



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural

EVALUACIÓN DEL PAPEL PRONÓSTICO DE LAS PROTEÍNAS PIWI EN CÁNCER MEDIANTE UNA ESTRATEGIA *IN SILICO*

Presentado por **Alejandra Omaira Morcillo Nieto**

Tutor académico: José Javier Forment Millet

Cotutor: Francisco García García

Universitat Politècnica de València
Escola Tècnica Superior d'Enginyeria Agronòmica i del Medi Natural

Trabajo Final de Grado - Biotecnología
Curso académico: 2019/2020
Valencia, junio 2020

Evaluation of PIWI proteins as prognostic biomarker in cancer by *in silico* strategies.

ABSTRACT

Cancer is a group of diseases characterized by uncontrolled growth, proliferation and spread of abnormal cells. Its origin is genetic due to mutations that alter in three genes: oncogenes, tumour suppressor genes and DNA repair genes. Nowadays, cancer is considered the number one cause of mortality in developed countries, the most common is lung cancer, especially non-small cell lung cancer. Discovery and development of new clinical strategies is crucial to improve the detection and stage identification of cancer. The aim of this is to improve the efficacy of the conventional methods. In this field, the role of Bioinformatics is quite relevant, since it allows a better analysis and interpretation of omic data, available in different open-source repositories.

In the present work, the prognostic role of family proteins, called PIWI, is evaluated for cancer detection. Due to the volume of emerging studies correlation PIWI proteins gene expression levels with underlying mechanism of cancer, these proteins are potential biomarkers in cancer. The characterization of these proteins has been performed via *in silico* strategy, based on data review in cancer webtools and databases. Besides, this information has been evaluated and compared to various studies of lung adenocarcinoma. Results suggest the down-regulated expression of PIWIL4 as prognostic biomarker in cancer.

KEYWORDS: cancer, biomarkers, PIWI proteins, lung adenocarcinoma.

Alejandra Omaira Morcillo Nieto

Academic tutor: **Prof. José Javier Forment Millet**

External cotutor: **D. Francisco García García**

Valencia, June 2020

Evaluación del papel pronóstico de las proteínas PIWI en cáncer mediante una estrategia *in silico*.

RESUMEN

El cáncer es un conjunto de enfermedades que se caracteriza por el crecimiento y proliferación descontrolada de un grupo de células y una posterior diseminación a otros tejidos del organismo. Su origen es genético debido a que las alteraciones que contribuyen a la aparición de esta enfermedad afectan principalmente a tres tipos de genes: oncogenes, genes supresores de tumores y genes reparadores de DNA. Además, esta enfermedad se ha posicionado en el primer puesto de causas de mortalidad en países desarrollados, siendo el cáncer de pulmón, en concreto el cáncer de pulmón de células no pequeñas, el más común. Por ello, se requiere el descubrimiento y desarrollo de nuevas estrategias clínicas para mejorar su detección y así determinar su estadio, con el fin de mejorar la eficacia de los tratamientos convencionales. En esta línea de investigación juegan un papel importante las herramientas bioinformáticas, pues facilitan el análisis e interpretación de datos ómicos depositados en los distintos repositorios públicos.

En el presente trabajo, se evalúa el papel pronóstico de una familia de proteínas, denominadas PIWI, en la detección de cáncer. La elección de estas proteínas se basa en estudios emergentes que las muestran como potenciales biomarcadores del cáncer, puesto que se han correlacionado sus niveles de expresión con los diversos mecanismos subyacentes al cáncer. Para realizar esta caracterización de las proteínas PIWI, se ha diseñado y llevado a cabo una estrategia *in silico*, basada en la revisión de información contenida en recursos y bases de datos en cáncer y, también, en la evaluación y comparación de esta información en una selección de estudios de adenocarcinoma de pulmón. Los resultados obtenidos sugieren la infraexpresión de PIWIL4 como un biomarcador pronóstico en varios tipos de cáncer.

PALABRAS CLAVE: cáncer, biomarcador, proteínas PIWI, adenocarcinoma de pulmón.

Alejandra Omaira Morcillo Nieto

Tutor académico: **Prof. José Javier Forment Millet**

Cotutor: **D. Francisco García García**

Valencia, junio 2020

AGRADECIMIENTOS

En primer lugar, me gustaría dar las gracias a Paco por brindarme la oportunidad de comenzar en el mundo de la Bioinformática, al abrirme las puertas de la Unidad de Bioinformática y Bioestadística (UBB) del Centro de Investigación Príncipe Felipe (CIPF). Quiero agradecer su confianza en mí para el desarrollo de este proyecto y por su ayuda y tiempo dedicados en la elaboración del mismo, pero sobre todo por transmitirme en cada reunión su entusiasmo y positivismo, tan necesarios en este periodo de confinamiento.

Agradecer también a mis compañeras Paula, Gloria, Olga, Mar, Inés, Marta y Raquel no solo por estos cuatro años de amistad repletos de experiencias inolvidables, sino también por su ayuda en este trabajo tanto en aspectos técnicos como estéticos, pero sobre todo por aguantar mis nervios y quejas. No puedo olvidar a mis amigas de la infancia Sofía, Elena, Alicia y Andrea, las cuales no han dudado en animarme en las circunstancias más complicadas y en echarme una mano en todo momento, a pesar de que no están acostumbradas a esta rama de conocimiento. Este trabajo no hubiese tenido los mismos resultados sin ellas.

Por último, dar las gracias a mi familia que siempre ha creído en mí y me ha apoyado y aconsejado en cualquier decisión y meta propuestas. En especial, a mi madre y a mi padre, pues fueron los que me animaron a escoger la rama de ciencias. A pesar de que en su momento no fuese mi primera opción, gracias a ellos continué cuatro años después creciendo en este área cada vez con más ilusión. No me olvido de mencionar a mi *chacha* Mari Nieves, mi madrina, la cual me ha dicho una y otra vez que estoy cumpliendo un sueño, su sueño de ir a la universidad y convertirme en una mujer independiente, libre y luchadora.

“El aspecto más triste de la vida es que la ciencia reúne el conocimiento más rápido que la sociedad la sabiduría”

Isaac Asimov (1920 – 1992)

ÍNDICE DE CONTENIDOS

Índice de figuras	II
Índice de tablas	IV
1. INTRODUCCIÓN	1
1.1. Caracterización molecular y epidemiología del cáncer	1
1.2. Papel de los biomarcadores en cáncer	2
1.3. Estudios <i>in silico</i> en Biomedicina	3
1.4. Proteínas PIWI y piRNA en cáncer	4
2. HIPÓTESIS Y OBJETIVOS	6
3. MATERIAL Y MÉTODOS	7
3.1. Revisión actualizada de información clínica y biológica de las proteínas PIWI en recursos y bases de datos de cáncer	7
3.2. Evaluación del papel de las proteínas PIWI en estudios transcriptómicos de adenocarcinoma de pulmón	8
4. RESULTADOS	9
4.1. Revisión actualizada de información clínica y biológica de las proteínas PIWI en recursos y bases de datos de cáncer	9
4.1.1. <i>GEPIA</i>	9
4.1.2. <i>cBioPortal</i>	14
4.1.3. <i>GTEEx Portal</i>	18
4.1.4. <i>The Human Protein Atlas</i>	20
4.2. Evaluación del papel de las proteínas PIWI en estudios transcriptómicos de adenocarcinoma de pulmón	26
4.2.1. <i>GSE10072</i>	26
4.2.2. <i>GSE19188</i>	27
4.2.3. <i>GSE31210</i>	27
4.2.4. <i>GSE32863</i>	27
4.2.5. <i>GSE63459</i>	28
4.2.6. <i>GSE75037</i>	28
4.2.7. <i>GSE81089</i>	28
4.2.8. <i>GSE87340</i>	29
5. DISCUSIÓN	32
6. CONCLUSIONES	35
7. BIBLIOGRAFÍA	36
A. ANEXO	39
I. Figuras	39

ÍNDICE DE FIGURAS

Figura 1.1. Incidencia y mortalidad global del cáncer.	2
Figura 4.1. Expresión diferencial de PIWIL1 en los grupos tumorales descritos en GEPIA.	9
Figura 4.2. Expresión diferencial de PIWIL2 en los grupos tumorales descritos en GEPIA.	10
Figura 4.3. Expresión diferencial de PIWIL3 en los grupos tumorales de GEPIA.....	10
Figura 4.4. Expresión diferencial de PIWIL4 en los grupos tumorales descritos en GEPIA.	11
Figura 4.5. Comparación del conjunto de genes PIWI en los grupos tumorales descritos en GEPIA.	11
Figura 4.6. Curvas de supervivencia de PIWIL1 generadas por GEPIA en los tumores: a) TGCT y b) THCA.....	12
Figura 4.7. Curvas de supervivencia de PIWIL2 generadas por GEPIA en los tumores: a) OV, b) TGCT, c) UCEC y d) UCS.....	12
Figura 4.8. Curvas de supervivencia de PIWIL4 generadas por GEPIA en los tumores: a) BRCA, b) CESC, c) LAMIL, d) LUAL, e) LUSC, f) OV, g) TGCT, h) UCEC e i) UCS.	13
Figura 4.9. Curva de supervivencia de PIWIL1 generada por cBioPortal.	16
Figura 4.10. Curva de supervivencia de PIWIL2 generada por cBioPortal.	16
Figura 4.11. Curva de supervivencia de PIWIL3 generada por cBioPortal.	17
Figura 4.12. Curva de supervivencia de PIWIL4 generada por cBioPortal.	17
Figura 4.13. Expresión del gen PIWIL1 en los diferentes tejidos registrados en GTEX Portal..	18
Figura 4.14. Expresión del gen PIWIL2 en los diferentes tejidos registrados en GTEX Portal..	18
Figura 4.15. Expresión del gen PIWIL3 en los diferentes tejidos registrados en GTEX Portal..	19
Figura 4.16. Expresión del gen PIWIL4 en los diferentes tejidos registrados en GTEX Portal..	19
Figura 4.17. Expresión del RNA y de la proteína PIWIL1 en los diferentes tejidos registrados en THPA.	20
Figura 4.18. Expresión de la proteína PIWIL1 en los diferentes tejidos registrados en THPA..	20
Figura 4.19. Expresión del RNA de PIWIL1 en los diferentes tejidos registrados en THPA. ...	20
Figura 4.20. Expresión del RNA y de la proteína PIWIL2 en los diferentes tejidos registrados en THPA.	21
Figura 4.21. Expresión del RNA de PIWIL2 en los diferentes tejidos registrados en THPA. ...	21
Figura 4.22. Expresión del RNA y de la proteína PIWIL3 en los diferentes tejidos registrados en THPA.	21
Figura 4.23. Expresión del RNA de PIWIL3 en los diferentes tejidos registrados en THPA. ...	22
Figura 4.24. Expresión del RNA y de la proteína PIWIL4 en los diferentes tejidos registrados en THPA.	22
Figura 4.25. Expresión de la proteína PIWIL4 en los diferentes tejidos registrados en THPA..	22
Figura 4.26. Expresión del RNA de PIWIL4 en los diferentes tejidos registrados en THPA. ...	23
Figura 4.27. Expresión de la proteína PIWIL1 en los grupos tumorales descritos en THPA. ...	23
Figura 4.28. Expresión del RNA de PIWIL1 en los grupos tumorales descritos en THPA.	24
Figura 4.29. Expresión del RNA de PIWIL2 en los grupos tumorales descritos en THPA.	24
Figura 4.30. Expresión de la proteína PIWIL3 en los grupos tumorales descritos en THPA. ...	25
Figura 4.31. Expresión del RNA de PIWIL3 en los grupos tumorales descritos en THPA.	25
Figura 4.32. Expresión de la proteína PIWIL4 en los grupos tumorales descritos en THPA. ...	25
Figura 4.33. Expresión del RNA de PIWIL4 en los grupos tumorales descritos en THPA.	26
Figura 4.34. Niveles de expresión de PIWIL1 según la distribución de individuos por sexo y enfermedad.	29
Figura 4.35. Niveles de expresión de PIWIL2 según la distribución de individuos por sexo y enfermedad.	30
Figura 4.36. Niveles de expresión de PIWIL3 según la distribución de individuos por sexo y enfermedad.	30

Figura 4.37. Niveles de expresión de PIWIL4 según la distribución de individuos por sexo y enfermedad.....	31
Figura A.1. Nivel de expresión de las proteínas PIWI en GSE10072.....	39
Figura A.2. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE10072.....	39
Figura A.3. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE10072.....	40
Figura A.4. Nivel de expresión de las proteínas PIWI en GSE19188.....	40
Figura A.5. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE19188.....	40
Figura A.6. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE19188.....	41
Figura A.7. Nivel de expresión de las proteínas PIWI en GSE31210.....	41
Figura A.8. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE31210.....	41
Figura A.9. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE31210.....	42
Figura A.10. Nivel de expresión de las proteínas PIWI en GSE32863.....	42
Figura A.11. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE32863.....	42
Figura A.12. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE32863.....	43
Figura A.13. Nivel de expresión de las proteínas PIWI en GSE63459.....	43
Figura A.14. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE63459.....	43
Figura A.15. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE63459.....	44
Figura A.16. Nivel de expresión de las proteínas PIWI en GSE75037.....	44
Figura A.17. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE75037.....	44
Figura A.18. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE75037.....	45
Figura A.19. Nivel de expresión de las proteínas PIWI en GSE81089.....	45
Figura A.20. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE81089.....	45
Figura A.21. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE81089.....	46
Figura A.22. Nivel de expresión de las proteínas PIWI en GSE87340.....	46
Figura A.23. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE87340.....	46
Figura A.24. Figura 59. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE87340.....	47

ÍNDICE DE TABLAS

Tabla 1.1. Resumen de las proteínas PIWI asociadas al cáncer.....	5
Tabla 3.1. Estudios seleccionados tras la revisión sistemática.	8
Tabla 4.1. Porcentaje de alteración de PIWIL1 en los grupos tumorales descritos en cBioPortal.	14
Tabla 4.2. Porcentaje de alteración de PIWIL2 en los grupos tumorales descritos en cBioPortal.	14
Tabla 4.3. Porcentaje de alteración de PIWIL3 en los grupos tumorales descritos en cBioPortal.	15
Tabla 4.4. Porcentaje de alteración de PIWIL4 en los grupos tumorales descritos en cBioPortal.	15
Tabla 4.5. Individuos incluidos en el estudio GSE10072.....	26
Tabla 4.6. Individuos incluidos en el estudio GSE19288.....	27
Tabla 4.7. Individuos incluidos en el estudio GSE31210.....	27
Tabla 4.8. Individuos incluidos en el estudio GSE32863.....	27
Tabla 4.9. Individuos incluidos en el estudio GSE63459.....	28
Tabla 4.10. Individuos incluidos en el estudio GSE75037.....	28
Tabla 4.11. Individuos incluidos en el estudio GSE81089.....	28
Tabla 4.12. Individuos incluidos en el estudio GSE87340.....	29

GLOSARIO DE ABREVIATURAS

AGO: Argonauta.

