
Papel de los Genes del Reloj Circadiano en los Trastornos del Sueño

Creación y explotación de un grafo RDF

Autora: Laura Llamas López

Asignatura: Explotación Semántica de Datos

Curso Académico: 2025-2026

Titulación: Máster en Bioinformática

Universidad de Murcia

Enero 2026

Índice

1. Introducción	2
1.1. Contexto Biológico	2
1.1.1. Genes del reloj circadiano	2
1.1.2. Trastornos del sueño	2
1.1.3. Variantes genéticas asociadas a trastornos circadianos	3
1.1.4. Tratamientos farmacológicos	3
1.2. Fundamentos de grafos RDF y web semántica	4
2. Objetivos	4
3. Materiales y Métodos	4
3.1. Búsqueda Bibliográfica	4
3.2. Creación del grafo RDF	5
3.2.1. Definición de prefijos y URIs	5
3.2.2. Modelado de propiedades semánticas	5
3.2.3. Modelado de entidades	5
3.2.4. Relaciones del loop circadiano	6
3.3. Carga del RDF dataset en Blazegraph	6
3.4. Generación de consultas SPARQL	7
4. Resultados y Discusión	7
4.1. Grafo creado y validación	7
4.2. Resultado de las consultas realizadas	8
4.2.1. Consulta 0: Verificación de carga	8
4.2.2. Consulta 1: ¿Qué fenotipos clínicos caracterizan cada trastorno del sueño?	8
4.2.3. Consulta 2: ¿Qué fármacos están indicados para cada trastorno circadiano?	9
4.2.4. Consulta 3: ¿Qué mutaciones missense patogénicas causan ASPD?	9
4.2.5. Consulta 4a: ¿Qué relaciones de interacción física y regulación existen en el TTFL?	10
4.2.6. Consulta 4b: ¿Qué proteínas actúan como hubs regulatorios en la red circadiana?	11
4.2.7. Consulta 5: ¿Qué nombres oficiales reporta UniProt para las proteínas del grafo?	12
5. Conclusiones	12

1. Introducción

1.1. Contexto Biológico

Los ritmos circadianos son oscilaciones biológicas con periodicidad aproximada de 24 horas que regulan funciones esenciales como el ciclo sueño-vigilia, la temperatura corporal y el metabolismo. Los trastornos del ritmo circadiano del sueño (CRSDs) se caracterizan por una desalineación entre el reloj endógeno y el ciclo ambiental, pudiendo originarse por mutaciones en genes del reloj circadiano que producen fenotipos hereditarios con manifestaciones clínicas distintivas [1, 2]. El estudio de estas variantes genéticas ha permitido comprender los mecanismos moleculares de la regulación temporal del sueño y ha abierto nuevas vías para el desarrollo de terapias dirigidas [3].

1.1.1. Genes del reloj circadiano

Los ritmos circadianos se generan mediante un bucle de retroalimentación transcripcional traduccional (TTFL, *Transcriptional-Translational Feedback Loop*) (Figura 1) donde los factores CLOCK y BMAL1 forman un heterodímero que activa la transcripción de los genes *Period* (*PER1*, *PER2*, *PER3*) y *Cryptochrome* (*CRY1*, *CRY2*). Las proteínas PER y CRY se acumulan en el citoplasma, se traslocan al núcleo y reprimen su propia transcripción al inhibir el complejo CLOCK:BMAL1, completando el ciclo en aproximadamente 24 horas [4, 5]. BMAL1 es el único componente esencial cuya ausencia produce arritmicidad completa [6], mientras que NPAS2 puede compensar parcialmente la función de CLOCK en ciertas regiones cerebrales [7].

La regulación post-traducciona es fundamental para el control temporal del reloj. Las caseína quinasas CSNK1D y CSNK1E fosforilan las proteínas PER marcándolas para degradación proteasomal, mientras que el receptor nuclear NR1D1 (REV-ERB α) reprime la transcripción de BMAL1 contribuyendo a la robustez del oscilador [8, 9].

