PropertyRange	1	11	**1**

A continuación, se muestra en la Figura 28 un OntoGraf del miRNA de ejemplo hsa-let-7a-5p (recordemos que los recuadros con un punto amarillo simbolizan clases y los lilas individuos). En ella podemos observar como la MIRNATO funciona correctamente.

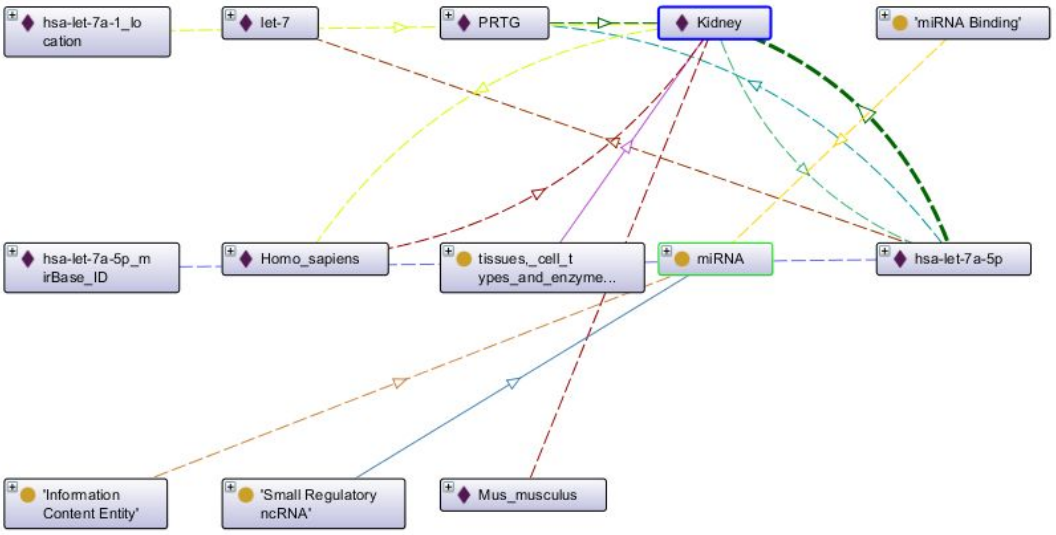


FIGURA 28 Gráfico con información acerca del miRNA de ejemplo hsa-let-7a-5p. Consultar leyenda de las propiedades en Figura 28.2

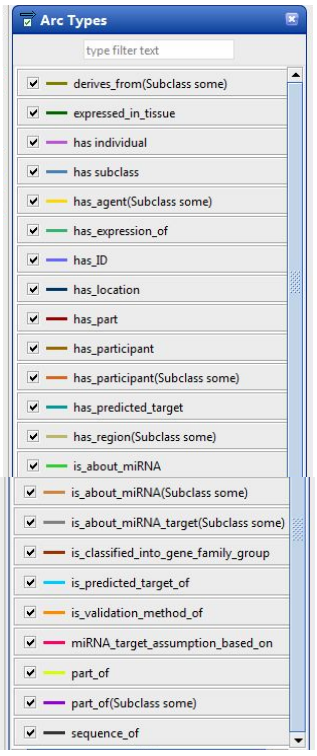


FIGURA 28.2 Leyenda de la figura 28

En este ejemplo se han modelado conceptos como la familia de miRNA, las especies y órganos en los que se expresa o la localización en la que se encuentra. La información se muestra de forma correcta y las relaciones (o propiedades) son lo suficientemente específicas.

Como se ha observado en la Figura 27 del apartado anterior, La plataforma funciona correctamente para realizar consultas que permitan estudiar los targets de un miRNA concreto, u cualquier otro tipo de meta-información que se desee obtener.

Por lo tanto aceptamos los resultados como válidos. Se ha logrado extender una ontología de miRNA (y sus targets) que contiene más clases y más relaciones que su antecesora, aumentando así su capacidad de representación. Además, su puesta a prueba con almacenaje de información de un miRNA de ejemplo y consultas SPARQL en una plataforma Fuseki Jena Apache ha dado resultados positivos.

6. Conclusiones

En este trabajo se ha realizado la extensión de una ontología de miRNA ya existente llamada miRNAO (Dritsou et al., 2012) añadiendo clases y propiedades basadas en fuentes bibliográficas o en otras ontologías a miRNAO (como OMIT o NCRO) que han aumentado su capacidad de representación y modelaje de conceptos acerca de los miRNA y sus targets en comparación a su versión original. La nueva ontología ha sido llamada MIRNATO. En segundo lugar, también se ha desarrollado una plataforma de consulta basada en un servidor local Fuseki Jena Apache que permite realizar consultas SPARQL a la ontología desarrollada, otorgándole así una mayor utilidad más allá de ser una herramienta que facilita el intercambio de información y define una terminología común.

Estas partes del trabajo responden a los dos objetivos principales: "*Extender una ontología para la representación de microRNAs*" y "*Implementar una plataforma y una base de datos para gestionar la información en la ontología creada mediante sparQL*", ambos cumplidos en su totalidad.

Más concretamente, la metodología seguida se basó en los subobjetivos "*identificar las limitaciones existentes en la representación de microRNAs"* y la "*Identificación de las características que serán necesarias dentro de una ontología de miRNAS*". Estos subojetivos se cumplieron gracias a una búsqueda bibliográfica (con especial énfasis en las características más comunes en bases de datos y herramientas de computacionales) y el contraste de información con las ontologías existentes relacionadas con los miRNA (miRNAO, NCRO y OMIT).

Los subobjetivos "*Diseñar una plataforma de consulta que permita el acceso a la información de la ontología*" y "*Compilar toda la información sobre miRNA targets relevante para la ontología en una base de datos*" fueron también cumplidos. En el primer caso se instaló y ejecutó un servidor Fuseki Jena Apache, que permite realizar consultas SPARQL de manera local. A pesar de que no se realizó un código web para integrar una interfaz de usuario, los mecanismos lógicos de consulta vienen por defecto en el servidor. En el segundo subobjetivo, se compiló gran parte de la información del miRNA de ejemplo hsa-let-7a-5p. A pesar de que se compiló menos informació de la esperada debido a falta de tiempo, sí fue suficiente para probar que la ontología funciona correctamente y es útil para obtener meta-información.