BRCA: *Breast invasive carcinoma*, cáncer de mama.

CESC: *Cervical squamous cell carcinoma*, cáncer de células escamosas del cérvix.

COAD: *Colon adenocarcinoma*, adenocarcinoma de colón.

GDC: *Genomic Data Commons*.

GEO: *Gene Expression Omnibus*.

GEPIA: *Gene Expression Profiling Interactive Analysis*.

GTEx Portal: *Genotype-Tissue Expression*.

LAMIL: *Acute Myeloid Leukemia*, leucemia mieloide aguda.

LUAD: *Lung adenocarcinoma*, adenocarcinoma de pulmón.

LUSC: *Lung squamous cell carcinoma*, carcinoma de células escamosas de pulmón.

NSCLC: *non-small cell lung cancer*, cáncer de pulmón de células no pequeñas.

OV: *Ovarian serous cystadenocarcinoma*, cáncer epitelial de ovario.

piRNA: *PIWI-interacting RNAs*.

PIWI: *P-element induced wimpy testis*.

PIWIL1: *piwi like RNA-mediated gene silencing 1*.

PIWIL2: *piwi like RNA-mediated gene silencing 2*.

PIWIL3: *piwi like RNA-mediated gene silencing 3*.

PIWIL4: *piwi like RNA-mediated gene silencing 4*.

RNA-seq: secuenciación de RNA.

READ: *Rectum adenocarcinoma*, adenocarcinoma de recto.

TCGA: *The Cancer Genome Atlas*.

TGCT: *Testicular Germ Cell Tumor*, cáncer de testículo.

THCA: *Thyroid carcinoma*, cáncer de tiroides.

THPA: *The Human Protein Atlas*.

UCEC: *Uterine Corpus Endometrial Carcinoma*, cáncer de endometrio.

UCS: *Uterine Carcinosarcoma*, carcinosarcoma uterino.

1. INTRODUCCIÓN

1.1. Caracterización molecular y epidemiología del cáncer

El cáncer es un conjunto de enfermedades que se caracteriza por un crecimiento excesivo y descontrolado de un grupo de células que invaden y dañan tejidos y órganos. Se trata de un aumento en la proliferación de un grupo de células, lo que provoca la formación de un tumor o neoplasia. Posteriormente, estas células adquieren una capacidad invasiva que les permite migrar desde su lugar de origen a otros tejidos u órganos, un proceso conocido como metástasis (SEOM: Sociedad Española de Oncología Médica, 2019).

El cáncer es considerado una enfermedad genética, puesto que se experimentan cambios en los genes que controlan el funcionamiento de las células, especialmente aquellos que están relacionados con procesos como la proliferación y la diferenciación celular. Estos cambios pueden ser transmitidos por la línea germinal de forma hereditaria u originarse por mutaciones espontáneas, además pueden deberse a cambios en la secuencia de DNA o a cambios epigenéticos que afectan a la expresión de los genes. Los distintos procesos moleculares que contribuyen a la formación y progresión tumoral afectan a tres tipos principales de genes: los oncogenes, los genes supresores de tumores y los genes reparadores de DNA (National Cancer Institute, n.d.; SEOM: Sociedad Española de Oncología Médica, 2019).

La aparición de estos cambios genéticos que desembocan en cáncer, se asocian principalmente a la exposición a factores hormonales, ambientales como carcinógenos físicos (radiaciones ultravioletas e ionizantes), químicos (amoníaco, componentes del tabaco, aflatoxinas, arsénico) y biológicos (determinados virus, bacterias y parásitos) y al envejecimiento, ya que la incidencia de esta enfermedad aumenta con la edad debido a la pérdida de eficacia de los mecanismos de reparación celular, entre otras cosas (World Health Organization, 2018).

De acuerdo con lo estimado por la *World Health Organization* (WHO) en el *Global Cancer Observatory* (GLOBOCAN), en 2018 aparecieron 18,1 millones de nuevos casos de cáncer y fue el responsable de 9,6 millones de muertes lo que le posiciona como la principal causa de muerte en los países desarrollados por delante de las enfermedades cardiovasculares y respiratorias (Bray et al., 2018). Este hecho se ha relacionado con la falta de eficacia de los tratamientos, en aquellos casos en los que la detección de la enfermedad ha sido tardía, los cuales también presentan altas tasas de metástasis y recurrencia (Scully et al., 2012). Por ello, surge la necesidad de nuevas estrategias para la detección de los diferentes tipos de tumores.

Por otro lado, el cáncer más común es el de pulmón puesto que se estima que el 11,6% de los casos diagnosticados en cáncer son de este tipo y, además, es el cáncer más mortal ya que provoca el 18,4% del total de muertes por cáncer (Bray et al., 2018) (Figura 1.1).

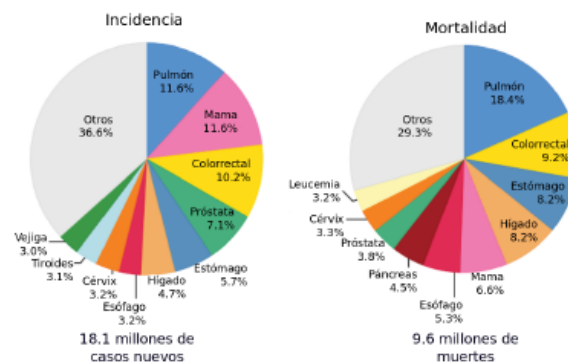


Figura 1.1. Incidencia y mortalidad global del cáncer (Bray et al., 2018).

1.1.1. Cáncer de pulmón

El cáncer de pulmón se divide en dos tipos principales: cáncer de pulmón de células pequeñas (SCLC) y cáncer de pulmón de células no pequeñas (NSCLC). Este último representa un 80% de los casos y se clasifica a su vez en tres subtipos: adenocarcinoma (LUAD), carcinoma de células escamosas (LUSC) y carcinoma de células grandes. El LUAD se ha relacionado con más del 50% de diagnosticados siendo el subtipo más frecuente en cáncer de pulmón (Teramoto et al., 2017).

Además, en LUAD se han observado diferencias en las tasas de incidencia y mortalidad ligadas al sexo, es decir, se considera que el sexo es un posible factor que influye en la supervivencia del cáncer (Kinoshita et al., 2017). A pesar de que en mujeres este subtipo es más frecuente que en hombres, el sexo femenino cuenta con un factor pronóstico favorable ya que presenta mayores tasas de supervivencia con respecto al masculino, independientemente de la etapa del tumor y del consumo de tabaco. Lo que sugiere que estas diferencias puedan estar influenciadas por factores genéticos y hormonales de ambos sexos (Radkiewicz et al., 2019).

Por tanto, una mejor comprensión de los factores ligados al sexo junto con la identificación de nuevos biomarcadores, permitirá una detección temprana que podría lograr un enfoque más personalizado en la terapia y, por consiguiente, mejorar la supervivencia en este tipo de cáncer (Radkiewicz et al., 2019).

1.2. Papel de los biomarcadores en cáncer

El reconocimiento de nuevos biomarcadores es crucial para el diagnóstico y pronóstico del cáncer. Un biomarcador se define como una característica de un sistema biológico que se puede medir y evaluar de forma objetiva, con el fin de obtener un indicador de procesos biológicos normales, patológicos o respuestas farmacológicas a una intervención terapéutica (World Health Organization, 2001). La respuesta medida puede ser funcional, fisiológica o bioquímica, puede ocurrir tanto a nivel celular como molecular y, además, está asociada con la probabilidad del desarrollo de una enfermedad (Arango, 2012). Los biomarcadores se diferencian en marcadores de respuesta pronóstica, predictiva y terapéutica.

En el contexto del cáncer, los marcadores pronósticos permiten predecir el curso natural de un tumor distinguiendo entre indolentes y agresivos; los marcadores predictivos se emplean para evaluar la probabilidad de que un paciente se beneficie de un tratamiento particular; y los marcadores terapéuticos miden los efectos a corto plazo de un tratamiento sobre un tumor (Phan et al., 2009).

En conclusión, los biomarcadores desempeñan un papel fundamental en la mejora del proceso de desarrollo de fármacos. Por lo que comprender la relación entre los procesos biológicos medibles y los resultados clínicos, es vital para expandir los tratamientos para todas las enfermedades y para profundizar nuestra comprensión de la fisiología normal y saludable (Strimbu y Tavel, 2010).

1.3. Estudios *in silico* en Biomedicina

En la búsqueda de nuevos biomarcadores, se requieren herramientas bioinformáticas para el análisis y la interpretación de datos ómicos. Los avances emergentes en las tecnologías de perfiles ómicos permiten el análisis sistemático y la caracterización de genes, RNA, proteínas y metabolitos, ofreciendo de esta forma una posibilidad de descubrir nuevos biomarcadores (Baumgartner et al., 2011). La bioinformática clínica juega un papel clave en el descubrimiento y desarrollo de nuevos diagnósticos y terapias para las enfermedades, combinando la información clínica, la bioinformática y las ciencias ómicas (X. Wang, 2011).

La aplicación de las técnicas de alto rendimiento, como los microarrays y la secuenciación masiva, han impulsado los estudios de diferentes ómicas como abordajes de interés en las áreas biomédicas. Además, su uso ha generado un gran volumen de información de tipo biológico, accesible en diferentes repositorios públicos. Esto ofrece la posibilidad de emplear y combinar diversas aproximaciones analíticas, con el fin de predecir el papel pronóstico de nuevas proteínas (Alandes, 2017; Penttinen et al., 2003).

Entre las distintas bases de datos públicas se encuentra *Gene Expression Omnibus* (GEO), que almacena conjuntos de datos genómicos funcionales de alto rendimiento generados mediante tecnologías basadas en microarrays y secuenciación (Edgar et al., 2002; Wilhite y Barrett, 2012). También existen consorcios del cáncer como *The Cancer Genome Atlas* (TCGA), un programa cuyo objetivo es almacenar datos genómicos, epigenómicos, transcriptómicos y proteómicos de miles de muestras tumorales (National Cancer Institute, 2006); y *Genomic Data Commons* (GDC), este proyecto consolidó todos los datos clínicos y genómicos de TCGA y otros programas de *The National Cancer Institute* (NCI) (Jensen et al., 2017).

Sin embargo, debido al libre acceso de datos científicos, surgió la necesidad de una serie de buenas prácticas para la publicación de datos que fuesen específicas y ampliamente compartidas y aplicadas. Por ello, el 15 de marzo de 2016 se publicó en la revista *Scientific Data* de *Nature* el artículo “*The FAIR Guiding Principles for scientific data management and stewardship*”. Estos principios ofrecen un conjunto de cualidades precisas y medibles que toda publicación de datos debe seguir: los datos deben ser encontrables (*Findable*), accesibles, (*Accesible*), interoperables (*Interoperable*) y reutilizables (*Reusable*) (Wilkinson, 2016).

Toda esta información *in silico* proporciona a la comunidad de investigación del cáncer un repositorio unificado. De esta forma, se permite el intercambio de datos entre los distintos estudios ómicos del cáncer en apoyo a una medicina personalizada y de precisión (Grossman et al., 2016).

1.4. Proteínas PIWI y piRNA en cáncer

Actualmente, un número emergente de estudios muestran el papel de piRNA y las proteínas PIWI como potenciales biomarcadores en el pronóstico del cáncer (Liu et al., 2019).

Las proteínas PIWI fueron descritas originalmente en *Drosophila* y su denominación proviene de la abreviación de *P-element Induced Wimpy testis*. Pertenecen a la familia de proteínas Argonauta junto con la subfamilia Argonauta (AGO). Tienen tres motivos proteicos principales: los dominios PAZ (Piwi-Argonaute-Zwile), MID y PIWI. El dominio PAZ es un motivo de reconocimiento al extremo 3' del ácido nucleico guía, mientras que el dominio MID es el responsable del reconocimiento 5' y el dominio PIWI es un dominio RNasaH con actividad endonucleasa (Dayeh et al., 2018; Seto et al., 2007). Estas proteínas juegan un papel importante en el mantenimiento y autorrenovación de las células madre, la interferencia con RNA y la regulación de la traducción de proteínas (Li et al., 2010).