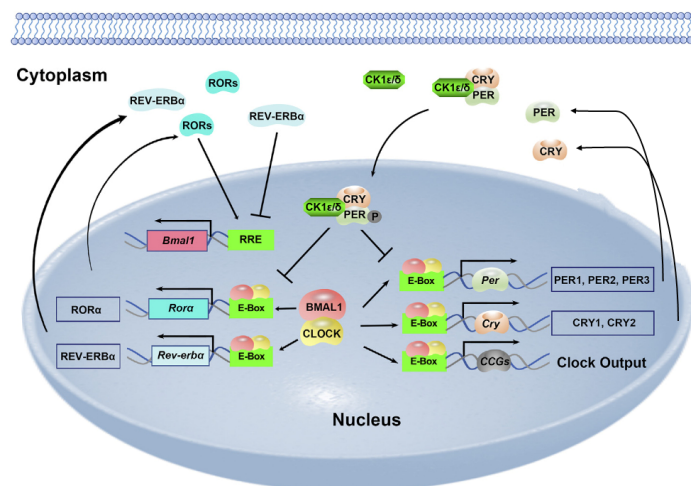


Figura 1: Esquema del bucle de retroalimentación transcripcional-traducciona (TTFL) del reloj circadiano [10].

1.1.2. Trastornos del sueño

Los principales trastornos modelados en este grafo son:

- **Síndrome de Fase de Sueño Adelantada (ASPD):** Cursa con somnolencia temprana (18:00-21:00h) y despertar espontáneo (02:00-05:00h), reflejando un periodo circadiano acortado de ~ 23 horas [11].
- **Trastorno de Fase de Sueño Retrasada (DSPD):** El más prevalente entre los CRSDs, se manifiesta con incapacidad para conciliar el sueño antes de las 02:00-04:00h y despertar durante el mediodía, debido a un periodo endógeno prolongado superior a 24.5 horas [12].
- **Trastorno de Ritmo Sueño-Vigilia Irregular:** Presenta múltiples episodios breves de sueño fragmentado sin patrón consolidado, prevalente en enfermedades neurodegenerativas [13].
- **Trastorno No-24 Horas:** Caracterizado por un ciclo de curso libre, afecta al 50-70 % de personas con ceguera total que carecen de sincronización lumínica [14].
- **Insomnio Crónico:** Definido como dificultad persistente para iniciar o mantener el sueño durante al menos tres meses, tiene etiología multifactorial incluyendo alteraciones en genes circadianos [15].

1.1.3. Variantes genéticas asociadas a trastornos circadianos

Las variantes en genes del reloj circadiano alteran el ciclo sueño-vigilia produciendo trastornos mendelianos con fenotipos definidos:

- **PER2 p.Ser662Gly:** Localizada en un sitio de fosforilación por CSNK1E, reduce la degradación de PER2 acumulándola en el núcleo y acortando el periodo circadiano a ~ 23 horas, manifestándose como síndrome de fase adelantada con herencia autosómica dominante [16].
- **CRY1 c.1657+3A>C:** Variante de splicing que produce elongación de 24 aminoácidos en la región C-terminal, estabilizando la proteína y prolongando el periodo circadiano hasta 24.5-25 horas, constituyendo la causa más frecuente de fase retrasada [17, 18].
- **CSNK1D p.Thr44Ala:** Mutación missense que reduce la actividad quinasa, causando ASPD con herencia autosómica dominante [16].
- **PER3 VNTR:** El polimorfismo de repeticiones en tándem (4 o 5 repeticiones de 54 nucleótidos) modula el cronotipo individual [19].
- **CLOCK 3111T>C:** Variante sinónima asociada con cronotipo vespertino [20].

1.1.4. Tratamientos farmacológicos

El abordaje terapéutico de los CRSDs se fundamenta en cronobióticos que modifican el periodo o fase circadiana (Cuadro 1).

Fármaco	Mecanismo	Indicación principal
Melatonina	Agonista MT1/MT2	DSPD, ASPD, Insomnio
Ramelteon	Agonista selectivo MT1/MT2	Insomnio de conciliación
Tasimelteon	Agonista dual MT1/MT2	Trastorno Non-24h

Cuadro 1: Principales fármacos utilizados en el tratamiento de los CRSDs.

La melatonina administrada en dosis de 0.5-5 mg entre 1-2 horas antes del sueño deseado adelanta la fase en trastornos de fase retrasada [21, 22]. Para el ASPD la melatonina se toma por la mañana, mientras que para el DSPD se administra por la noche. Ramelteon es un agonista selectivo de los receptores de melatonina MT1 y MT2 con mayor afinidad que la melatonina endógena, indicado para el insomnio de conciliación [23]. El tasimelteon es el primer fármaco aprobado específicamente para el trastorno Non-24 horas en personas ciegas [24].

1.2. Fundamentos de grafos RDF y web semántica

Los grafos de conocimiento basados en Resource Description Framework (RDF) ofrecen una alternativa flexible para la representación e integración de datos biomédicos heterogéneos. Un grafo RDF se construye sobre tripletas sujeto-predicado-objeto, donde cada afirmación se representa como una relación dirigida entre dos recursos. Esta estructura permite construir redes de conocimiento altamente interconectadas donde cada entidad puede actuar como sujeto de unas relaciones y objeto de otras, facilitando el descubrimiento de conexiones implícitas [25].