Como autor del trabajo, bajo mi punto de vista personal pienso que los resultados han sido satisfactorios y, si se siguieran las líneas de trabajo futuro (ver Apartado 5.3) MIRNATO podría llegar a ser de gran utilidad para suplir la falta de una ontología que cubra de manera satisfactoria todas las características de los miRNA y sus targets (mencionada en la

Justificación del trabajo).

Como se ha dicho anteriormente, el desarrollo de una ontología puede conllevar muchos meses o incluso años. MIRNATO es un paso importante en la dirección correcta, y si ésta siguiera bajo mantenimiento junto con las líneas de trabajo futuras podría llegar a tener un gran impacto en el ámbito de la investigación de miRNA y sus targets.

La continuación del desarrollo de MIRNATO facilitaría el intercambio de información determinando una terminología común para los miRNA, pues miRNAO está relativamente obsoleta y, NCRO y OMIT no cubren de manera satisfactoria todas las características de los miRNA.

6.1 Lecciones aprendidas

Las ontologías son necesarias en el ámbito de la investigación biológica (y especialmente la biomédica) porque aportan un lenguaje estructurado que permite el intercambio de información. Sin embargo, su diseño no siempre es sencillo sino que requiere un gran dominio del conocimiento del área que se desea cubrir.

Dicho diseño consta de distintos pasos y tiene unos principios que deben cumplirse en la medida de lo posible, como evitar usar conceptos abstractos en lugar términos demasiado concretos o evitar la duplicidad de clases tomando en consideración ontologías ya creadas. Además, es importante es añadir propiedades para poder determinar los relacionantes existentes entre las clases / conceptos, a ser posible indicando siempre las cualidades de la propiedad (reflexiva, funcional, simétrica, etc)

Una vez finalizado el diseño, debe tenerse en cuenta que el mantenimiento de una ontología es importante para que no quede obsoleta y pueda seguir siendo útil para compartir información. En el caso de este proyecto, la información acerca de los miRNA está en constante expansión, pues cada día se publican muchos artículos relacionados.

Por otra parte, existen múltiples herramientas computacionales que usan distintas características de los miRNA para predecir sus targets. También existen múltiples bases de datos de miRNA y genes diana validados experimentalmente.

Tanto las ontologías de miRNA como las herramientas computacinoales juegan un papel clave en el estudio de miRNA. Su interés científico se debe a que los miRNA tienen un importante papel en la regulación génica, y en un futuro aparecerán nuevas (y mejores) herramientas con nuevos algoritmos de predicción de targets y cada vez será más necesaria una ontología sólida.

Finalmente, la plataforma de consulta permite dar un uso práctico y útil a la ontología, más allá de ser simplemente una declaración de conocimiento con un lenguaje semántico y estructurado. Permite almacenar información y luego llegar a ciertas conclusiones que no se pueden obtener con BBDD que simplemente almacena ficheros de texto planos.

6.2 Reflexión & Análisis crítico

Se han cumplido todos los objetivos principales y todos los subobjetivos a excepción de uno.

Sobre los dos objetivos principales, se ha realizado con éxito la extensión de una ontología (MIRNAO) con clases y propiedades que antes no tenía y permiten modelar de forma mucho más exacta y completa los conceptos acerca de miRNA y sus targets. También se ha logrado usar una plataforma que permite realizar consultas a MIRNATO, además de permitir añadir información acerca de casos concretos de miRNA para luego poder realizar consultas y extraer conclusiones que no se pueden extraer de una simple base de datos.

La metodología usada ha sido adecuada. La forma para saber qué conceptos se pueden modelar en una ontología y qué limitaciones tiene fue seleccionar un miRNA de ejemplo e intentar modelar toda su información disponible. No toda la información del miRNA de ejemplo fue representada con individuos en la ontología, pero sus conceptos sí fueron añadidos con propiedades y clases (algunas prestadas de otras ontologías existentes, otras nuevas fueron creadas) después de contrastarlos con la bibliografía.

El segundo objetivo principal era lograr diseñar una plataforma que permitiera realizar consultas SPARQL online a la ontología. Se ha cumplido el objetivo empleando un servidor Fuseki Jena Apache, que permite hacer consultas de manera local. Sólo un subobjetivo quedó sin cumplir, que fue el desarrollo de un código web. Esta parte correspondería a la Interfaz gráfica, dado que es las estructuras lógicas de la plataforma ya vienen incorporadas en Fuseki Apache Jena.

La causa principal ha sido la demora temporal en algunas de las tareas, que ha provocado un efecto dominó teniendo como consecuencia la reducción del tiempo dedicado las tareas finales. Los cambios realizados también en los objetivos principales sí han permitido lograr los objetivos, prácticamente en su totalidad.

Mencionar finalmente que sí hubo una tarea que se añadió durante el transcurso del proyecto: contactar con los autores de OMIT y NCRO para pedir permiso para editarla y modificarla. El permiso fue concedido y la única condición que pusieron fue la de citar su artículo en caso de

que se publicara un artículo.

6.3 Líneas de trabajo futuro

Para un trabajo futuro deberían explorarse opciones que permitan modelar conceptos acerca de miRNA y sus dianas que no se han modelado en esta ontología a pesar de ser importantes, como el concepto “múltiples lugares de unión” (o *multiple binding sites*) que puede tener miRNA dentro de un gen diana. Esto tiene una influencia directa en su afinidad por la diana, y es una característica que algunos algoritmos tienen en cuenta para predecir targets (Tabla 3)

También debería realizarse un mantenimiento periódico de la ontología, para eliminar clases que hayan caído en desuso, añadir sinónimos, añadir propiedades o clases que modelen conceptos de hallazgos recientes relacionados con miRNA y sus dianas, etc

El primer punto de trabajo que debería abordarse al finalizar el trabajo es la implementación de una interfície web que facilite el acceso a la plataforma SPARQL que contiene la ontología. Esto facilitaría el acceso a la información modelada en este TFM y permitiría el cumplimiento del último subobjetivo del proyecto.