Esta familia de genes está formada por cuatro miembros *PIWIL1* (*piwi like RNA-mediated gene silencing 1*), *PIWIL2* (*piwi like RNA-mediated gene silencing 2*), *PIWIL3* (*piwi like RNA-mediated gene silencing 3*) y *PIWIL4* (*piwi like RNA-mediated gene silencing 4*), cuyos homólogos en humano se denominan *HIWI*, *HILI*, *PIWIL3* y *HIWI2*, respectivamente. Solo se han identificado en animales y se expresan fundamentalmente en células de la línea germinal. Son proteínas nucleares que cuentan con funciones importantes, como el silenciamiento de retrotransposones y el control de la línea germinal masculina, además de la espermatogénesis (Han et al., 2017; Li et al., 2010).

Además, estas proteínas son capaces de interactuar con unos pequeños RNA no codificantes de 21-35 nucleótidos conocidos como piRNA (*PIWI-interacting RNA*), distintos de los microRNA (miRNA) y RNA interferentes (siRNA) que interactúan con las proteínas AGO. Mientras que los miRNA y siRNA derivan de precursores de RNA bicatenarios, los piRNA se procesan a partir de largas cadenas de RNA monocatenario. Estos últimos, son los encargados de guiar a las proteínas PIWI para escindir el RNA objetivo, formando el complejo piRNA/PIWI (Ozata et al., 2019).

Los piRNA desempeñan diversas funciones, entre las que se encuentran el silenciamiento del transposón dependiente de PIWI, la regulación epigenética, la regulación de genes y proteínas, el reordenamiento genómico y el mantenimiento de las células madre germinales (Aravin et al., 2007). También desempeñan un papel fundamental en la reproducción y la regulación de la fertilidad (Cox et al., 2000).

En estudios previos, se ha demostrado una relación entre la expresión aberrante de los piRNA y los miembros de la familia PIWI en el desarrollo del cáncer. Estos niveles diferenciales están involucrados en la proliferación de células cancerosas, apoptosis, metástasis, invasión y pueden ser potenciales biomarcadores de diagnóstico y pronóstico durante el desarrollo del cáncer (Han et al., 2017) (Tabla 1.1).

Tabla 1.1. Resumen de las proteínas PIWI asociadas al cáncer (Han et al., 2017).

<i>PIWI</i>	<i>CÁNCER</i>	<i>EXPRESIÓN</i>	<i>FUNCIÓN</i>
<i>PIWIL1/HIWI</i>	Seminoma	Alta	Promueve la proliferación celular
	Cáncer de estómago	Alta	Promueve la proliferación celular. Relacionado con un mal pronóstico
	Sarcoma	Alta	Promueve la proliferación celular
	Cáncer de pulmón	Alta	Promueve la proliferación celular
	Cáncer de mama	Alta	Promueve la proliferación celular
	Glioma	Alta	Promueve la proliferación celular, metástasis y migración
	Carcinoma hepatocelular	Baja	Reduce la proliferación celular y migración. Relacionado con pronóstico
	Cáncer epitelial de ovario	Baja	Reprime la invasión celular
	Cáncer colorrectal	Alta	Inhíbe la migración e invasión. Relacionado con un mal pronóstico sin metástasis en nódulos linfáticos
	Cáncer de endometrio	Baja	Relacionado con hipermetilación de DNA
<i>PIWIL2/HILI</i>	Sarcoma de tejidos blandos	Baja	Reduce la proliferación celular y aumento de apoptosis
	Cáncer severo	Alta	Promueve la supervivencia y proliferación de las células tumorales
	Carcinoma de células escamosas del cérvix	Alta	Inhíbe la apoptosis
	Cáncer de colon	Alta	Promueve la metástasis
	Cáncer de mama	Alta	Promueve la proliferación celular
	Cáncer colorrectal	Baja	Inhíbe la migración e invasión. Relacionado con un mal pronóstico
	Cáncer de estómago	Alta	Relacionado con un mal pronóstico
	Cáncer de esófago	Alta	Relacionado con un mal pronóstico
<i>PIWIL3</i>	Cáncer de mama	Alta	Importante en pronóstico
<i>PIWIL4/HIWI2</i>	Cáncer de cérvix	Alta	Afecta a la apoptosis y promueve la invasión
	Cáncer de mama	Alta	Importante en pronóstico

También, se ha comprobado que la expresión de las proteínas PIWI se correlaciona negativamente con la supervivencia de pacientes. La inhibición de estas proteínas podría reducir el número de células en fase G2/M y mejorar la expresión de la proteína p53, inhibiendo así la proliferación y promoviendo la apoptosis de las células cancerosas (Lu et al., 2012). Además, las proteínas PIWI aumentan la resistencia de las células frente a la quimioterapia. Por tanto, una disminución en sus niveles puede aumentar la sensibilidad al tratamiento proporcionando potenciales estrategias para superar la quimiorresistencia de los pacientes (Wang et al., 2011).

A pesar de la existencia de muchos estudios sobre el patrón de expresión de piRNA y las proteínas PIWI en la tumorigénesis, no está claro si las proteínas PIWI regulan estos mecanismos subyacentes al cáncer de forma independiente, o gracias a la interacción con piRNA. Es por ello por lo que nace la necesidad de seguir explorando el complejo piRNA/PIWI (Li et al., 2010).

2. HIPÓTESIS Y OBJETIVOS

El objetivo de este trabajo es explorar y caracterizar el papel pronóstico de las proteínas PIWI mediante una estrategia *in silico* basada en la revisión de estudios ómicos en cáncer.

Para lograrlo, se llevarán a cabo los siguientes objetivos específicos:

- Búsqueda y estudio de la información descrita para esta familia de genes PIWI en humano: *PIWIL1* o *HIWI*, *PIWIL2* o *HILI*, *PIWIL3* y *PIWIL4* o *HIWI2*, en diferentes recursos y bases de datos sobre cáncer.
- Diseño y aplicación de una estrategia *in silico*, sobre una selección de estudios transcriptómicos de adenocarcinoma de pulmón, para evaluar la información obtenida en el primer objetivo específico.

3. MATERIAL Y MÉTODOS

La metodología de investigación en este estudio de caracterización de las proteínas PIWI en cáncer se realizará mediante i) la revisión actual de información clínica y biológica de estas proteínas en los principales recursos y bases de datos de cáncer y ii) evaluación de esta información en una selección de estudios de adenocarcinoma de pulmón, procedentes de una revisión sistémica.

3.1.Revisión actualizada de información clínica y biológica de las proteínas PIWI en recursos y bases de datos de cáncer

En primer lugar, es de interés reunir todo el conocimiento actual de esta familia de proteínas y su relación en los diferentes tipos de cáncer. Para llevar a cabo esta revisión se han empleado las siguientes bases de datos: *Gene Expression Profiling Interactive Analysis* (GEPIA), *cBioPortal*, *Genotype-Tissue Expression* (GTEx) y *The Human Protein Atlas* (THPA).

Por un lado, GEPIA es un servidor web interactivo que se ha desarrollado para analizar los datos de secuenciación de RNA (RNA-seq) de tumores y muestras normales de los proyectos TCGA y GTEx. Esta herramienta proporciona funciones personalizadas, como el análisis de expresión diferencial de tumor/normal, el perfil según los tipos de cáncer o etapas patológicas y el análisis de supervivencia del paciente, entre otras (Tang et al., 2017).

En el caso del servidor *cBioPortal*, este permite la exploración interactiva de conjuntos de datos genómicos multidimensionales sobre el cáncer, cuyo objetivo es proporcionar un acceso de alta calidad a los perfiles moleculares y los atributos clínicos de los proyectos de genómica del cáncer a gran escala (Cerami et al., 2012; Gao et al., 2013).

Por otro lado, la finalidad del proyecto GTEx es estudiar la expresión y regulación de genes específicos de tejidos libres de patologías (The GTEx Consortium, 2015).

Por último, THPA es un programa cuyo objetivo es mapear todas las proteínas humanas en células, tejidos y órganos mediante la integración de diversas tecnologías ómicas, como la proteómica, transcriptómica y biología de sistemas. Esta herramienta muestra la distribución de las proteínas a través de todos los tejidos y órganos principales del cuerpo humano de pacientes con cáncer (Uhlen et al., 2015; Uhlen et al., 2017).

Con el empleo de estas bases de datos, se pretende observar en qué tipos de tumores hay una expresión diferencial del conjunto de proteínas y si este nivel de expresión altera la tasa de supervivencia de los pacientes oncológicos.

3.2. Evaluación del papel de las proteínas PIWI en estudios transcriptómicos de adenocarcinoma de pulmón

Tras revisar el conocimiento actual del papel de las proteínas PIWI en los diferentes grupos tumorales, a partir de la revisión de los recursos y bases de datos en cáncer, se diseñó una estrategia *in silico* para conocer si esta información correspondía o podía ser ampliada en un determinado tipo de cáncer. Para ello, se utilizó una revisión sistemática y selección de ocho estudios transcriptómicos de adenocarcinoma de pulmón, donde disponíamos de los niveles de expresión génica en pacientes y controles de ambos sexos (Pérez, 2019).

El objetivo de la selección de estos estudios es evaluar las diferencias entre los niveles de expresión de la familia de proteínas PIWI y comprobar si estos niveles concuerdan con la caracterización global realizada en el apartado anterior. Además, también se pretende analizar si las diferencias observadas pueden deberse a algún factor ligado al sexo.

Tabla 3.1. Estudios seleccionados tras la revisión sistemática (Pérez, 2019).

<i>Estudio</i>	<i>Plataforma</i>
GSE10072	Affymetrix Human Genome U133A Array
GSE19188	Affymetrix Human Genome U133 Plus 2.0 Array
GSE31210	Affymetrix Human Genome U133 Plus 2.0 Array
GSE32863	Illumina HumanWG-6 v3.0 expression beadchip
GSE63459	Illumina HumanRef-8 v3.0 expression beadchip
GSE75037	Illumina HumanWG-6 v3.0 expression beadchip
GSE81089	Illumina HiSeq 2500
GSE87340	Illumina HiSeq 2000

Para la recopilación de estos estudios, hay que tener en cuenta que una revisión sistemática es una metodología que permite comparar toda la evidencia empírica que cumpla ciertos criterios de elegibilidad (Liberati et al., 2009). De esta forma, se tuvo que realizar previamente una búsqueda exhaustiva en las bases de datos públicas GEO y GDC sobre adenocarcinoma de pulmón. Se seleccionaron aquellos ensayos que contenían información sobre la expresión génica, el sexo, el estadio de la enfermedad y el hábito tabáquico de los pacientes. Los datos de estos estudios fueron procesados y sometidos a un análisis de expresión diferencial y de enriquecimiento funcional tanto en términos *Gene Ontology* como en rutas KEGG (*Kyoto Encyclopedia of Genes and Genomes*) y, finalmente, a un metaanálisis (Pérez, 2019).

En la evaluación que realizamos de estos estudios, nos centramos en los niveles de expresión correspondientes a las proteínas de interés. En el procesamiento de los datos, se ha comprobado que todas las medidas de expresión estuviesen en escala logarítmica y que no incluyesen valores negativos. Por ello, en el caso del estudio GSE31210, se ha tenido que realizar una transformación logarítmica en base 2; y en los estudios GSE19188, GSE81089 y GSE87320 se ha llevado a cabo un re-escalado, para que todas las medidas tomasen valores positivos.

Para evaluar la expresión diferencial de estos genes en las muestras disponibles, se ha llevado a cabo un análisis estadístico descriptivo en el que se han calculado la media y la desviación típica de los distintos estudios para cada gen y, además, se han generado unos gráficos de barras que muestran la expresión diferencial de los genes en los distintos individuos según enfermedad y sexo. Para realizar estos cálculos se ha empleado Microsoft Office Excel y el lenguaje de programación R (R Core Team, 2018).

4. RESULTADOS

En este apartado se muestran los resultados obtenidos, siguiendo el orden establecido en el apartado anterior.

4.1.Revisión actualizada de información clínica y biológica de las proteínas PIWI en recursos y bases de datos de cáncer

4.1.1. GEPIA

En la base de datos GEPIA, se han obtenido los perfiles de expresión individuales de cada gen, la expresión diferencial del conjunto de genes y, además, las curvas de supervivencia de aquellos tipos de cáncer que tienen una expresión significativa de cada gen.

En las Figuras 4.1 a 4.4, se observan los perfiles de expresión de cada gen en los distintos tumores y también aparecen destacados aquellos tipos de cáncer en los que los niveles de expresión de las proteínas son significativos. Para ello, GEPIA compara los datos procedentes de TCGA, como muestra enferma, y de GTEx Portal, como muestra control.

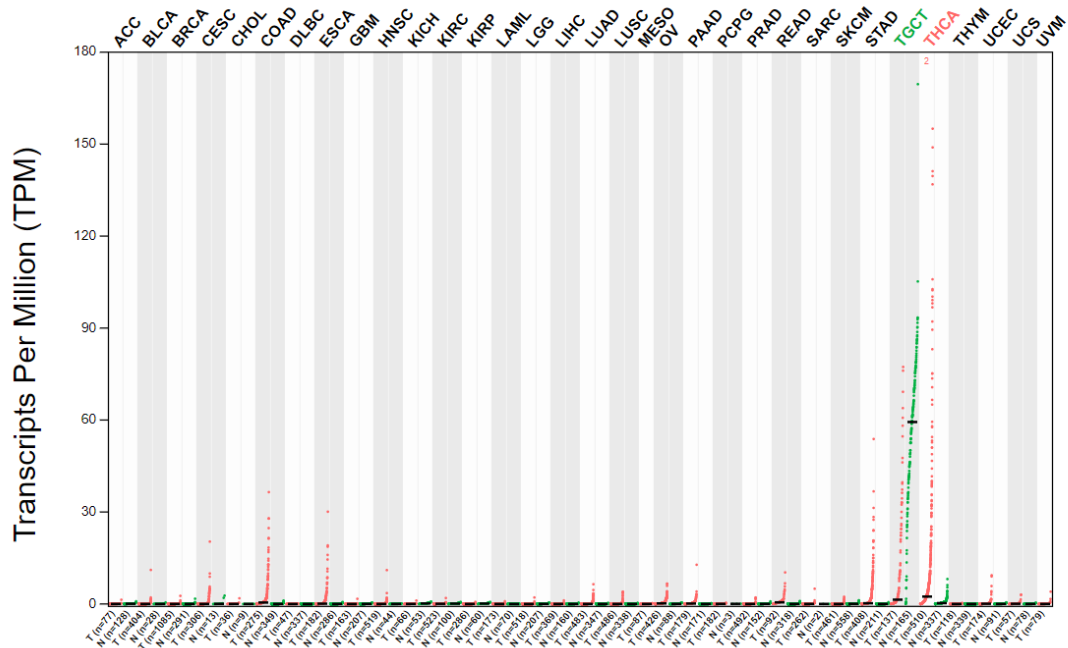


Figura 4.1. Expresión diferencial de PIWIL1 en los grupos tumorales descritos en GEPIA.