La potencia del modelo RDF se amplifica mediante ontologías, vocabularios controlados que definen formalmente las clases de entidades y las relaciones permitidas entre ellas. Ontologías biomédicas como Gene Ontology, Human Phenotype Ontology o BioLink Model proporcionan un lenguaje común que permite la integración semántica de datos de distintas bases de datos. Los grafos RDF se consultan mediante SPARQL, un lenguaje declarativo que navega libremente por la red de relaciones permitiendo formular consultas complejas que atraviesan múltiples entidades y recuperan subgrafos que cumplan criterios específicos.

2. Objetivos

El objetivo principal es demostrar las competencias adquiridas en Explotación Semántica de Datos mediante:

1. **Creación de un grafo RDF semánticamente rico** que represente las relaciones entre genes del reloj circadiano, proteínas, variantes patogénicas, trastornos del sueño, fenotipos clínicos y tratamientos farmacológicos. El grafo debe contener ≥ 30 tripletas y reutilizar vocabularios estándar.
2. **Diseño de consultas SPARQL complejas:** Implementar ≥ 5 consultas que empleen operadores avanzados, documentadas mediante script R reproducible.

3. Materiales y Métodos

3.1. Búsqueda Bibliográfica

Para la construcción del grafo de conocimiento se realizó una búsqueda bibliográfica avanzada en PubMed. La anotación semántica de las entidades del grafo se realizó consultando ontologías biomédicas estándar (BioLink Model, Relations Ontology, Gene Ontology, HPO, MONDO) y bases de datos de referencia internacional (UniProt, NCBI Gene, ClinVar, DrugBank, RefSeq, dbSNP).

3.2. Creación del grafo RDF

Para la construcción del grafo de conocimiento sobre cronobiología se utilizó el lenguaje Turtle (Terse RDF Triple Language) por su sintaxis clara y legible. El grafo se estructuró en 12 secciones temáticas claramente diferenciadas.

3.2.1. Definición de prefijos y URIs

Se establecieron prefijos mediante la directiva `@prefix` para facilitar la reutilización de URIs. Los espacios de nombres definidos se clasifican en tres categorías:

a) **Prefijos personalizados del dominio:**

- **crono:** Identifica recursos específicos del dominio (genes, proteínas, variantes, fenotipos, fármacos, trastornos, etc).
- **crono_o:** Define propiedades ontológicas personalizadas (codifica, localizadoEn, interactuaCon, causaTrastorno, etc.).

b) **Prefijos de infraestructura RDF/OWL:**

- **rdf:** Vocabulario fundamental de RDF (type).
- **rdfs:** RDF Schema para definición de clases y propiedades (label, comment, seeAlso).
- **owl:** Web Ontology Language para lógica descriptiva.

c) **Prefijos de ontologías y bases de datos biomédicas externas:** Permiten vincular las entidades locales con identificadores persistentes internacionales.

3.2.2. Modelado de propiedades semánticas

Se definieron 13 propiedades relacionales personalizadas mediante `owl:ObjectProperty`. Para garantizar la interoperabilidad con sistemas externos, la mayoría de propiedades locales se mapeó simultáneamente a BioLink Model y Relations Ontology utilizando `rdfs:seeAlso`:

```
crono_o:codifica rdf:type owl:ObjectProperty ;
rdfs:label "codifica"@es ;
rdfs:seeAlso biolink:has_gene_product ;
rdfs:seeAlso ro:0002205 .
```

El cuadro 2 resume el mapeo de algunas propiedades locales a vocabularios estándar:

Propiedad local	BioLink	RO	Descripción
crono_o:codifica	has_gene_product	RO:0002205	Gen → Proteína
crono_o:interactuaCon	interacts_with	RO:0002434	Interacción proteína proteína (PII)
crono_o:causaTrastorno	causes	RO:0003303	Variante → Enfermedad
crono_o:tieneFenotipo	has_phenotype	RO:0002200	Enfermedad → Fenotipo
crono_o:tratadoCon	treated_by	RO:0002302	Enfermedad → Fármaco

Cuadro 2: Mapeo de propiedades locales a vocabularios estándar.