7. Bibliografía

1. Hammond, S. M. (2015). An overview of microRNAs. *Advanced drug delivery reviews*, *87*, 3-14.
2. Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome biology*, *12*(4), 221.
3. Wang, X. J., Reyes, J. L., Chua, N. H., & Gaasterland, T. (2004). Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome biology*, *5*(9), R65.
4. Yue, D., Liu, H., & Huang, Y. (2009). Survey of computational algorithms for microRNA target prediction. *Current genomics*, *10*(7), 478-492.
5. Sohel, M. H. (2016). Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges. *Achievements in the Life Sciences*, *10*(2), 175-186.
6. Peterson, S. M., Thompson, J. A., Ufkin, M. L., Sathyanarayana, P., Liaw, L., & Congdon, C. B. (2014). Common features of microRNA target prediction tools. *Frontiers in genetics*, *5*, 23.
7. Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F., & Procopio, A. D. (2016). Bioinformatic tools for microRNA dissection. *Nucleic acids research*, *44*(1), 24-44.
8. Wang, X. (2008). miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *Rna*, *14*(6), 1012-1017.
9. Huang, J., Tan, M., Dou, D., He, L., Townsend, C., & Hayes, P. J. (2010, August). Ontology for microRNA target prediction in human cancer. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology* (pp. 472-474). ACM.
10. Dritsou, V., Topalis, P., Dialynas, E., Mitraka, E., & Louis, C. (2012). miRNAO: An Ontology for microRNAs. In *ICBO*.
11. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., ... & Leontis, N. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, *25*(11), 1251-1255.
12. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, *5*(2), 199-220.
13. Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.
14. Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, *11*(02), 93-136.
15. Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them?. *IEEE Intelligent Systems and their applications*, *14*(1), 20-26.
16. Betel, D., Wilson, M., Gabow, A., Marks, D. S., & Sander, C. (2008). The microRNA. org resource: targets and expression. *Nucleic acids research*, 36(suppl 1), D149-D153.

17. Ceusters, W. (2012, January). An information artifact ontology perspective on data collections and associated representational artifacts. In *MIE* (pp. 68-72).
18. Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., & Ashburner, M. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome biology*, 6(5), R44.
19. Peters, B., & OBI Consortium. (2009). Ontology for biomedical investigations.
20. Fensel, D. Ontologies: A silver bullet for knowledge management and electronic-commerce (2000). *Berlin: Spring-Verlag*.
21. Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, *4*, e05005.
22. Hoehndorf, R., Batchelor, C., Bittner, T., Dumontier, M., Eilbeck, K., Knight, R., ... & Zirbel, C. L. (2011). The RNA Ontology (RNAO): An ontology for integrating RNA sequence and structure data. *Applied Ontology*, *6*(1), 53-89.
23. Motik, B., Patel-Schneider, P. F., Parsia, B., Bock, C., Fokoue, A., Haase, P., ... & Smith, M. (2009). OWL 2 web ontology language: Structural specification and functional-style syntax. *W3C recommendation*, *27*(65), 159.
24. Huang, J., Eilbeck, K., Smith, B., Blake, J. A., Dou, D., Huang, W., ... & Jiang, G. (2016). The Non-Coding RNA Ontology (NCRO): a comprehensive resource for the unification of non-coding RNA biology. *Journal of biomedical semantics*, *7*(1), 24.
25. Huang, J., Townsend, C., Dou, D., Liu, H., & Tan, M. (2011). OMIT: a domain-specific knowledge base for MicroRNA target prediction. *Pharmaceutical research*, *28*(12), 3101-3104.
26. Friedländer, M.R., Chen,W., Adamidi,C., Maaskola,J., Einspanier,R., Knespel,S. and Rajewsky,N. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.,* 26, 407–415
27. Griffiths-Jones, S., Grocock, R. J., Van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research*, *34*(suppl 1), D140-D144.
28. Alexiou, P., Maragkakis, M., Papadopoulos, G. L., Reczko, M., & Hatzigeorgiou, A. G. (2009). Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, *25*(23), 3049-3055.
29. Vergoulis, T., Vlachos, I. S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., ... & Hatzigeorgiou, A. G. (2012). TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic acids research*, *40*(D1), D222-D229.
30. Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., ... & Su, Z. (2010). PMRD: plant microRNA database. *Nucleic acids research*, *38*(suppl 1), D806-D813.
31. Liu, H., Jin, T., Liao, R., Wan, L., Xu, B., Zhou, S., & Guan, J. (2012). miRFANs: an integrated database for Arabidopsis thaliana microRNA function annotations. *BMC plant biology*, *12*(1), 68.
32. Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human