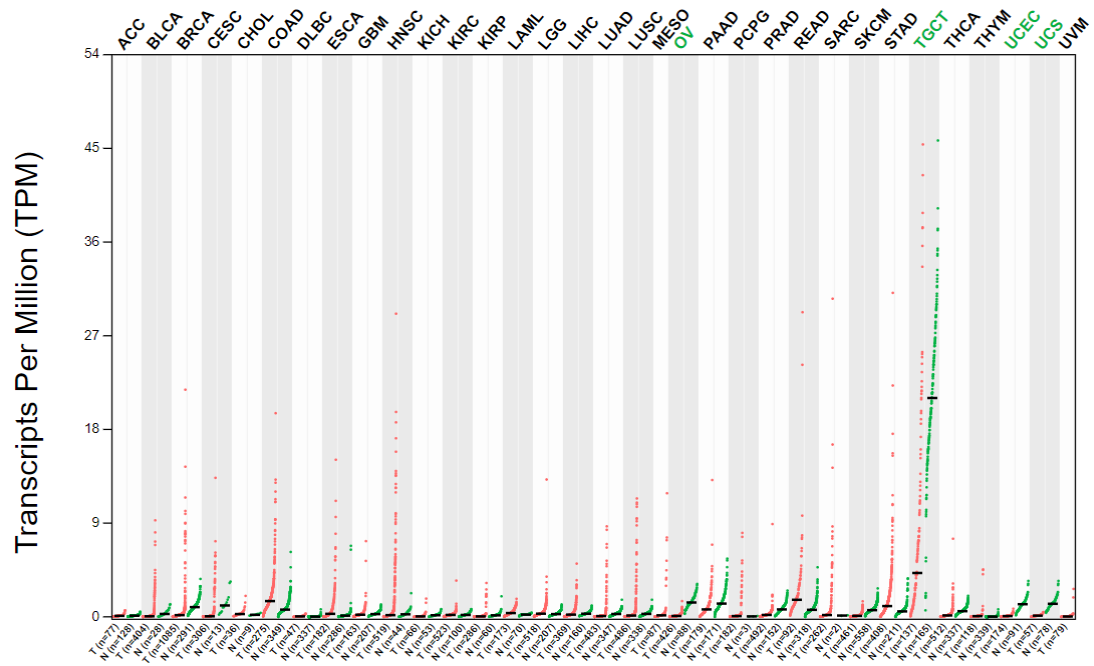


Figura 4.2. Expresión diferencial de PIWIL2 en los grupos tumorales descritos en GEPIA.



Figura 4.3. Expresión diferencial de PIWIL3 en los grupos tumorales de GEPIA.

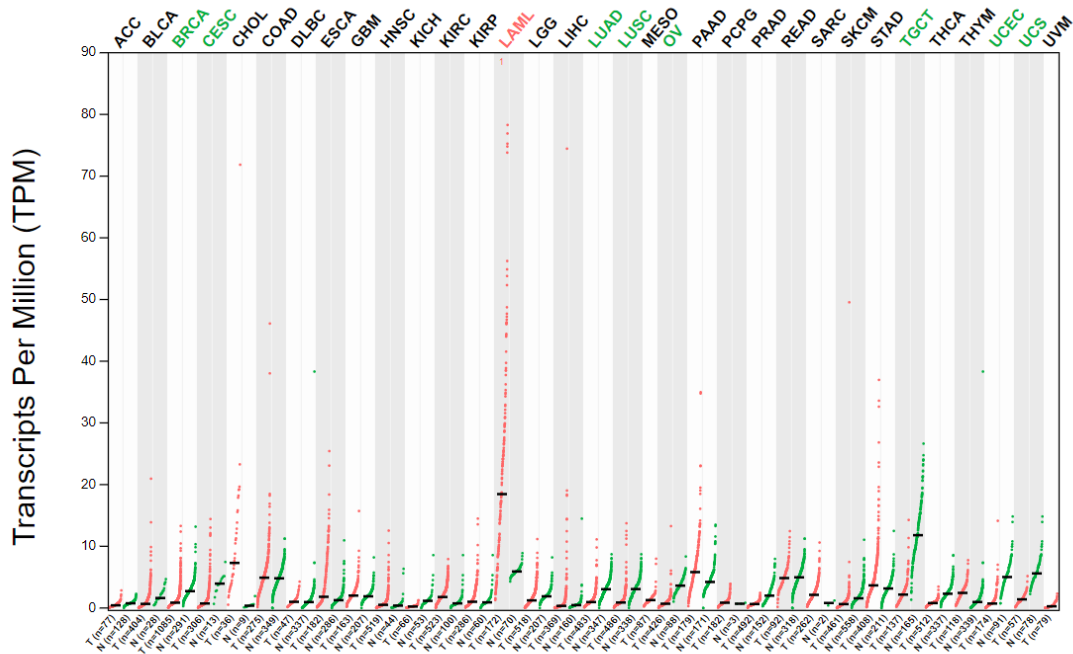


Figura 4.4. Expresión diferencial de PIWIL4 en los grupos tumorales descritos en GEPIA.

Por otro lado, el heatmap de la Figura 4.5 representa los niveles de expresión de cada gen en los grupos tumorales descritos. La intensidad del color está asociada a los niveles de expresión de cada gen.

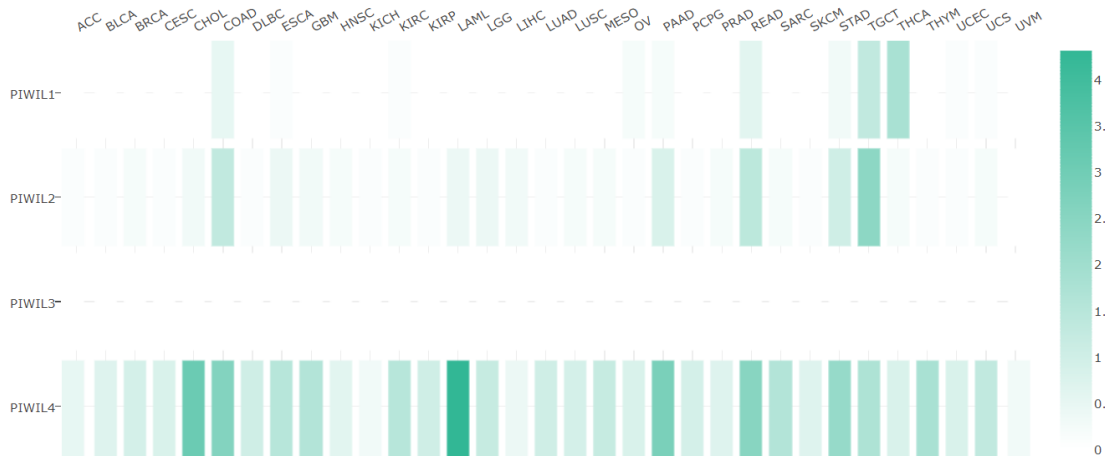


Figura 4.5. Comparación del conjunto de genes PIWI en los grupos tumorales descritos en GEPIA.

Además, atendiendo a los resultados obtenidos en los perfiles de expresión diferencial (Figuras 4.1 a 4.4), se han analizado aquellos tipos de cáncer con niveles diferenciales significativos. En las Figuras 4.6 a 4.8, se comparan las curvas de supervivencia entre los pacientes enfermos con alta y baja expresión en cada uno de los genes de interés mediante el análisis de supervivencia general (OS del inglés, *Overall Survival*). A continuación, se aplica el test Log-rank para comprobar la existencia de diferencias estadísticamente significativas entre ambos grupos.

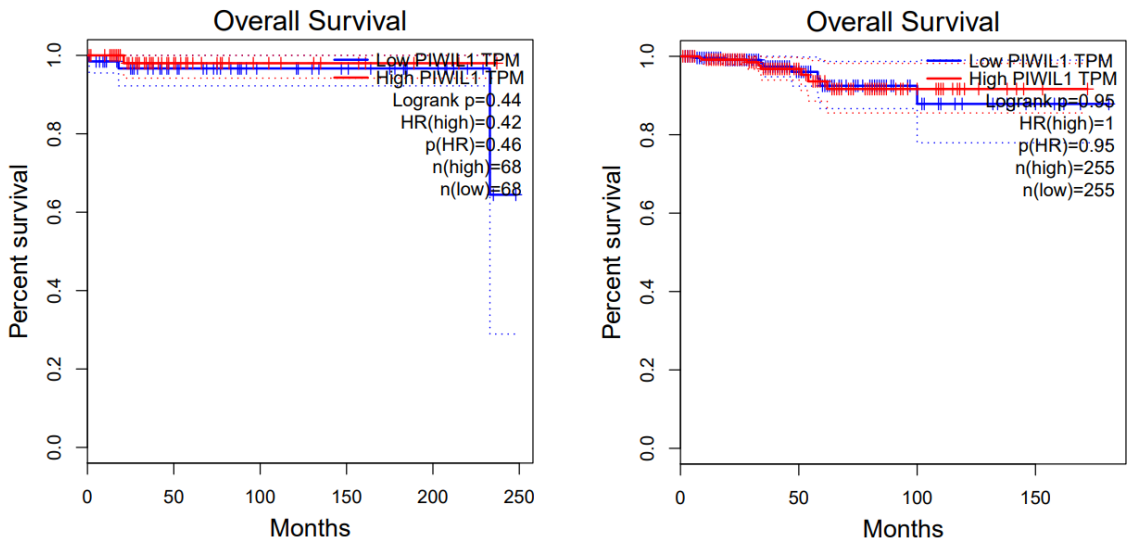


Figura 4.6. Curvas de supervivencia de PIWIL1 generadas por GEPIA en los tumores: a) TGCT y b) THCA.

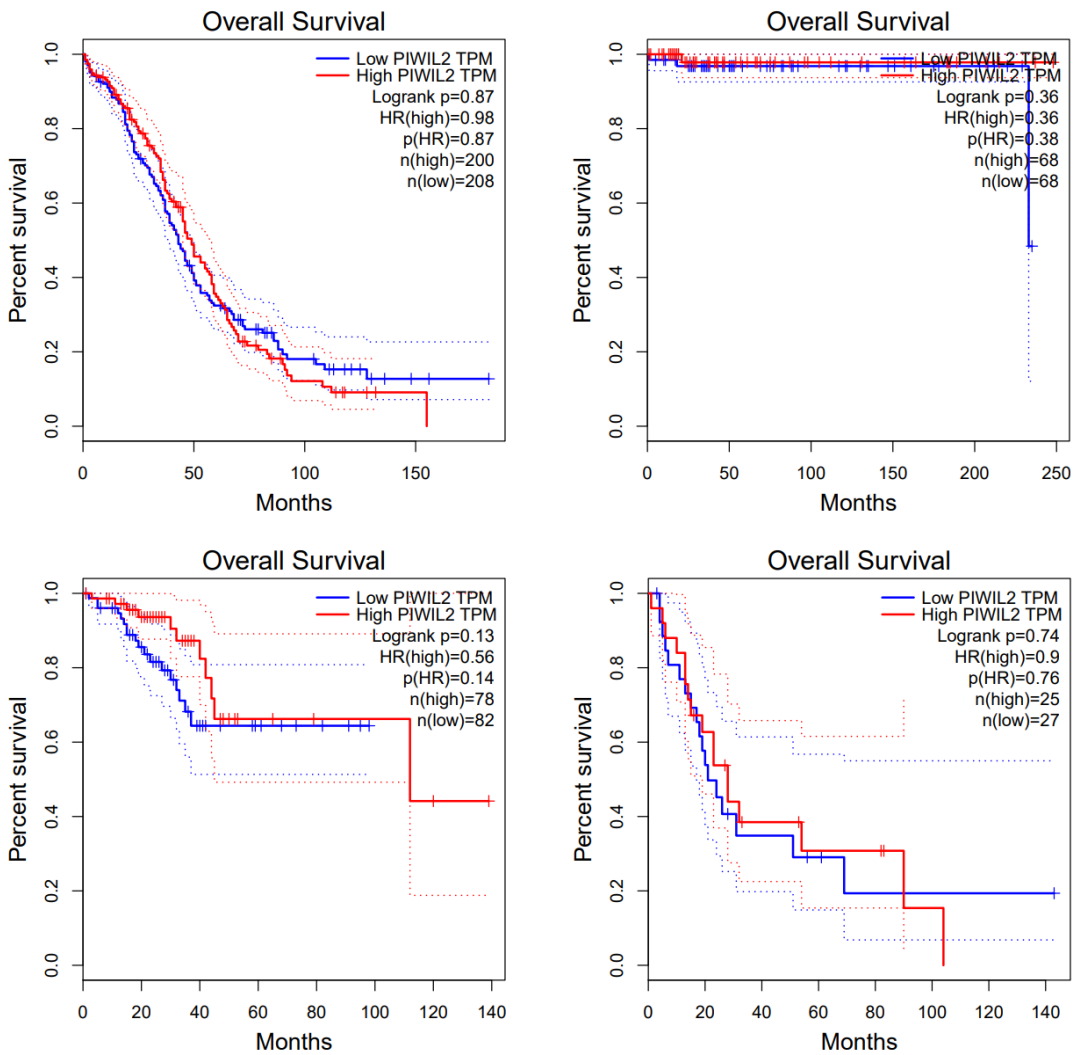


Figura 4.7. Curvas de supervivencia de PIWIL2 generadas por GEPIA en los tumores: a) OV, b) TGCT, c) UCEC y d) UCS.