3.2.3. Modelado de entidades

El grafo incluye 8 categorías principales de entidades, tipadas según BioLink Model y vinculadas a identificadores persistentes de recursos externos:

- **Procesos biológicos** (biolink:BiologicalProcess): 3 procesos vinculados a Gene Ontology.
- **Cromosomas** (biolink:NucleicAcidEntity): 7 cromosomas vinculados a RefSeq.
- **Fenotipos** (biolink:PhenotypicFeature): 10 fenotipos clínicos vinculados a HPO.
- **Fármacos** (biolink:Drug): 3 cronobióticos vinculados a DrugBank.
- **Trastornos** (biolink:Disease): 5 trastornos vinculados a MONDO con sus fenotipos asociados y tratamientos.
- **Genes** (biolink:Gene): 11 genes del reloj circadiano vinculados a NCBI Gene.
- **Proteínas** (biolink:Protein): 11 proteínas vinculadas a UniProt con sus interacciones.
- **Variantes** (biolink:SequenceVariant): 5 variantes genéticas clasificadas según Sequence Ontology con patogenicidad y patrón de herencia.

3.2.4. Relaciones del loop circadiano

Una característica distintiva del grafo es la representación explícita del bucle de retroalimentación transcripcional-traducciona (TTFL):

- **Loop positivo:** Las proteínas CLOCK y BMAL1 activan la transcripción de los genes del loop negativo (PER1, PER2, PER3, CRY1, CRY2, NR1D1).
- **Loop positivo alternativo:** NPAS2 (parálogo de CLOCK) activa genes del loop negativo en cerebro anterior.
- **Loop negativo:** Las proteínas PER1, PER2, CRY1 y CRY2 reprimen la actividad del heterodímero CLOCK:BMAL1.
- **Loop estabilizador:** NR1D1 (REV-ERB α) reprime la transcripción de BMAL1.

3.3. Carga del RDF dataset en Blazegraph

Con el objetivo de habilitar consultas semánticas sobre el grafo generado, se procedió a su despliegue en el triple store Blazegraph disponible en el servidor docente del máster (dayhoff.inf.um.es:3039). El proceso de carga consistió en:

1. **Conversión del formato:** El archivo Turtle se convirtió a formato RDF/XML mediante la herramienta EasyRDF¹.
2. **Validación sintáctica:** El grafo se validó utilizando el validador oficial del W3C² para verificar la corrección estructural.
3. **Creación del namespace:** Desde la interfaz web de Blazegraph, se creó un nuevo namespace denominado `cronobiologia`.
4. **Carga del grafo:** Se procedió a cargar el grafo RDF mediante la pestaña UPDATE de Blazegraph.

¹<https://www.easyrdf.org/converter>

²<https://www.w3.org/RDF/Validator/>

3.4. Generación de consultas SPARQL

Una vez cargado el grafo RDF en Blazegraph, se procedió a realizar consultas SPARQL desde el entorno de programación RStudio. Se utilizó la librería SPARQL de R para enviar consultas al endpoint remoto:

```
library(SPARQL)

# Endpoint de Blazegraph
endpoint <- "http://dayhoff.inf.um.es:3039/blazegraph/namespace/
cronobiologia/sparql"

# Funcion para ejecutar consultas
ejecutar_consulta <- function(query) {
  resultado <- SPARQL(endpoint, query)
  return(as.data.frame(resultado$results))
}
```

Se diseñaron 7 consultas SPARQL de complejidad creciente (Cuadro 3):

N°	Objetivo	Operadores SPARQL	Complejidad
0	Verificación de carga	COUNT, ORDER BY	Básica
1	Trastornos y fenotipos	OPTIONAL, FILTER	Media
2	Tratamientos farmacológicos	OPTIONAL, FILTER	Media
3	Variantes missense ASPD	Múltiples FILTER, JOIN	Alta
4a	Red de regulación TTFL	UNION, BIND	Alta
4b	Centralidad de red	GROUP BY, HAVING, agregación	Alta
5	Consulta federada UniProt	SERVICE, BIND, REPLACE	Muy alta

Cuadro 3: Resumen de las consultas SPARQL diseñadas.

4. Resultados y Discusión

El repositorio con todos los archivos generados está disponible en:

<https://github.com/laullamas/practica-esd-cronobiologia>

4.1. Grafo creado y validación

Métrica	Valor
Total de tripletas	436
Genes modelados	11
Proteínas modeladas	11
Variantes genéticas	5
Trastornos	5
Fenotipos clínicos	10
Fármacos	3
Propiedades personalizadas	13

Cuadro 4: Estadísticas del grafo RDF creado.

En el Cuadro 4 se resumen las características del grafo generado. La validación mediante el validador oficial del W3C confirmó que todas las tripletas presentan sintaxis válida conforme al estándar RDF 1.1, sin errores estructurales ni advertencias semánticas. El grafo fue cargado exitosamente en el namespace `cronobiologia` del servidor Blazegraph.