genes are microRNA targets. *cell*, *120*(1), 15-20.
33. Zou, Q., Mao, Y., Hu, L., Wu, Y., & Ji, Z. (2014). miRClassify: an advanced web server for miRNA family classification and annotation. *Computers in biology and medicine*, *45*, 157-160.
34. Ellwanger, D. C., Büttner, F. A., Mewes, H. W., & Stümpflen, V. (2011). The sufficient minimal set of miRNA seed types. *Bioinformatics*, *27*(10), 1346-1350.
35. Neilsen, C. T., Goodall, G. J., & Bracken, C. P. (2012). IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends in Genetics*, *28*(11), 544-549.
36. Xie, M., Zhang, S., & Yu, B. (2015). microRNA biogenesis, degradation and activity in plants. *Cellular and molecular life sciences*, *72*(1), 87-99.
37. Kai, Z. S., & Pasquinelli, A. E. (2010). MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nature structural & molecular biology*, *17*(1), 5-10.
38. Huntley, R. P., Sitnikov, D., Orlic-Milacic, M., Balakrishnan, R., D'Eustachio, P., Gillespie, M. E., ... & Petri, V. (2016). Guidelines for the functional annotation of microRNAs using the Gene Ontology. *RNA*, *22*(5), 667-676.
39. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature genetics*, *39*(10), 1278-1284.
40. Iwai, N., & Naraba, H. (2005). Polymorphisms in human pre-miRNAs. *Biochemical and biophysical research communications*, *331*(4), 1439-1444.
41. Castilla-Llorente, V., Nicastro, G., & Ramos, A. (2013). Terminal loop-mediated regulation of miRNA biogenesis: selectivity and mechanisms.
42. Krol, J., Sobczak, K., Wilczynska, U., Drath, M., Jasinska, A., Kaczynska, D., & Krzyzosiak, W. J. (2004). Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *Journal of Biological Chemistry*, *279*(40), 42230-42239.
43. Shirdel, E. A., Xie, W., Mak, T. W., & Jurisica, I. (2011). NAViGaTing the micronome–using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PloS one*, *6*(2), e17429.
44. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W., & Pasquinelli, A. E. (2016). Pairing beyond the seed supports microRNA targeting specificity. *Molecular Cell*, *64*(2), 320-333.
45. Cullen, B. R. (2009). Viral and cellular messenger RNA targets of viral microRNAs. *Nature*, *457*(7228), 421-425.
46. Ruby, J. G., Jan, C. H., & Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*, *448*(7149), 83-86.
47. Hu, Z., Chen, J., Tian, T., Zhou, X., Gu, H., Xu, L., ... & Chen, Y. (2008). Genetic variants of miRNA sequences and non–small cell lung cancer survival. *The Journal of clinical investigation*, *118*(7), 2600-2608.
48. Mendes, N. D., Freitas, A. T., & Sagot, M. F. (2009). Current tools for the identification of miRNA genes and their targets. *Nucleic acids research*, *37*(8), 2419-2433.

49. Haunsberger, S. J., Connolly, N., & Prehn, J. H. (2017). miRNAmeConverter: an R/bioconductor package for translating mature miRNA names to different miRBase versions. *Bioinformatics*, *33*(4), 592-593.
50. Saçar, M. D., Hamzeiy, H., & Allmer, J. (2013). Can MiRBase provide positive data for machine learning for the detection of MiRNA hairpins?. *Journal of Integrative Bioinformatics (JIB)*, *10*(2), 1-11.
51. Li, L., Xu, J., Yang, D., Tan, X., & Wang, H. (2010). Computational approaches for microRNA studies: a review. *Mammalian Genome*, *21*(1-2), 1-12.
52. Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, *4*, e05005.
53. Arenz, C., (2014) miRNA Maturation. Methods and Protocols. *Editorial Springer*. Volume 1095 2014. ISBN: 978-1-62703-702-0
54. Falconer, S. (2010). OntoGraf Protege Plugin. Place: Available at: http://protegewiki.stanford.edu/wiki/OntoGraf [Accessed: 21/03/2014].
55. Griffiths-Jones, S. (2010). MiRBase. Retrieved June 6, 2017, from http://www.mirbase.org/
56. Vergoulis et al., (2010). Retrieved June 06, 2017, from http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase%2Findex
57. Stanford Center for Biomedical Informatics Research. (n.d.). A free, open-source ontology editor and framework for building intelligent systems. Retrieved June 06, 2017, from http://protege.stanford.edu/
58. Wang, X. (2008). MiRDB - MicroRNA Target Prediction And Functional Study Database. Retrieved June 06, 2017, from http://mirdb.org/
59. Online microRNA prediction tool. (n.d.). Retrieved June 06, 2017, from http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html
60. Whitehead Institute for Biomedical Research (2017). Release 7.1. Retrieved June 06, 2017, from http://www.targetscan.org/vert_71/
61. Wikipedia (2017, May 28). Ontología Retrieved June 06, 2017, from https://es.wikipedia.org/wiki/Ontolog%C3%ADa
62. EMBL-EBI Ontology Lookup Service (2016, March 23). MicroRNA Ontology. Retrieved June 06, 2017, from http://www.ebi.ac.uk/ols/ontologies/mirnao
63. EMBL-EBI Ontology Lookup Service (2016, March 23). Non-Coding RNA Ontology. Retrieved June 06, 2017, from https://www.ebi.ac.uk/ols/ontologies/ncro
64. Falconer, S. (2010). OntoGraf. Retrieved June 06, 2017, from https://protegewiki.stanford.edu/wiki/OntoGraf
65. NCBI (n.d.). Taxonomy. Retrieved June 06, 2017, from https://www.ncbi.nlm.nih.gov/taxonomy
66. EMBL-EBI Ontology Lookup Service (2016, March 23). BRENDA tissue / enzyme source. Retrieved June 06, 2017, from http://www.ebi.ac.uk/ols/ontologies/bto (2017)
67. The Apache Software Foundation (2011). Apache Jena Fusek . Retrieved June 06, 2017, from https://jena.apache.org/documentation/fuseki2/index.html
68. Feigenbaum, L. et al., (2013). SPARQL 1.1 Protocol. Retrieved June 06,

2017, from http://www.w3.org/TR/sparql11-protocol/
69. Klyne, G. et al., (2014) RDF. Retrieved June 06, 2017, from https://www.w3.org/TR/rdf11-concepts/

8. Anexos

Tabla 4

	RNAO (Hoehndorf)	OMIT (Huang)	MIRNAO (Dritsou)
Resource description framework (RDF)	No	Sí	No
Web Ontology Language (OWL)	??	Sí	Sí
Imports upper-level terms from the Basic Formal Ontology (BFO)	Sí	Sí	16 (2.4%)
Imports common relations (shared across different ontologies) from the Relation Ontology (RO)	Sí	Sí	No
Imports ncRNA-related terms and relations from the Non-coding RNA Ontology (NCRO)	No	Sí	No
Imports gene product terms from the GO	Sí	Sí	28 (4.2%)
Imports sequence structural feature terms from the SO	Sí	Sí	348 (51.5%)
Imports life-science and clinical investigation terms from the Ontology for Biomedical Investigations (OBI)	No	Sí	1 (0.1%)
Imports molecular entity (especially small chemical compounds) terms from the Chemical Entities of Biological Interest Ontology (ChEBI)	Sí	Sí	4 (0.6%)
Imports information entity terms from the Information Artifact Ontology (IAO)	No	Sí	No
Imports cell line-relevant terms from the Cell Line Ontology (CLO)	No	Sí	No
Imports protein-related entity terms from the PRO	No	Sí	No
Imports cross-species anatomy terms from the Uberon multi-species anatomy ontology (UBERON)	No	Sí	No
Imports disease terms from the uman Disease	No	Sí	No