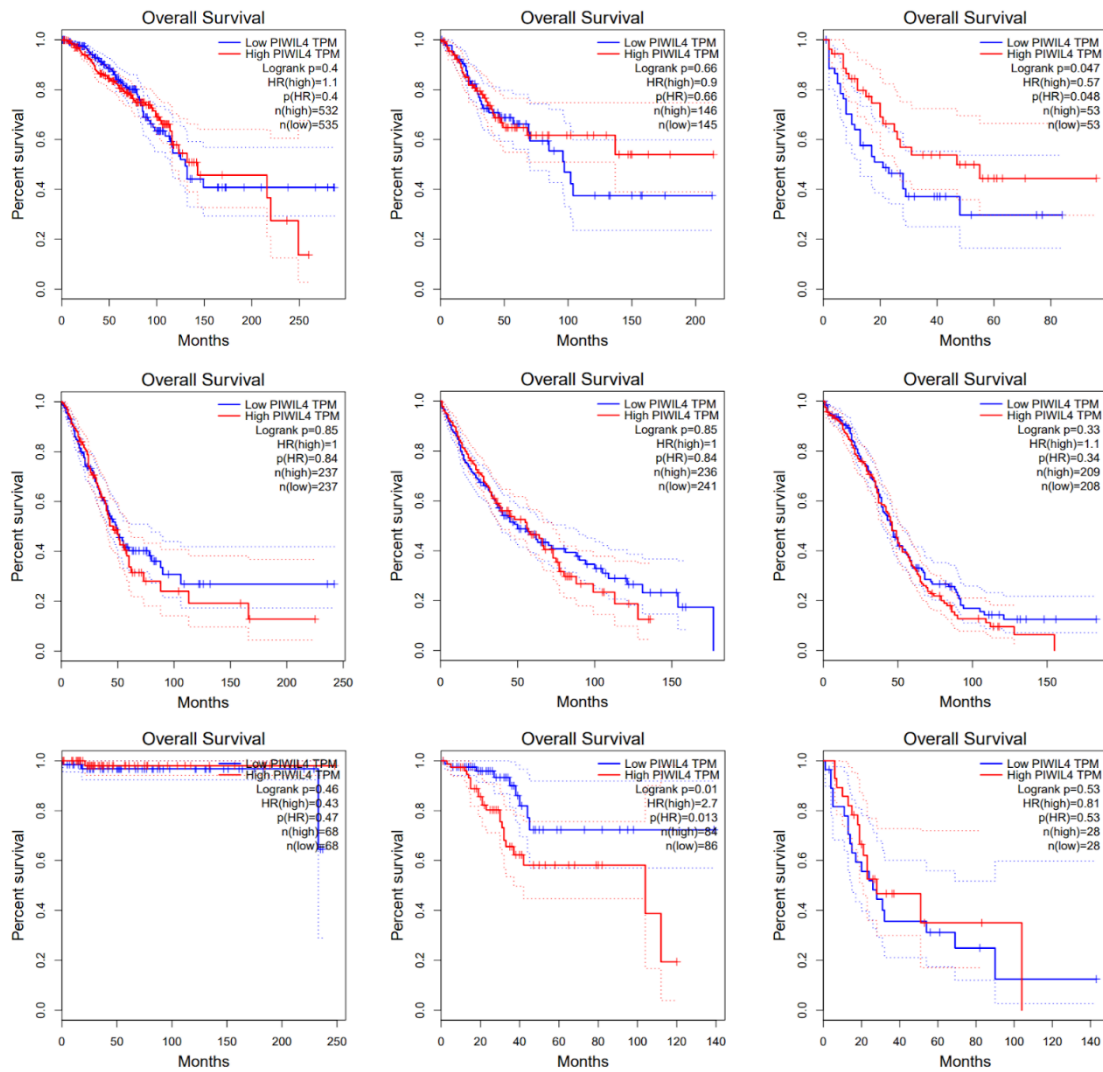


Figura 4.8. Curvas de supervivencia de PIWIL4 generadas por GEPIA en los tumores: a) BRCA, b) CESC, c) LAMIL, d) LUAL, e) LUSC, f) OV, g) TGCT, h) UCEC e i) UCS.

4.1.2. cBioPortal

Desde *cBioPortal*, se han obtenido las alteraciones más frecuentes para cada uno de los genes de interés en los diferentes tumores. Esta información se detalla en las siguientes tablas.

Tabla 4.1. Porcentaje de alteración de PIWIL1 en los grupos tumorales descritos en cBioPortal.

	<i>Casos totales</i>	<i>Porcentaje de alteración</i>	<i>Mutación</i>	<i>Amplificación</i>	<i>Deleción</i>
<i>Melanoma</i>	444	8,78%	8,11% (36 casos)	0,23% (1 caso)	0,45% (2 casos)
<i>Cáncer de endometrio</i>	586	7,85%	7,17% (42 casos)	0,51% (3 casos)	0,17% (1 caso)
<i>Linfoma de células B</i>	48	6,25%	6,25% (3 casos)		
<i>Cáncer de pulmón de células no pequeñas</i>	1053	4,65%	3,89% (41 casos)	0,28% (41 casos)	0,47% (5 casos)
<i>Adenocarcinoma colorrectal</i>	594	4,21%	4,04% (24 casos)	0,17% (1 caso)	

Tabla 4.2. Porcentaje de alteración de PIWIL2 en los grupos tumorales descritos en cBioPortal.

	<i>Casos totales</i>	<i>Porcentaje de alteración</i>	<i>Mutación</i>	<i>Amplificación</i>	<i>Deleción</i>	<i>Otras</i>
<i>Cáncer de endometrio</i>	586	9,90%	6,31% (37 casos)	0,51% (3 casos)	2,90% (17 casos)	0,17% (1 caso) Múltiples
<i>Carcinoma urotelial</i>	411	8,27%	1,95% (8 casos)	0,49% (2 casos)	5,84% (24 casos)	
<i>Adenocarcinoma de próstata</i>	494	8,10%	0,40% (2 casos)		7,49% (37 casos)	0,20% (1 caso) Múltiples
<i>Adenocarcinoma colorrectal</i>	594	7,74%	2,69% (16 casos)		5,05% (30 casos)	
<i>Melanoma</i>	444	7,43%	6,98% (31 casos)		0,45% (2 casos)	
<i>Carcinoma hepatocelular</i>	369	6,78%	0,54% (2 casos)		5,96% (22 casos)	0,27% (1 caso) Múltiples
<i>Cáncer de pulmón de células no pequeñas</i>	1053	6,74%	1,90% (20 casos)	0,09% (1 caso)	4,65% (49 casos)	0,09% (1 caso) Múltiples
<i>Cáncer epitelial de ovario</i>	584	6,51%	0,34% (2 casos)		5,99% (35 casos)	0,17% (1 caso) Fusión

Tabla 4.3. Porcentaje de alteración de PIWIL3 en los grupos tumorales descritos en cBioPortal.

	<i>Casos totales</i>	<i>Porcentaje de alteración</i>	<i>Mutación</i>	<i>Amplificación</i>	<i>Delección</i>
<i>Cáncer de endometrio</i>	586	7,34%	6,48% (38 casos)	0,85% (5 casos)	
<i>Melanoma</i>	444	5,86%	4,73% (21 casos)	0,90% (4 casos)	0,23% (1 caso)
<i>Carcinoma adrenocortical</i>	91	4,40%	2,20% (2 casos)	1,10% (1 caso)	1,10% (1 caso)

Tabla 4.4. Porcentaje de alteración de PIWIL4 en los grupos tumorales descritos en cBioPortal.

	<i>Casos totales</i>	<i>Porcentaje de alteración</i>	<i>Mutación</i>	<i>Amplificación</i>	<i>Delección</i>	<i>Otras</i>
<i>Cáncer de endometrio</i>	586	7,17%	6,31% (37 casos)	0,51% (3 casos)	0,34% (2 casos)	
<i>Melanoma</i>	444	6,98%	4,28% (19 casos)	1,35% (6 casos)	1,13% (5 casos)	0,23% (1 caso)
<i>Cáncer de células escamosas del cérvix</i>	251	4,78%	1,58% (4 casos)	2,39% (6 casos)	0,8% (2 casos)	

Además de contemplar qué alteraciones son las más frecuentes, se puede observar la tasa de supervivencia en aquellos pacientes oncológicos que sufren algún tipo de alteración en estos genes, con respecto a aquellos que no tienen ninguna alteración en la secuencia genética.

En las Figuras 4.9 a 4.12, se muestran las curvas de supervivencia para los grupos que presentan la proteína alterada frente a los que no. El servidor *cBioPortal* también emplea el test Log-Rank para evaluar si las diferencias de supervivencia son estadísticamente significativas entre ambos grupos.

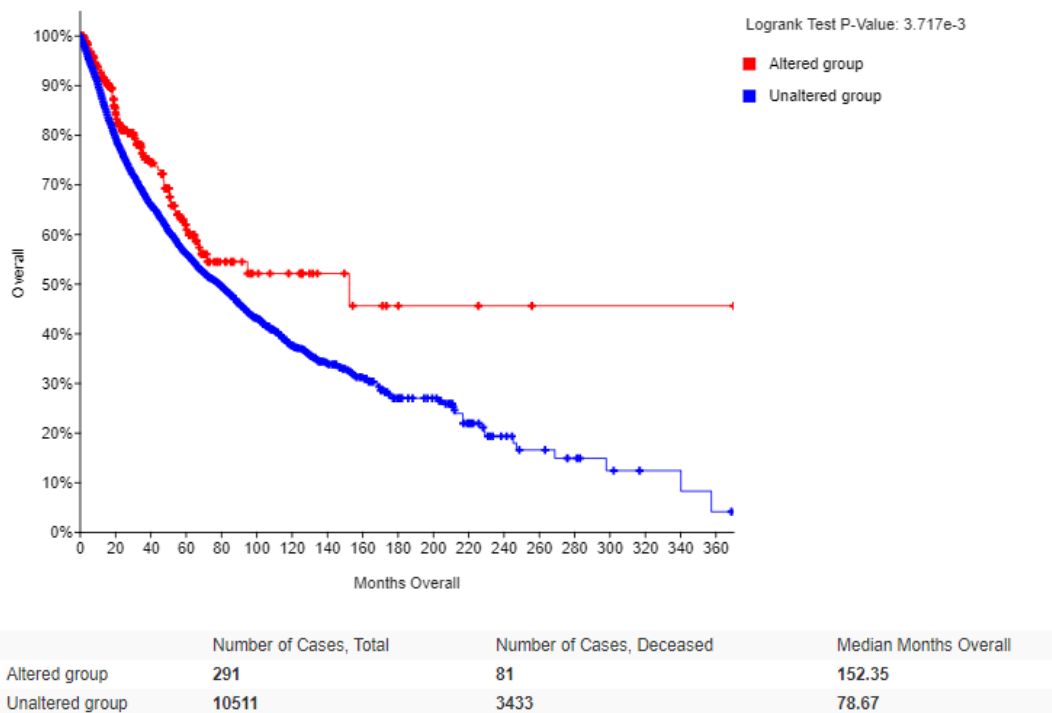


Figura 4.9. Curva de supervivencia de PIWIL1 generada por cBioPortal.

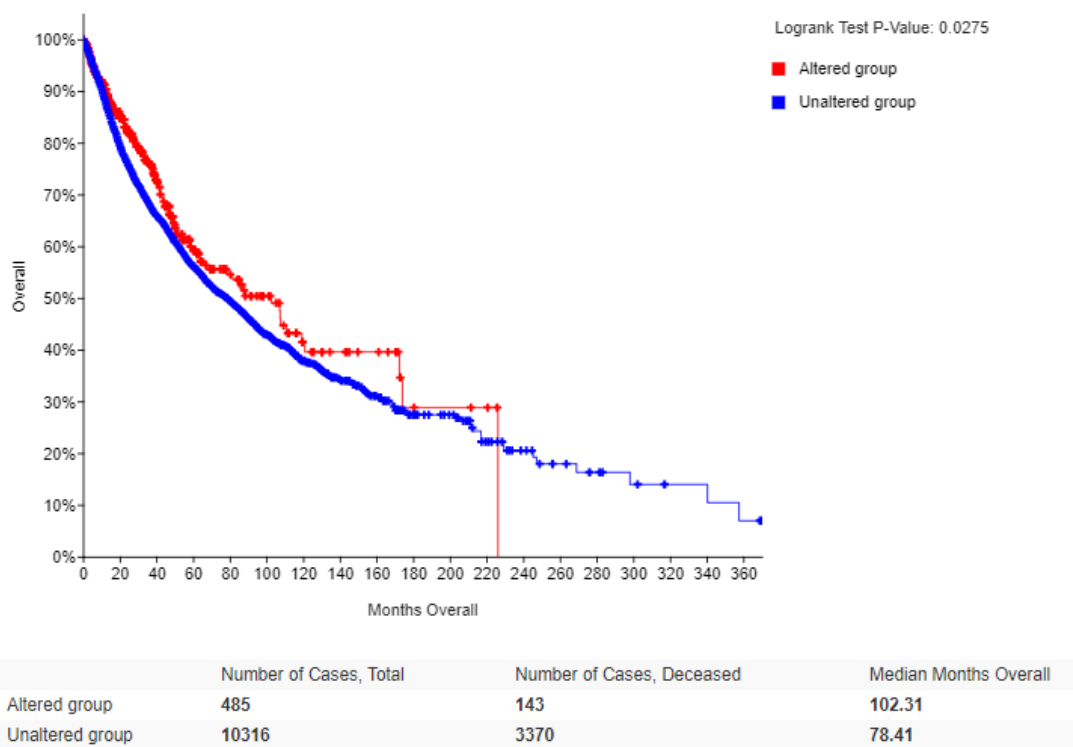


Figura 4.10. Curva de supervivencia de PIWIL2 generada por cBioPortal.