4.2. Resultado de las consultas realizadas

4.2.1. Consulta 0: Verificación de carga

```
SELECT (COUNT(*) AS ?total)
WHERE { ?sujeto ?predicado ?objeto . }
```

El grafo contiene 436 tripletas, superando ampliamente el mínimo requerido de 30. La recuperación exitosa de todas las tripletas confirma que la carga se realizó correctamente.

4.2.2. Consulta 1: ¿Qué fenotipos clínicos caracterizan cada trastorno del sueño?

La correlación genotipo-fenotipo es fundamental para el diagnóstico diferencial, la medicina de precisión y la interoperabilidad con bases de datos clínicas internacionales mediante HPO.

```
consulta1 <- '
PREFIX crono: <http://cronobiologia.um.es/recurso/>
PREFIX crono_o: <http://cronobiologia.um.es/ontologia/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX biolink: <https://w3id.org/biolink/vocab/>

SELECT ?trastorno ?nombreTrastorno ?fenotipo ?nombreFenotipo ?descripcionFenotipo ?hpo
WHERE {
  ?trastorno rdf:type biolink:Disease ;
    rdfs:label ?nombreTrastorno ;
    crono_o:tieneFenotipo ?fenotipo .

  ?fenotipo rdfs:label ?nombreFenotipo ;
    rdfs:comment ?descripcionFenotipo .

  OPTIONAL {
    ?fenotipo rdfs:seeAlso ?hpo .
    FILTER(CONTAINS(STR(?hpo), "HP_"))
  }
}
ORDER BY ?nombreTrastorno ?nombreFenotipo
'
```

La consulta devuelve los 5 trastornos con sus fenotipos asociados. Se observa un solapamiento significativo entre trastornos (dificultad para conciliar el sueño, desincronización circadiana, sueño fragmentado), lo que refleja su base fisiopatológica común. Sin embargo, existen fenotipos distintivos: el ASPD es el único con somnolencia temprana, mientras que el DSPD presenta despertar tardío como característica específica.

4.2.3. Consulta 2: ¿Qué fármacos están indicados para cada trastorno circadiano?

```

consulta2 <- '
PREFIX crono: <http://cronobiologia.um.es/recurso/>
PREFIX crono_o: <http://cronobiologia.um.es/ontologia/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX biolink: <https://w3id.org/biolink/vocab/>

SELECT ?nombreTrastorno ?nombreFarmaco ?mecanismo ?drugbank
WHERE {
  ?trastorno rdf:type biolink:Disease ;
    rdfs:label ?nombreTrastorno ;
    crono_o:tratadoCon ?farmaco .

  ?farmaco rdfs:label ?nombreFarmaco ;
    rdfs:comment ?mecanismo .

  OPTIONAL {
    ?farmaco rdfs:seeAlso ?drugbank .
    FILTER(CONTAINS(STR(?drugbank), "drugbank"))
  }
}
ORDER BY ?nombreTrastorno ?nombreFarmaco
'

```

La consulta devuelve 7 asociaciones fármaco-trastorno. La melatonina emerge como eje terapéutico central (indicada en 4 de 5 trastornos). Crucialmente, el momento de administración determina su efecto: por la mañana en ASPD para retrasar la fase, por la noche en DSPD para adelantarla. Tasimelteon es específico para Non-24 horas, mientras que Ramelteon se indica para DSPD e Insomnio.

4.2.4. Consulta 3: ¿Qué mutaciones missense patogénicas causan ASPD?

```

consulta3 <- '
PREFIX crono: <http://cronobiologia.um.es/recurso/>
PREFIX crono_o: <http://cronobiologia.um.es/ontologia/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX biolink: <https://w3id.org/biolink/vocab/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?variante ?nombreVariante ?nombreGen ?patogenicidad ?herencia ?clinvar ?descripcion
WHERE {
  ?variante rdf:type biolink:SequenceVariant ;
    rdfs:label ?nombreVariante ;
    rdfs:comment ?descripcion ;
    crono_o:afectaGen ?gen ;
    crono_o:causaTrastorno crono:Trastorno_ASPD ;
    crono_o:tipoMutacion crono:TipoMutacion_Missense .

  ?variante crono_o:tienePatogenicidad ?patogenicidad .
  FILTER(?patogenicidad = "Patogénica"^^xsd:string)

  ?gen rdfs:label ?nombreGen .

  ?variante crono_o:tienePatronHerencia ?herenciaUri .
  ?herenciaUri rdfs:label ?herencia .

  OPTIONAL {
    ?variante rdfs:seeAlso ?clinvar .
    FILTER(CONTAINS(STR(?clinvar), "clinvar"))
  }
}
ORDER BY ?nombreGen ?nombreVariante
'

```

Se recuperan 2 variantes missense patogénicas: **PER2 p.Ser662Gly** (ClinVar:6345) y **CSNK1D p.Thr44Ala** (ClinVar:8741), ambas con herencia autosómica dominante. Ambas están relacionadas con la vía de fosforilación de PER: la primera elimina un sitio de fosforilación, la segunda reduce la actividad de la quinasa fosforiladora.