Ontology (DOID)			
Gene Regulation Ontology (GRO)	No	No	2 (0.3%)
NCI Thesaurus (NCIThesaurus)	No	No	1 (0.1%)
Ontology for General Medical Science (OGMS)	No	No	1 (0.1%)
RNAO	-	No	9 (1.3%)
miRNAO	-	No	5 (0.7%)

8.2 Glosario de términos ontológicos

Como en el trabajo se ha hablado de muchos términos y el lector no tiene porqué estar familiarizado con ellos, a continuación se muestra un glosario con términos de ontologías. Está estructurado de la siguiente manera:

***

- **Identificator (con hiperenlace) - Nombre - "Definición"**
  - **Subclasse**

***

- IAO_0000030 (Information Artifact Ontology):
  - **Definition** A generically dependent continuant that is about some thing.
  - **Example of usage:** Examples of information content entites include journal articles, data, graphical layouts, and graphs.
- DbXref
- IAO:0000078 The curation status of the term.
  - The allowed values come from an enumerated list of predefined terms. See the specification of these instances for more detailed definitions of each enumerated value.
  - {ready for release, metadata incomplete, example to be eventually removed, uncurated, pending final vetting, metadata complete, organizational term, requires discussion, to be replaced with external ontology term}
- OMIT:0000030 PubMed_id
- OMIT:0000041 PubMed_summary
- OMIT:0000003 information_from_PubMed
- OMIT:0000001 computationally_asserted_evidence
  - OMIT:0000002 information_from_miRNA_target_prediction_databases
    - *has_miRNA_target_gene* some NCRO_0000025 miRNA_target_gene
    - *has_human_miRNA* some NCRO_0000810 human_miRNA

        - OMIT:0000022 prediction_from_DIANA_microT
            - *has_part* some gene_miTG_score_in_DIANA_microT
        - OMIT:0000023 prediction_from_DIANA_mirPath
            - has_part some mirSVR_score_in_DIANA_mirPath
        - OMIT:0000024 prediction_from_Diana_TarBase
            - has_part some methods_in_DIANA_TarBase
        - OMIT:0000019 prediction_from_TargetScan
            - has_part some gene_context_score_in_TargetScan
            - has_part some gene_transcript_id_in_TargetScan
        - OMIT:0000020 prediction_from_miRDB
            - has_part some OMIT:0000012 gene_three_prime_UTR_sequence_in_miRDB
            - has_part some OMIT:0000009 gene_id_in_miRDB
            - has_part some target_score_in_miRDB
            - has_part some OMIT:0000010 gene_transcript_id_in_miRDB
        - OMIT:0000021 prediction_from_miRanda
            - has_part some mirSVR_score_in_miRanda
    - OMIT:0000173 information_from_miRNA_target_validation_database
        - OMIT:0000174 validation_from_miRTarBase
- OMIT:0000004 information_from_KEGG
- OMIT:0000005 information_from_NCBI_gene
    - has_part OMIT:0000035 PubMed_summary_in_NCBI_gene
    - *has_part* OMIT:0000039 summary_in_NCBI_gene
    - *has_part* OMIT:0000040 gene_id_in_NCBI_gene
- OMIT:0000006 information_from_NCBI_nucleotide
    - *has_part* OMIT:0000034 accession_in_NCBI_nucleotide
    - has_part OMIT:0000036 fasta_in_NCBI_nucleotide
    - has_part OMIT:0000037 PubMed_in_NCBI_nucleotide
        - ← *has_gene_id_in_NCBI_gene* OMIT:0000009 gene_id_in_miRDB
- OMIT:0000032 KEGG_pathway_in_DIANA_mirPath
- OMIT:0000031 ensembl_gene_id_in_DIANA_microT
- OMIT:0000009 gene_id_in_miRDB
- OMIT:0000013 gene_transcript_id_in_TargetScan
- OMIT:0000018 methods_in_DIANA_TarBase
- OMIT:0000038 miRNA_target_gene_symbol
- OMIT:0000033 p_value_in_DIANA_mirPath

2

- OBI:0100026 A material entity that is an individual living system, such as animal, plant, bacteria or virus, that is capable of replicating or reproducing, growth and maintenance in the right environment. An organism may be unicellular or made up, like humans, of many billions of cells divided into specialized tissues and organs.

- NCBITaxon:131567 Cellular organism
  - NCBITaxon:554915 Amebozoa
  - NCBITaxon:33208 Metazoa
  - Plant (from Thesaurus.owl) Any living organism that typically synthesizes its food from inorganic substances, possesses cellulose cell walls, responds slowly and often permanently to a stimulus, lacks specialized sense organs and nervous system, and has no powers of locomotion. (EPA Terminology Reference System).
- NCBITaxon:10239 Virus An infectious agent which consists of two parts, genetic material and a protein coat. These organisms lack independent metabolism, and they must infect the cells of other types of organisms to reproduce. Most viruses are capable of passing through fine filters that retain bacteria, and are not visible through a light microscope.
- NCRO_0000015 animal_study_model → no subclasses
- NCRO:0000008 non_mammalian_miRNA
  - NCRO:0000007 virus_miRNA
  - NCRO:0000004 fly_miRNA
  - NCRO:0000005 c_elegans_miRNA
  - NCRO:0000006 bacterial_miRNA
- NCRO:0000009 mammalian_miRNA
  - NCRO:0000810 human_miRNA
  - NCRO:0000701 mouse_miRNA
- NCRO:0000481 Cellular Organism "An organism of microscopic or submicroscopic size, especially a bacterium or protozoan"
  - NCRO:0000482 archea “microorganisms that are similar to bacteria in size and simplicity of structure but radically different in molecular organization. They are now believed to constitute an ancient intermediate group between the bacteria and eukaryotes” → no subclasses
  - NCRO:0000483 bacteria “a member of a large group of unicellular microorganisms that have cell walls but lack organelles and an organized nucleus, including some that can cause diseas” → no subclasses
- NCRO:0000091 Virus “Small infectious agent that replicates only inside the living cells of other organisms” → no subclasses

- UBERON:0000465 material_anatomical_entity “Anatomical entity that has mass.”
  - UBERON:0000061 anatomical_structure “Material anatomical entity that is a single connected structure with inherent 3D shape generated by coordinated expression of the organism's own genome.”
    - UBERON:0000062 organ “Anatomical structure that performs a specific function or group of functions [WP].”