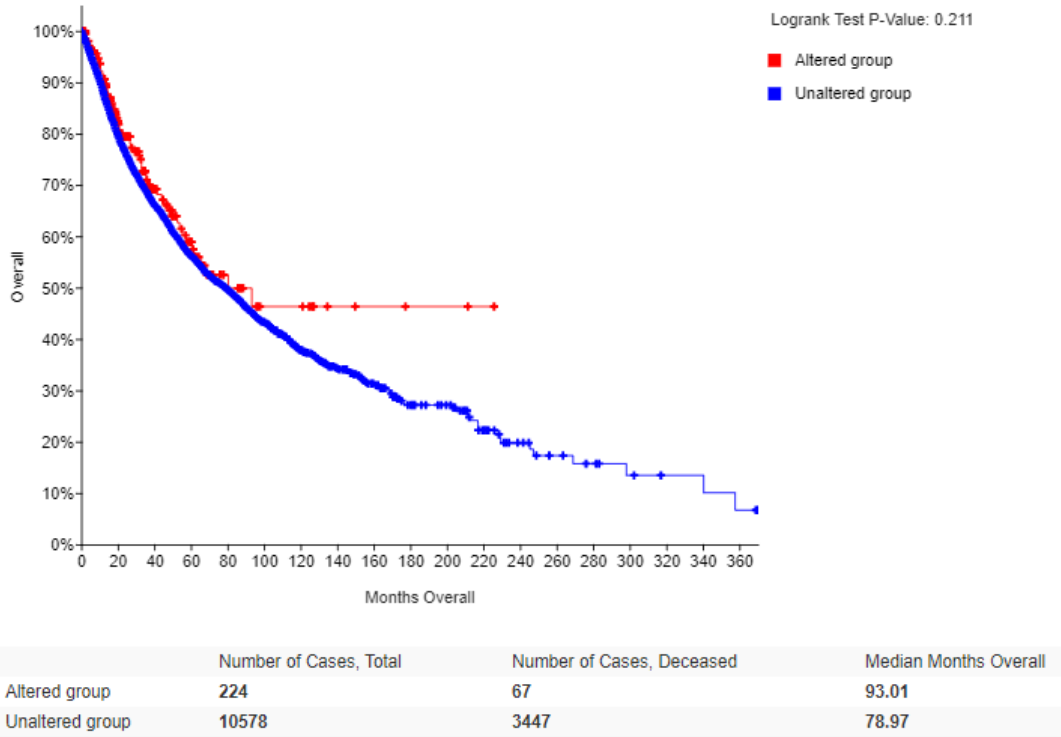


Figura 4.11. Curva de supervivencia de PIWIL3 generada por cBioPortal.

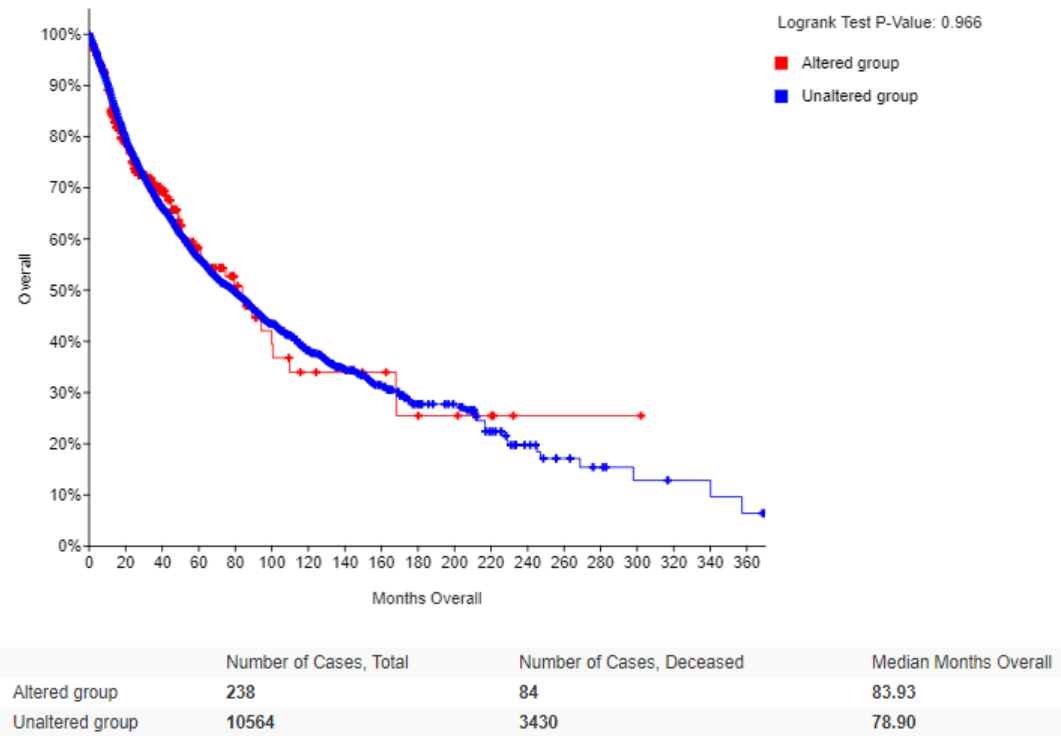


Figura 4.12. Curva de supervivencia de PIWIL4 generada por cBioPortal.

4.1.3. GTEx Portal

Gracias a GTEx Portal, se ha podido observar la expresión y regulación de esta familia de genes de interés en los diferentes tejidos libres de tumor, con el fin de poder comparar estos resultados con los obtenidos en los anteriores servidores relacionados con cáncer. En las Figuras 4.13 a 4.16, se contempla la distribución de la expresión de estos genes en los diferentes tejidos.

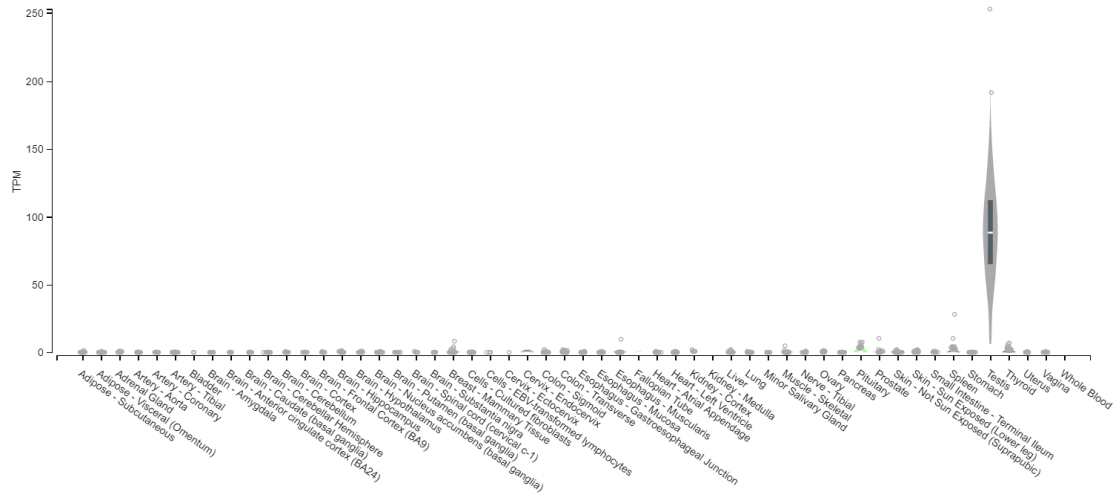


Figura 4.13. Expresión del gen PIWIL1 en los diferentes tejidos registrados en GTEx Portal.

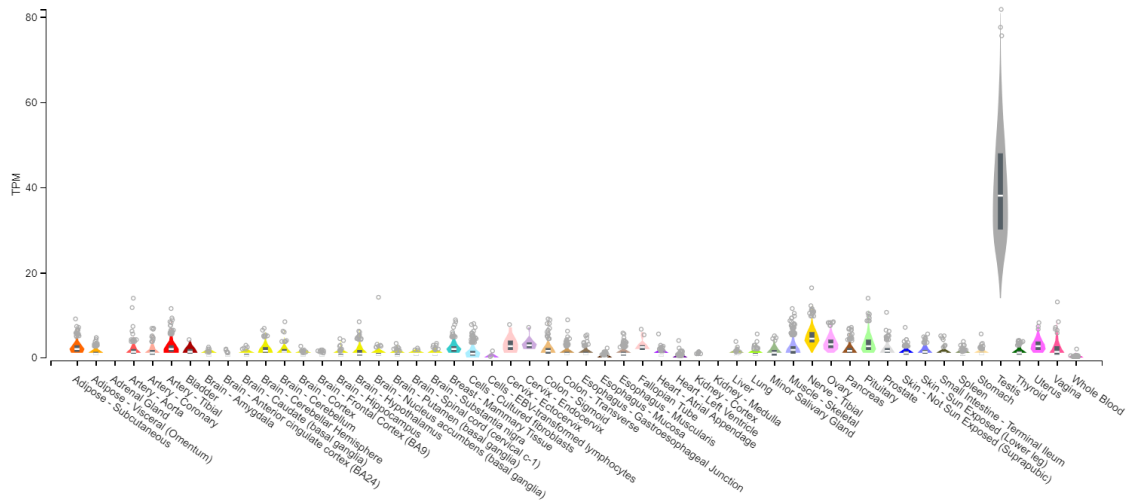


Figura 4.14. Expresión del gen PIWIL2 en los diferentes tejidos registrados en GTEx Portal.

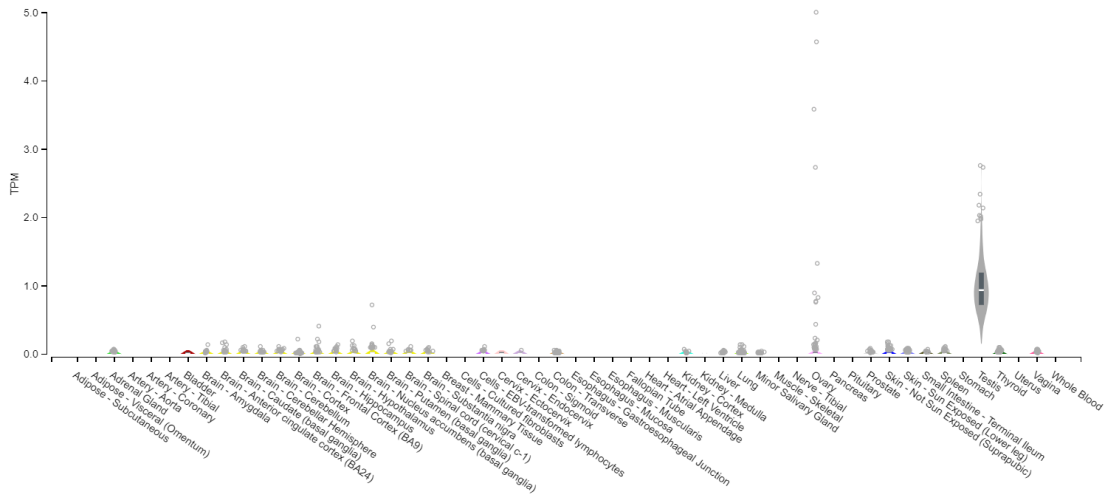


Figura 4.15. Expresión del gen PIWIL3 en los diferentes tejidos registrados en GTEx Portal.

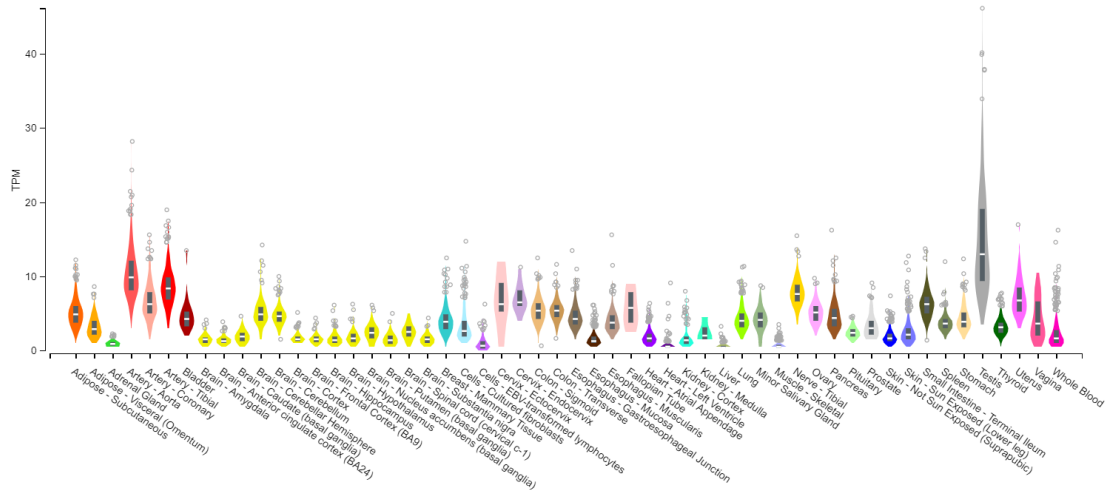


Figura 4.16. Expresión del gen PIWIL4 en los diferentes tejidos registrados en GTEx Portal.