4.2.5. Consulta 4a: ¿Qué relaciones de interacción física y regulación existen en el TTFL?

```
consulta4a <- '
PREFIX crono: <http://cronobiologia.um.es/recurso/>
PREFIX crono_o: <http://cronobiologia.um.es/ontologia/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?origenLabel ?tipoRelacion ?destinoLabel ?destinoTipo
WHERE {
  {
    # Interacciones físicas proteína-proteína
    ?origen crono_o:interactuaCon ?destino .
    BIND("Interacción física (PPI)" AS ?tipoRelacion)
  }
  UNION
  {
    # Regulación transcripcional positiva
    ?origen crono_o:regulaPositivamente ?destino .
    BIND("Regulación positiva" AS ?tipoRelacion)
  }
  UNION
  {
    # Regulación transcripcional negativa
    ?origen crono_o:regulaNegativamente ?destino .
    BIND("Regulación negativa" AS ?tipoRelacion)
  }
}

?origen rdfs:label ?origenLabel .
?destino rdfs:label ?destinoLabel .
OPTIONAL { ?destino rdf:type ?destinoTipo . }
}
ORDER BY ?tipoRelacion ?origenLabel ?destinoLabel
'
```

La consulta devuelve 52 relaciones clasificadas en: interacciones físicas (CLOCK-BMAL1, PER-CRY, PER-CSNK1D/E), regulación positiva (CLOCK/BMAL1 → genes PER, CRY, NR1D1) y regulación negativa (PER/CRY → proteínas CLOCK/BMAL1, NR1D1 → gen BMAL1). El modelo distingue correctamente entre regulación transcripcional (proteína → gen) y regulación post-traducciona (proteína → proteína).

4.2.6. Consulta 4b: ¿Qué proteínas actúan como hubs regulatorios en la red circadiana?

```

consulta4b <- '
PREFIX crono: <http://cronobiologia.um.es/recurso/>
PREFIX crono_o: <http://cronobiologia.um.es/ontologia/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX biolink: <https://w3id.org/biolink/vocab/>

SELECT ?proteina ?nombreProteina
      (COUNT(DISTINCT ?interactor) AS ?numInteracciones)
      (COUNT(DISTINCT ?genRegulado) AS ?genesRegulados)
      ((?numInteracciones + ?genesRegulados) AS ?gradoCentralidad)
WHERE {
  ?proteina rdf:type biolink:Protein ;
            rdfs:label ?nombreProteina .

  # Contar interacciones físicas proteína-proteína
  OPTIONAL {
    ?proteina crono_o:interactuaCon ?interactor .
  }

  # Contar genes que regula (positiva o negativamente)
  OPTIONAL {
    ?proteina (crono_o:regulaPositivamente|crono_o:regulaNegativamente) ?genRegulado .
  }
}
GROUP BY ?proteina ?nombreProteina
HAVING(?gradoCentralidad > 0)
ORDER BY DESC(?gradoCentralidad)
'

```

BMAL1 emerge como el hub más conectado (centralidad=8), seguido de CLOCK (7). Esto es consistente con la evidencia experimental: BMAL1 es el único componente cuya ausencia causa arritmicidad completa, mientras que otros pueden ser parcialmente compensados.

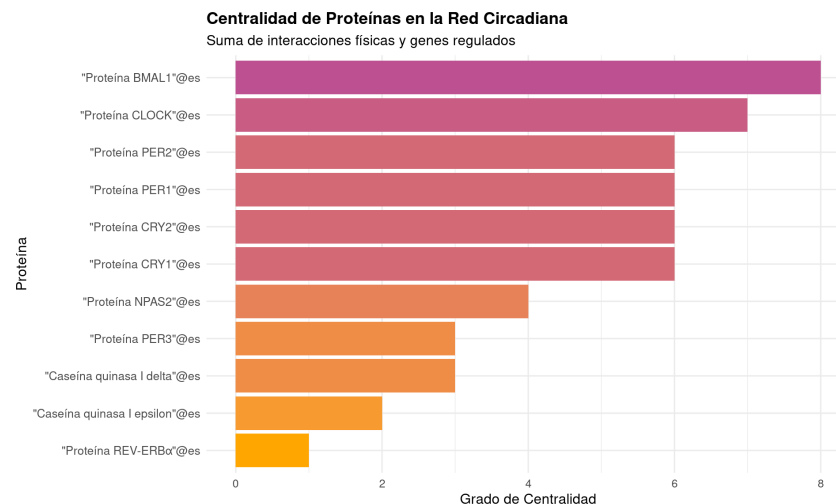


Figura 2: Visualización de la centralidad de red de las proteínas del reloj circadiano.