- UBERON:0000479 tissue “Multicellular anatomical structure that consists of many cells of one or a few types, arranged in an extracellular matrix such that their long-range organisation is at least partly a repetition of their short-range organisation”
    - NCRO:0000098 adult_tissue
        - NCRO:0000100 connective_tissue “Tissue that connects, supports, binds, or separates other tissues or organs, typically having relatively few cells embedded in an amorphous matrix, often with collagen or other fibers, and including cartilaginous, fatty, and elastic tissues.” → UBERON:0002384
            - NCRO:0000104 adipose_tissue “A kind of body tissue containing stored fat that serves as a source of energy; it also cushions and insulates vital organs.” → UBERON:0001013
            - NCRO:0000105 blood “Bodily fluid in animals that delivers necessary substances such as nutrients and oxygen to the cells and transports metabolic waste products away.” → UBERON:0000178
            - NCRO:0000106 bone “Rigid organs that constitute part of the endoskeleton of vertebrates.” → UBERON:0002481
            - NCRO:0000107 cartilage “A tough, elastic, fibrous connective tissue found in various parts of the body, such as the joints, outer ear, and larynx.” → UBERON:0002418
        - NCRO:0000101 epithelia_tissue “Membranous tissue covering internal organs and other internal surfaces of the body.” → UBERON:0000483
        - NCRO:0000102 muscle_tissue “Muscle tissue is one of four basic biological tissues present in animals. It is a soft tissue that composes muscles. It is formed during embryo development in a process known as myogenesis.” → UBERON:0002385
            - NCRO:0000108 cardiac_muscle “Type of involuntary striated muscle found in the walls and histological foundation of the heart.” → UBERON:0001133
            - NCRO:0000109 skeletal_muscle “Form of striated muscle tissue which is under the control of the somatic nervous system.” → UBERON:0004755

            - NCRO:0000110 smooth_muscle "Responsible for the contractility of hollow organs, such as blood vessels, the gastrointestinal tract, the bladder, or the uterus." → UBERON:0001135
        - NCRO:0000103 nerve_tissue "Tissue composed of neurons." → UBERON:0003714
    - NCRO:0000099 embryonic_tissue "tissue in an embryo" → UBERON:0005291
        - NCRO:0000111 ectoderm "Outermost of the three primary germ layers of an embryo, from which the epidermis, nervous tissue, and, in vertebrates, sense organs develop." → UBERON:0000924
        - NCRO:0000112 endoderm "Innermost of the three primary germ layers of an animal embryo, developing into the gastrointestinal tract, the lungs, and associated structures." → UBERON:0000925
        - NCRO:0000113 mesoderm "Middle embryonic germ layer, lying between the ectoderm and the endoderm, from which connective tissue, muscle, bone, and the urogenital and circulatory systems develop." → UBERON:0000926
- NCRO:0001667 miRNA_gene_family
- FiatObjectPart "A material entity that is part of an object but is not demarcated by any physical discontinuities."
    - SO:0000001 Region "A sequence feature with an extent greater than zero. A nucleotide region is composed of bases and a polypeptide region is composed of amino acids."
        - SO:0001411 Biological Region "A region defined by its disposition to be involved in a biological process."
            - SO:0000704 Gene
                - NCRO:0000025 miRNA Target Gene
                - SO:0001263 ncRNA_gene
                - SO:0001265 miRNA_gene
            - SO:0000409 Binding Site
                - SO:0001655 Nucleotide Binding Site
                - SO:0001429 DNA Binding Site
                - SO:0000934 miRNA Target Site
                - SO:0000410 Protein Binding Site
            - SO:0000833 Transcript Region
                - SO:0001243 miRNA Primary Transcript Region
            - SO:0000830 chromosome_part
- BFO:0000030 object
    - GO:0005575 cellular component
        - GO:0005694 chromosome

GO:0044747 mature miRNA 3'-end processing “Any process involved in forming distinct **miRNA isoforms** from a mature miRNA that differ at their 3'-ends.”
OMIT:0027585 RNA Isoforms

3

MIRNATO

MIRNAO
SO:0000647 pri-miRNA
GO:0006351 Transcription, DNA-dependent
GO:0001172 Transcription, RNA-dependent
GO:0003677 DNA Binding
GO:0003723 RNA Binding
  GO:0070878 pri-miRNA Binding
  GO:0070883 pre-miRNA binding
  GO:0035198 miRNA Binding
Transcription Factor

NCRO & OMIT
TRANSCRIPTION
GO:0006351 transcription, DNA-templated
NCRO:0000011 protein_miRNA_promoter_binding
NCRO:0000012 promoter_of_miRNA
NCRO:0000013 miRNA_transcription_initiation
NCRO:0000014 transcription_of_miRNA
  GO:0061614 pri-miRNA transcription from RNA polymerase II promoter
    GO:1902893 regulation of pri-miRNA transcription from RNA polymerase II promoter
    GO:1902894 negative regulation of pri-miRNA transcription from RNA polymerase II promoter “Any process that stops, prevents or reduces the frequency, rate or extent of pri-miRNA transcription from RNA polymerase II promoter”
    GO:1902895 positive regulation of pri-miRNA transcription from RNA polymerase II promoter “Any process that activates or increases the frequency, rate or extent of pri-miRNA transcription from RNA polymerase II promoter.”
GO:0010445 nuclear dicing body “A small round nuclear body, measuring 0.2-0.8 microns in diameter that is diffusely distributed throughout the nucleoplasm. Several proteins known to be involved in miRNA processing have been localized to these structures. D-bodies are thought to be involved in primary-miRNA processing and/or storage/assembly of miRNA processing complexes”