4.1.4. The Human Protein Atlas

Por último, se ha utilizado el programa THPA. Por un lado, se ha empleado el atlas tisular, con el que se ha podido observar tanto la expresión como localización del conjunto de genes en los diferentes tejidos y órganos. De esta forma, se ha podido contrastar esta información con la obtenida en GTEX Portal. En las Figuras 4.17 a 4.26, se recopila la expresión tanto de RNA como de proteína de esta familia de genes.

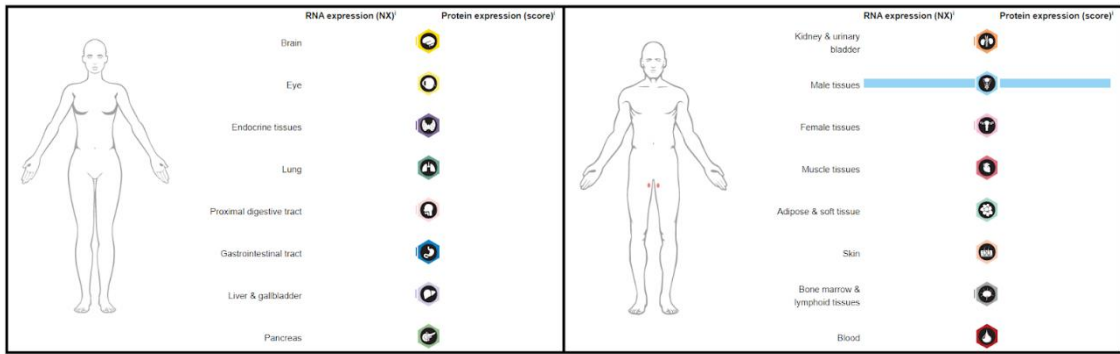


Figura 4.17. Expresión del RNA y de la proteína PIWIL1 en los diferentes tejidos registrados en THPA.

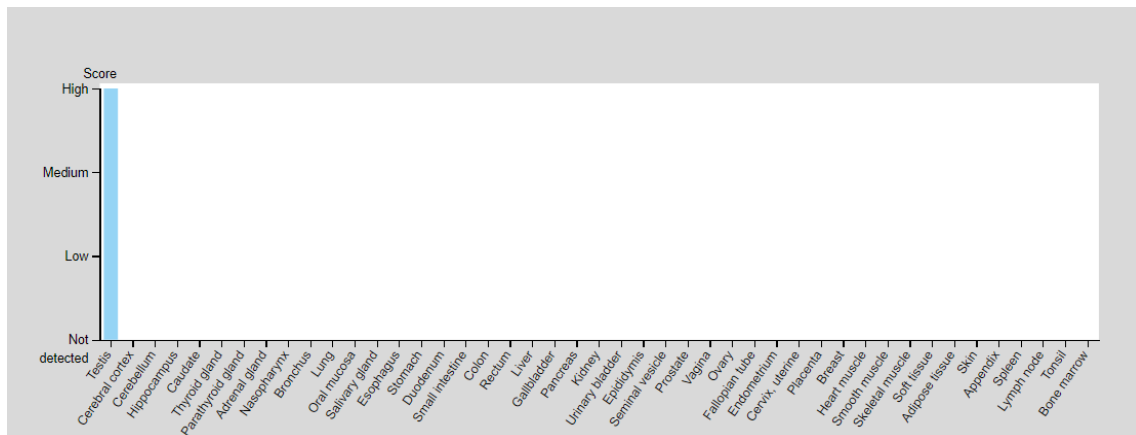


Figura 4.18. Expresión de la proteína PIWIL1 en los diferentes tejidos registrados en THPA.

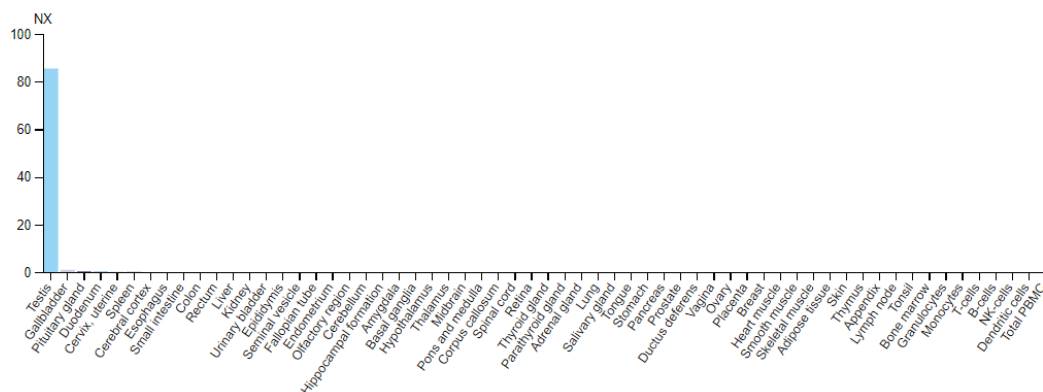


Figura 4.19. Expresión del RNA de PIWIL1 en los diferentes tejidos registrados en THPA.

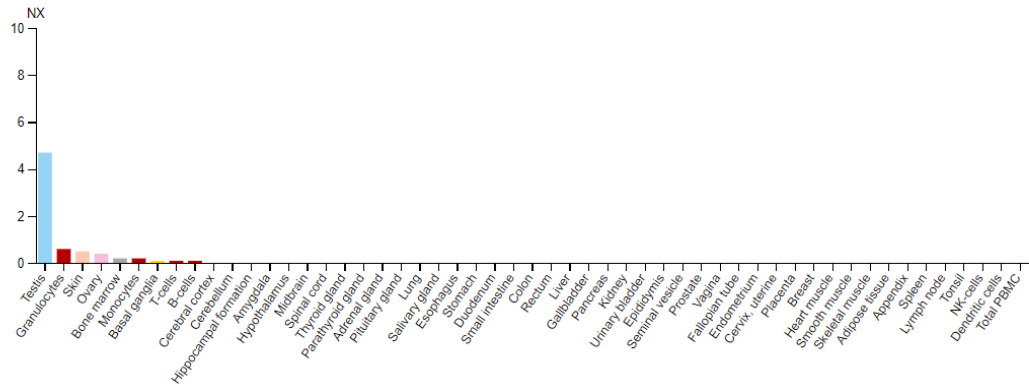


Figura 4.23. Expresión del RNA de PIWIL3 en los diferentes tejidos registrados en THPA.

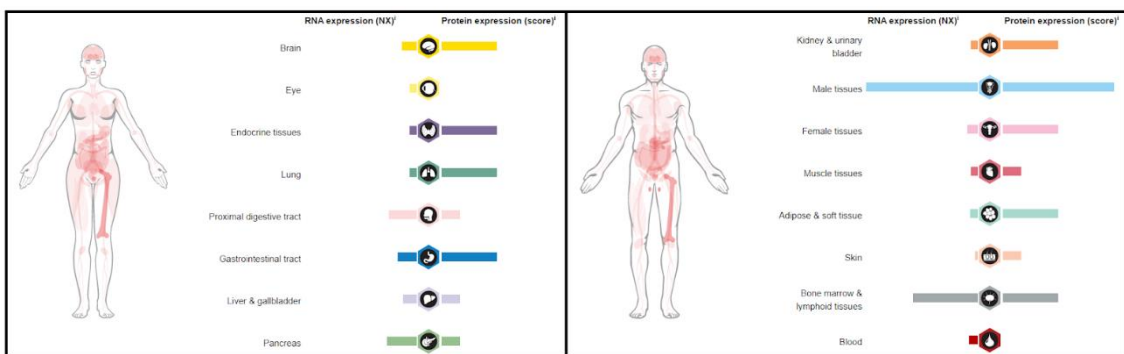


Figura 4.24. Expresión del RNA y de la proteína PIWIL4 en los diferentes tejidos registrados en THPA.

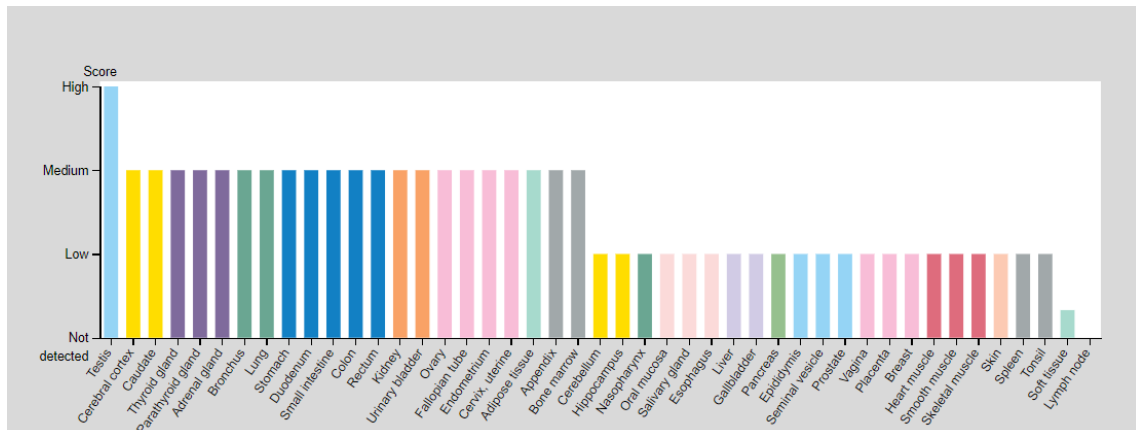


Figura 4.25. Expresión de la proteína PIWIL4 en los diferentes tejidos registrados en THPA.

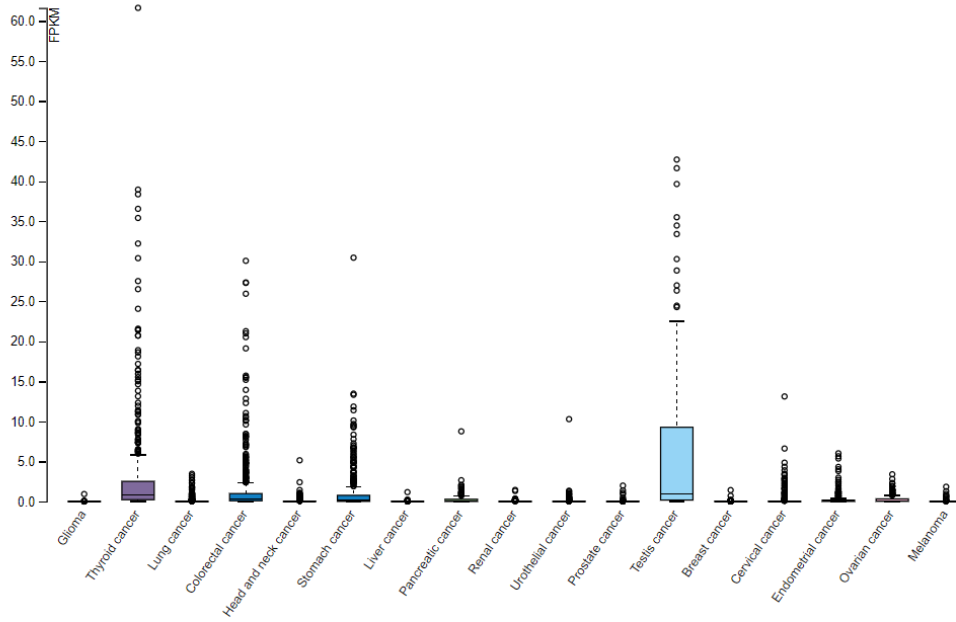


Figura 4.28. Expresión del RNA de PIWIL1 en los grupos tumorales descritos en THPA.

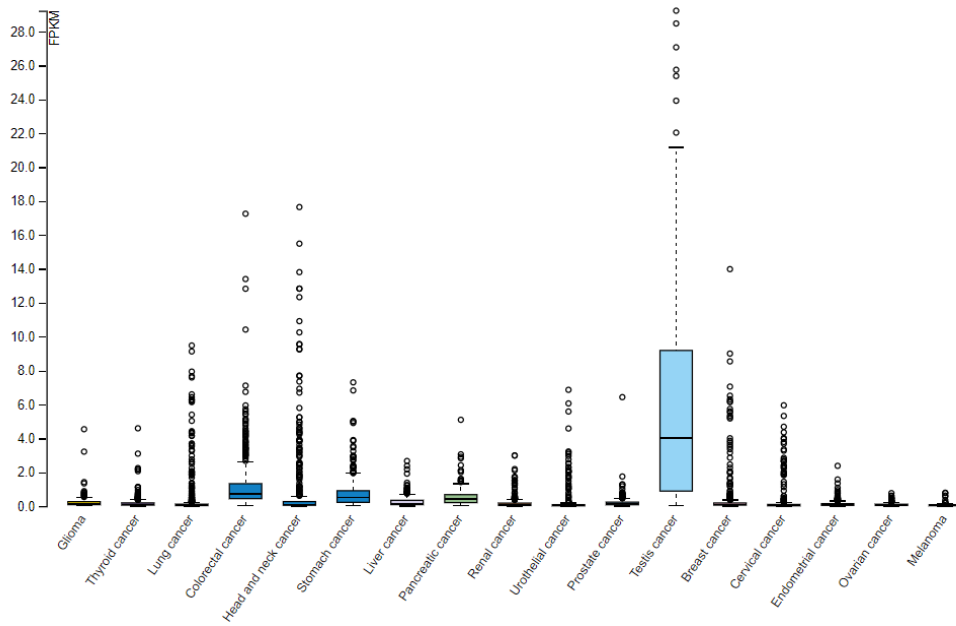


Figura 4.29. Expresión del RNA de PIWIL2 en los grupos tumorales descritos en THPA.