4.2.7. Consulta 5: ¿Qué nombres oficiales reporta UniProt para las proteínas del grafo?

```

consulta5_federada <- '
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX biolink: <https://w3id.org/biolink/vocab/>
PREFIX up: <http://purl.uniprot.org/core/>

SELECT ?nombreLocal ?uniprotId ?nombreUniprot
WHERE {
  ?proteínaLocal a biolink:Protein ;
    rdfs:label ?nombreLocal ;
    rdfs:seeAlso ?uniprotUri .

  FILTER(CONTAINS(STR(?uniprotUri), "uniprot"))
  BIND(REPLACE(STR(?uniprotUri), "http://purl.uniprot.org/uniprot/", "") AS ?uniprotId)

  SERVICE SILENT <https://sparql.uniprot.org/sparql> {
    ?uniprotUri a up:Protein .
    OPTIONAL { ?uniprotUri up:recommendedName/up:fullName ?nombreUniprot . }
  }
}
ORDER BY ?nombreLocal
'
```

La consulta federada recupera exitosamente los nombres oficiales de las 11 proteínas, validando la correcta anotación del grafo. Por ejemplo: CLOCK (O15516) → “Circadian locomoter output cycles protein kaput”; BMAL1 (O00327) → “Basic helix-loop-helix ARNT-like protein 1”.

5. Conclusiones

El presente trabajo ha demostrado la aplicación práctica de las técnicas de explotación semántica de datos para la representación e integración de conocimiento biomédico en el dominio de la cronobiología y los trastornos del ritmo circadiano del sueño.

Principales logros:

1. **Grafo RDF de alta calidad semántica:** Se ha construido un grafo con 436 tripletas que integra múltiples tipos de entidades biomédicas (genes, proteínas, variantes, trastornos, fenotipos, fármacos) utilizando ontologías y vocabularios estándar.
2. **Consultas SPARQL de complejidad creciente:** Las 7 consultas diseñadas demuestran diferentes capacidades: filtrado múltiple, uso de UNION, agregaciones con GROUP BY, cálculo de métricas de red, y federación con endpoints externos.
3. **Aplicabilidad clínica:** Las consultas permiten responder a preguntas relevantes para el diagnóstico diferencial, la interpretación de variantes y la selección de tratamientos.

En conclusión, este trabajo ilustra cómo los grafos de conocimiento RDF constituyen una herramienta poderosa para la integración de datos biomédicos heterogéneos y su explotación mediante consultas semánticas.