CLEAVAGE
GO:0031053 pri-miRNA processing (all processes till getting mature miRNA)
  GO:2000636 positive regulation of primary miRNA processing
  GO:2000635 negative regulation of primary miRNA processing
GO:0070877 microprocessor complex

OMIT:0041308 DROSHA
OMIT:0042326 DGCR8
GO:0016891 endoribonuclease activity, producing 5'-phosphomonoesters (including RNase III)
GO:1990729 primary miRNA modification "The covalent alteration of one or more nucleotides within a primary miRNA molecule to produce a primary miRNA molecule with a sequence that differs from that coded genetically"

GO:1903800 positive regulation of production of miRNAs involved in gene silencing by miRNA
GO:1903799 negative regulation of production of miRNAs involved in gene silencing by miRNA

NUCLEAR EXPORT
GO:0070883 pre-miRNA binding
GO:0031054 pre-miRNA processing
GO:2000631 regulation of pre-miRNA processing
  GO:2000632 negative regulation of pre-miRNA processing
  GO:2000633 positive regulation of pre-miRNA processing
GO:0090631 pre-miRNA transporter activity
GO:0035281 pre-miRNA export from nucleus
GO:0071430 pre-miRNA-containing ribonucleoprotein complex export from nucleus
OMIT:0043831 XPO5
OMIT:0020755 ran GTP-binding protein

CLEAVAGE PRE-MIRNA
GO:2000631 regulation of pre-miRNA processing
GO:2000632 negative regulation of pre-miRNA processing
GO:2000633 positive regulation of pre-miRNA processing
OMIT:0039525 DICER1
OMIT:0035592 TARBP2

ARGONAUTE (RISC FORMATION)
GO:0035198 miRNA Binding
GO:0016442 RISC complex
GO:0070578 RISC-loading complex
GO:0035280 miRNA loading onto RISC involved in gene silencing by miRNA
GO:0035068 micro-ribonucleoprotein complex
OMIT:0027799 Argonaute Proteins
OMIT:0040228 AGO1
OMIT:0040517 AGO2
OMIT:0049796 AGO3
OMIT:0049797 AGO4

OTHER GO TERMS TO TAKE IN CONSIDERATION (Huntley et al., 2016)
GO:0001034 RNA polymerase III transcription factor activity, sequence-specific DNA

binding

TF REGULATORS
GO:1902893 regulation of pri-miRNA transcription from RNA polymerase II promoter
GO:0001085 RNA polymerase II transcription factor binding "Interacting selectively and non-covalently with an RNA polymerase II transcription factor, any protein required to initiate or regulate transcription by RNA polymerase II"
[GO:0001025 RNA polymerase III transcription factor binding "Interacting selectively and non-covalently with an RNA polymerase III transcription factor, any protein required to initiate or regulate transcription by RNA polymerase III."]
[GO:1903095 ribonuclease 3 complex]

TFs
GO:1902893 regulation of pri-miRNA transcription from RNA polymerase II promoter
GO:0000981 RNA polymerase II transcription factor activity, sequence-specific DNA binding

MICROPROCESSOR
GO:0070878 pri-miRNA Binding
GO:0031053 pri-miRNA processing (all processes till getting mature miRNA)
GO:0070877 microprocessor complex "A protein complex that binds to heme and to pri-miRNAs, and is required for the formation of a pre-microRNA (pre-miRNA), the initial step of microRNA (miRNA) biogenesis. The complex is composed of the double-stranded-RNA-specific RNase **Drosha** (also called RNASEN) and the RNA-binding protein **DGCR8** (heme-free or heme-bound forms)"
GO:0004525 ribonuclease III activity "Catalysis of the endonucleolytic cleavage of RNA with 5'-phosphomonoesters and 3'-OH termini; makes two staggered cuts in both strands of dsRNA, leaving a 3' overhang of 2 nt."
GO:0090502 RNA phosphodiester bond hydrolysis, endonucleolytic "The chemical reactions and pathways involving the hydrolysis of internal 3',5'-phosphodiester bonds in one or two strands of ribonucleotides."

EXPORTIN-5 COMPLEX
GO:0035281 pre-miRNA export from nucleus
GO:0070883 pre-miRNA binding
GO:0042565 RNA nuclear export complex
XPO-5:
GO:0008536 Ran GTPase binding
GO:0090631 pre-miRNA transporter activity
RAN
contribute_to GO:0090631 pre-miRNA transporter activity
contribute_to GO:0070883 pre-miRNA binding

RISC loading complex
GO:0098851 double-stranded miRNA binding
GO:0010501 RNA secondary structure unwinding

GO:0035280 miRNA loading onto RISC involved in gene silencing by miRNA
GO:0035196 production of miRNAs involved in gene silencing by miRNA
GO:0070578 RISC-loading complex

DICER1
GO:0004525 Ribonuclease III activity
GO:0090502 RNA phosphodiester bond hydrolysis, endonucleolytic

ARGONAUTES
GO:0035198 miRNA Binding
GO:0010501 RNA secondary structure undwinding
GO:0016442 RISC Complex
GO:0035278 miRNA mediated inhibition of translation
GO:0098806 deadenylation involved in gene silencing by miRNA

AGO2
GO:0090624 endoribonuclease activity, cleaving miRNA-paired mRNA
GO:0035279 mRNA cleavage involved in gene silencing by miRNA

GO HUNTLEY - PLANTS
TF REGULATORS =

miPEPs
GO:1902893 regulation of pri-miRNA transcription from RNA polymerase II promoter
TFs =

Nuclear dicing body proteins (ocu, SE, l-lYL1, DDL, TGl-l):
GO:0031053 pri-miRNA processing (all processes till getting mature miRNA)
GO:0070878 pri-miRNA Binding
GO:0035196 production of miRNAs involved in gene silencing by miRNA
GO:0031054 pre-miRNA processing
GO:0070883 pre-miRNA binding
GO:0010445 nuclear dicing body**

DCL1:
GO:0004525 ribonuclease III activity
GO:0090502 RNA phosphodiester bond hydrolysis, endonucleolytic

Cap binding complex proteins (CBPZO and came):
GO:0031053 pri-miRNA processing (all processes till getting mature miRNA)
GO:0035196 production of miRNAs involved in gene silencing by miRNA
GO:0005846 nuclear cap binding complex**

HEN1:
GO:0008757 S-adenosylmethionine-dependent methyltransferase activity‘

GO:0061715 miRNA 2' O—methylation

HST:
GO:0061716 miRNA export from nucleus
GO:0061717 miRNA transporter activity RISC complex U U ‘j
GO:0042565 RNA nuclear export complex (= Exportin-5 complex)

HSP90 and SQN:
GO:0098851 double-stranded miRNA binding
GO:0035280 miRNA loading onto RISC involved in gene silencing by miRNA (=RISC)

Argonautes (AGO1-10):
GO:0098851 double-stranded miRNA binding
GO:0035198 miRNA binding
GO:0010501 RNA secondary structure unwinding
GO:0035278 miRNA mediated inhibition of translation
GO:0016442 RISC Complex

AG01, 2, 4, 7, 10:
GO:0090624 endoribonuclease activity, cleaving miRNA-paired mRNA
GO:0035279 mRNA cleavage involved in gene silencing by miRNA

4

5

MIRNAO
SO:0001243 miRNA Primary Transcript Region
  SO:0001473 miRNA Antiguide “A region of the pri miRNA that basepairs with the guide to form the hairpin”
  SO:0001246 miRNA loop “The loop of the hairpin loop formed by folding of the pre-miRNA”
  SO:0001245 miRNA stem “The stem of the hairpin loop formed by folding of the pre-miRNA”
  SO:0000163 5' cis Splice Site
    SO:0000677 Canonical 5' Splice Site
    SO:0000679 Non-canonical 5' Splice SIte
SO:0000122 RNA Secondary Structure
  RNAO:0000098 Terminal Loop
  RNAO:0000106 Stem junction
  RNAO:0000097 Unbroken Stem
  RNAO:0000107 Strand End
  RNAO:0000096 Unpaired Region
  RNAO:0000099 Internal loop
SO:0000836 Mature Transcript Region
  SO:0000205 Three Prime UTR
    SO:0001180 AU rich element “A cis-acting element found in the 3' UTR of some

mRNA which is rich in AUUUA pentamers. Messenger RNAs bearing multiple AU-rich elements are often unstable."

SO:0001655 Nucleotide Binding Site
SO:0000934 miRNA Binding Site "A miRNA target site is a binding site where the molecule is a micro RNA"

NCRO
All under "biological region" class
SO:0000830 chromosome part
SO:0000704 gene
SO:0001265 miRNA gene
NCRO:0000025 miRNA_target_gene
SO:0000831 gene_member_region
SO:0000233 mature_transcript
NCRO:0000002 miRNA_target_gene_mRNA
SO:0000185 primary_transcript
SO:0000647 miRNA_primary_transcript "A primary transcript encoding a micro RNA."
NCRO:0000001 miRNA_target_gene_primary_transcript

6

MIRNAO
SO:0000409 Binding Site
SO:0001655 Nucleotide Binding Site
SO:0000934 miRNA Target Site

RNAO:0000095 paired region

NCRO
NCRO:0000002 miRNA_target_gene_mRNA
SO:0000239 flanking_region

7

NCRO
GO:0035279 mRNA cleavage involved in gene silencing by miRNA
GO:0035195 gene silencing by miRNA

GO
GO:0098806 deadenylation involved in gene silencing by miRNA
GO:0035278 miRNA mediated inhibition of translation

8

JERARQUÍA DE LA CLASSE “CONTINUANT” DE MIRNAO

- **Continuant** (or snapshot) ontologies, comprehending continuant entities such as three-dimensional enduring objects
    - **Dependent Continuant** A continuant [snap:Continuant] that is either dependent on one or other independent continuant [snap:IndependentContinuant] bearers or inheres in or is borne by other entities.
        - **Generically Dependent Continuant** A continuant [snap:Continuant] that is dependent on one or other independent continuant [snap:IndependentContinuant] bearers. For every instance of A requires some instance of (an independent continuant [snap:IndependentContinuant] type) B but which instance of B serves can change from time to time.
            - DNA Sequence
            - RNA Sequence
        - **Specifically Dependent continuant** A continuant [snap:Continuant] that inheres in or is borne by other entities. Every instance of A requires some specific instance of B which must always be the same.
            - **Quality** A specifically dependent continuant [snap:SpecificallyDependentContinuant] that is exhibited if it inheres in an entity or entities at all (a categorical property).
            - **Realizable Entity** A specifically dependent continuant [snap:SpecificallyDependentContinuant] that inheres in continuant [snap:Continuant] entities and are not exhibited in full at every time in which it inheres in an entity or group of entities. The exhibition or actualization of a realizable entity is a particular manifestation, functioning or process that occurs under certain circumstances.
                - **Function** A realizable entity [snap:RealizableEntity] the manifestation of which is an essentially end-directed activity of a continuant [snap:Continuant] entity in virtue of that continuant [snap:Continuant] entity being a specific kind of entity in the kind or kinds of contexts that it is made for.
                - Molecular Function
                - Binding
                - Nucleic Acid Binding
                - DNA binding
                - RNA binding
                - → miRNA Binding
                - → pre-miRNA Binding

            - → pri-miRNA Binding
        - **Role**
            - **Transcription** factor A role played by a protein that binds to specific DNA sequences, thereby controlling the transcription of genetic information from DNA to mRNA.
    - **Independent Continuant** A continuant [snap:Continuant] that is a bearer of quality [snap:Quality] and realizable entity [snap:RealizableEntity] entities, in which other entities inhere and which itself cannot inhere in anything.
        - Immaterial Entity
        - Material Entity
            - Collection of Material Entities
            - **Fiat Object Part** A material entity that is part of an object but is not demarcated by any physical discontinuities.

occurrent ontologies, comprehending processes conceived as extended through (or as spanning) time.