Referencias

- [1] Abbott SM, Reid KJ, Zee PC. Circadian rhythm sleep-wake disorders. *Psychiatr Clin North Am.* 2015;38(4):805-823. doi:10.1016/j.psc.2015.07.001
- [2] Kuhlman SJ, Craig LM, Duffy JF. Introduction to chronobiology. *Cold Spring Harb Perspect Biol.* 2018;10(9):a033613. doi:10.1101/cshperspect.a033613
- [3] Patke A, Young MW, Axelrod S. Molecular mechanisms and physiological importance of circadian rhythms. *Nat Rev Mol Cell Biol.* 2020;21(2):67-84. doi:10.1038/s41580-019-0179-2
- [4] Lowrey PL, Takahashi JS. Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet.* 2004;5:407-441. doi:10.1146/annurev.genom.5.061903.175925
- [5] Mohawk JA, Green CB, Takahashi JS. Central and peripheral circadian clocks in mammals. *Annu Rev Neurosci.* 2012;35:445-462. doi:10.1146/annurev-neuro-060909-153128
- [6] Bunger MK, Wilsbacher LD, Moran SM, et al. Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell.* 2000;103(7):1009-1017. doi:10.1016/S0092-8674(00)00205-1
- [7] DeBruyne JP, Weaver DR, Reppert SM. CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nat Neurosci.* 2007;10(5):543-545. doi:10.1038/nn1884
- [8] Preitner N, Damiola F, Lopez-Molina L, et al. The orphan nuclear receptor REV-ERB α controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell.* 2002;110(2):251-260. doi:10.1016/S0092-8674(02)00825-5
- [9] Xu Y, Toh KL, Jones CR, et al. Modeling of a human circadian mutation yields insights into clock regulation by PER2. *Cell.* 2007;128(1):59-70. doi:10.1016/j.cell.2006.11.043
- [10] Liu C, Tang X, Gong Z, Zeng W, Hou Q, Lu R. Circadian Rhythm Sleep Disorders: Genetics, Mechanisms, and Adverse Effects on Health. *Front Genet.* 2022;13:875342. doi:10.3389/fgene.2022.875342
- [11] Jones CR, Campbell SS, Zone SE, et al. Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. *Nat Med.* 1999;5(9):1062-1065. doi:10.1038/12502
- [12] Micic G, Lovato N, Gradisar M, et al. The etiology of delayed sleep phase disorder. *Sleep Med Rev.* 2016;27:29-38. doi:10.1016/j.smrv.2015.06.004
- [13] Ancoli-Israel S, Gehrman P, Martin JL, et al. Increased light exposure consolidates sleep and strengthens circadian rhythms in severe Alzheimer's disease patients. *Behav Sleep Med.* 2003;1(1):22-36. doi:10.1207/S15402010BSM0101_4
- [14] Lockley SW, Skene DJ, Arendt J, et al. Relationship between melatonin rhythms and visual loss in the blind. *J Clin Endocrinol Metab.* 1997;82(11):3763-3770. doi:10.1210/jcem.82.11.4355
- [15] Levandovski R, Dantas G, Fernandes LC, et al. Depression scores associate with chronotype and social jetlag in a rural population. *Chronobiol Int.* 2011;28(9):771-778. doi:10.3109/07420528.2011.602445
- [16] Xu Y, Padiath QS, Shapiro RE, et al. Functional consequences of a CK1 δ mutation causing familial advanced sleep phase syndrome. *Nature.* 2005;434(7033):640-644. doi:10.1038/nature03453

- [17] Patke A, Murphy PJ, Onat OE, et al. Mutation of the human circadian clock gene CRY1 in familial delayed sleep phase disorder. *Cell*. 2017;169(2):203-215. [doi:10.1016/j.cell.2017.03.027](https://doi.org/10.1016/j.cell.2017.03.027)
- [18] Brennan KC, Bates EA, Shapiro RE, et al. Casein kinase δ mutations in familial migraine and advanced sleep phase. *Sci Transl Med*. 2013;5(183):183ra56. [doi:10.1126/scitranslmed.3005784](https://doi.org/10.1126/scitranslmed.3005784)
- [19] Archer SN, Robilliard DL, Skene DJ, et al. A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep*. 2003;26(4):413-415. [doi:10.1093/sleep/26.4.413](https://doi.org/10.1093/sleep/26.4.413)
- [20] Garaulet M, Corbalán-Tutau MD, Madrid JA, et al. CLOCK gene is implicated in weight reduction in obese patients. *Int J Obes*. 2010;34(3):516-523. [doi:10.1038/ijo.2009.255](https://doi.org/10.1038/ijo.2009.255)
- [21] Auger RR, Burgess HJ, Emens JS, et al. Clinical practice guideline for the treatment of intrinsic circadian rhythm sleep-wake disorders. *J Clin Sleep Med*. 2015;11(10):1199-1236. [doi:10.5664/jcsm.5100](https://doi.org/10.5664/jcsm.5100)
- [22] Buscemi N, Vandermeer B, Hooton N, et al. Efficacy and safety of exogenous melatonin for secondary sleep disorders. *BMJ*. 2006;332(7538):385-393. [doi:10.1136/bmj.38731.532766.F6](https://doi.org/10.1136/bmj.38731.532766.F6)
- [23] Kato K, Hirai K, Nishiyama K, et al. Neurochemical properties of ramelteon (TAK-375), a selective MT1/MT2 receptor agonist. *Neuropharmacology*. 2005;48(2):301-310. [doi:10.1016/j.neuropharm.2004.09.007](https://doi.org/10.1016/j.neuropharm.2004.09.007)
- [24] Lockley SW, Dressman MA, Licamele L, et al. Tasimelteon for non-24-hour sleep-wake disorder in totally blind people. *Lancet*. 2015;386(10005):1754-1764. [doi:10.1016/S0140-6736\(15\)60031-9](https://doi.org/10.1016/S0140-6736(15)60031-9)
- [25] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management. *Sci Data*. 2016;3:160018. [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)