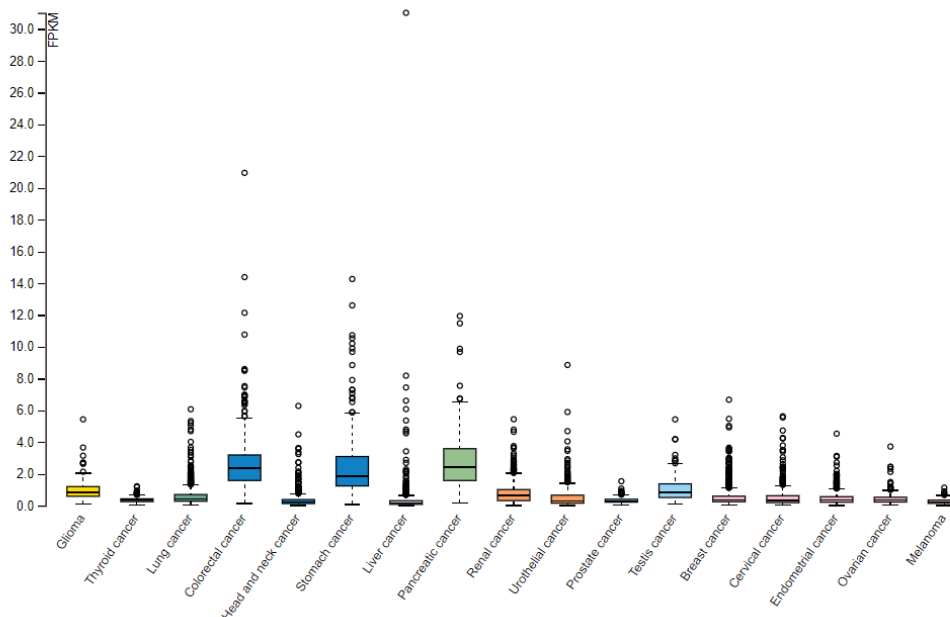


Figura 4.33. Expresión del RNA de PIWIL4 en los grupos tumorales descritos en THPA.

En esta revisión, la información clínica de esta familia de proteínas, que ha sido proporcionada por los servidores y bases de datos de cáncer ya mencionadas en este apartado, será relacionada y discutida en el siguiente punto del presente trabajo.

4.2. Evaluación del papel de las proteínas PIWI en estudios transcriptómicos de adenocarcinoma de pulmón

La aplicación de la estrategia *in silico* para conocer el papel de las proteínas PIWI en el adenocarcinoma de pulmón, incluyó como primer paso la revisión y procesamiento de los datos de cada uno de los estudios seleccionados. Se verificó que todos estos grupos de datos presentaran niveles de expresión en escala logarítmica y en un rango de valores positivos. A continuación, se llevó a cabo un análisis estadístico descriptivo que determinó los niveles medios de expresión en los grupos de enfermos y controles por sexo, así como su variabilidad, con el objetivo de conocer el papel de las proteínas PIWI en este grupo de estudios. Se obtuvieron los siguientes resultados:

4.2.1. GSE10072

La distribución de individuos por sexo y enfermedad de este estudio se muestra en la Tabla 4.5.

Tabla 4.5. Individuos incluidos en el estudio GSE10072.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	8	29	37
<i>Adenocarcinoma</i>	13	30	43
<i>Total</i>	21	59	80

En este estudio sólo hay información de *PIWIL1* y *PIWIL2*, en los gráficos de barras del Anexo I (Figuras A.1 a A.3) se observa que la expresión de *PIWIL2* es mayor que la de *PIWIL1*. Por otro lado, no se observan diferencias entre los distintos individuos, es decir, la expresión de estas proteínas es independiente del tumor y del sexo.

4.2.2. GSE19188

La Tabla 4.6 recoge la distribución de individuos por sexo y enfermedad.

Tabla 4.6. Individuos incluidos en el estudio GSE19288.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	11	41	52
<i>Adenocarcinoma</i>	11	21	32
<i>Total</i>	22	62	84

En este estudio, también aparecen medidas de expresión de *PIWIL4*. En este caso, los gráficos (Fig. A.4 a A.6) muestran que los niveles de expresión de las tres proteínas son muy similares. Además, al igual que en GSE10072 no hay diferencias entre individuos.

4.2.3. GSE31210

La Tabla 4.7 muestra la distribución de individuos por sexo y enfermedad.

Tabla 4.7. Individuos incluidos en el estudio GSE31210.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	9	11	20
<i>Adenocarcinoma</i>	121	105	226
<i>Total</i>	130	116	246

A diferencia del estudio anterior, los niveles de expresión de *PIWIL4* son superiores con respecto a *PIWIL1* y *PIWIL2*, independientemente de las características del individuo. Estos niveles se muestran en las Figuras A.7 a A.9 del Anexo I.

4.2.4. GSE32863

La distribución de individuos por sexo y enfermedad se contempla en la Tabla 4.8.

Tabla 4.8. Individuos incluidos en el estudio GSE32863.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	35	10	45
<i>Adenocarcinoma</i>	35	10	45
<i>Total</i>	70	20	90

En este estudio aparecen medidas para las cuatro proteínas de la familia PIWI. Los gráficos de barras generadas (Figuras A.10 a A.12) muestran los niveles de expresión de cada proteína, en orden descendente serían *PIWIL4*, *PIWIL3*, *PIWIL2* y *PIWIL1*. Los resultados no muestran ninguna diferencia con relación al sexo y a la enfermedad.

4.2.5. GSE63459

La Tabla 4.9 recopila la distribución de individuos por sexo y enfermedad.

Tabla 4.9. Individuos incluidos en el estudio GSE63459.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	17	14	31
<i>Adenocarcinoma</i>	17	15	32
<i>Total</i>	34	29	63

En este estudio, se observa que *PIWIL4* tiene mayores niveles de expresión, seguido de *PIWIL2*. *PIWIL1* y *PIWIL3* presentan niveles similares. Ninguna de las cuatro proteínas presenta diferencias entre los diferentes individuos. Los datos están contemplados en las Figuras A.13 a A.15 del Anexo I.

4.2.6. GSE75037

La Tabla 4.10 muestra la distribución de individuos por sexo y enfermedad.

Tabla 4.10. Individuos incluidos en el estudio GSE75037.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	59	24	83
<i>Adenocarcinoma</i>	59	24	83
<i>Total</i>	118	48	166

En los gráficos de las Figuras A.16 a A.18 del Anexo I, se observa que la proteína *PIWIL4* presenta una mayor expresión con respecto a las otras tres proteínas de esta familia, de las cuales *PIWIL1* es la segunda con mayor expresión. Al igual que en los estudios anteriores, los niveles de expresión no varían con respecto a la enfermedad y al sexo.

4.2.7. GSE81089

La distribución de individuos por sexo y enfermedad se contempla en la Tabla 4.11.

Tabla 4.11. Individuos incluidos en el estudio GSE81089.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	8	11	19
<i>Adenocarcinoma</i>	53	24	77
<i>Total</i>	61	35	96

En este estudio, *PIWIL4* es la proteína con mayor expresión, seguida de *PIWIL2*. En este caso, se observa un ligero aumento en los niveles de *PIWIL1* y *PIWIL3* en los pacientes enfermos con respecto a los sanos. Los datos están representados en las Figuras A.19 a A.21 del Anexo I.

4.2.8. GSE87340

En la Tabla 4.12 se recopila la distribución de individuos por sexo y enfermedad.

Tabla 4.12. Individuos incluidos en el estudio GSE87340.

	<i>Mujeres</i>	<i>Hombres</i>	<i>Total</i>
<i>Control</i>	23	4	27
<i>Adenocarcinoma</i>	23	4	27
<i>Total</i>	46	8	54

En este estudio, en los gráficos de barras de las Figuras A.22 a A.24, se observa que las proteínas según los niveles de expresión en orden descendente serían: *PIWIL4*, *PIWIL2*, *PIWIL3* y *PIWIL1*. Además, se observa un ligero aumento en los niveles de *PIWIL1* en individuos sanos con respecto a enfermos.

Por otro lado, también se ha realizado un análisis individual a nivel de proteína para los diferentes estudios. Estos resultados quedan contemplados en las Figuras 4.34 a 4.37.

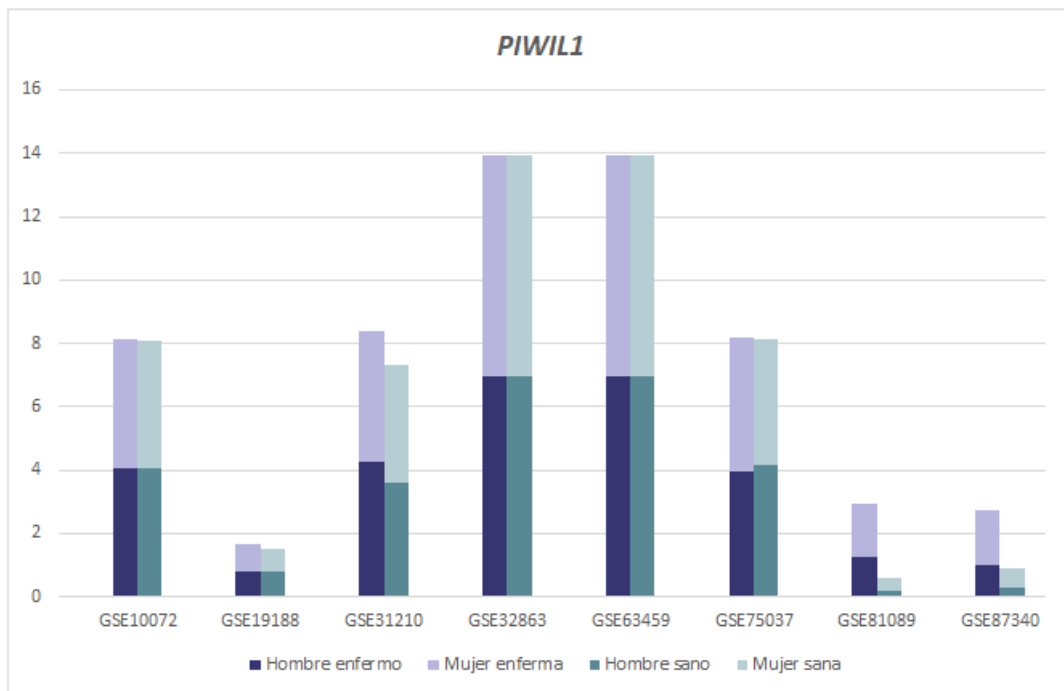


Figura 4.34. Niveles de expresión de PIWIL1 según la distribución de individuos por sexo y enfermedad.

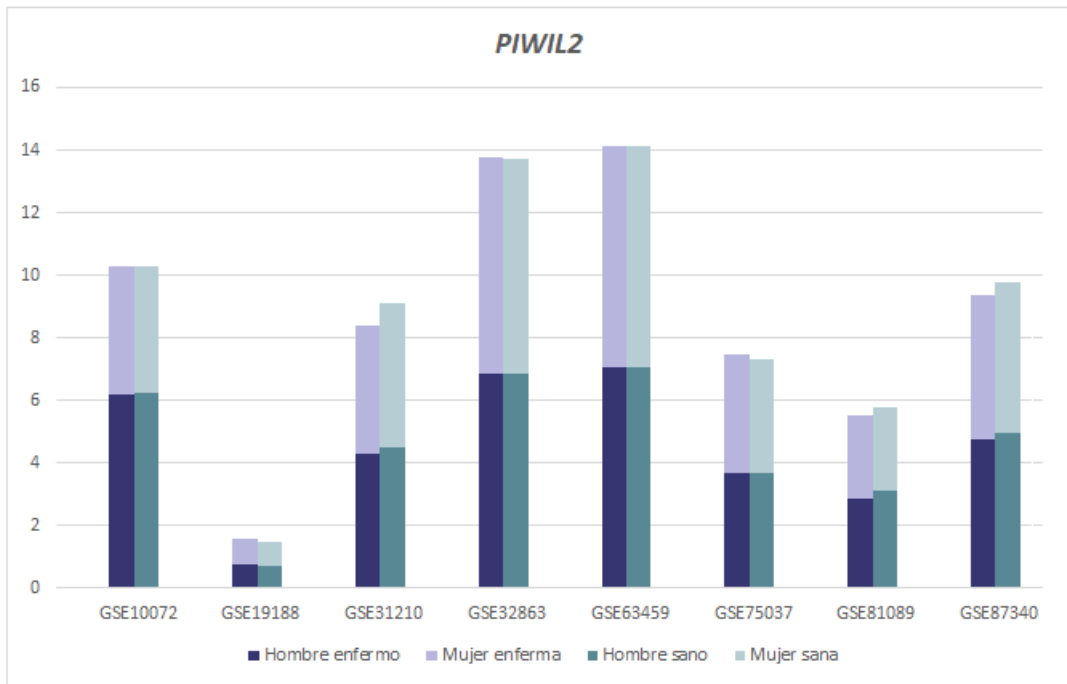


Figura 4.35. Niveles de expresión de PIWIL2 según la distribución de individuos por sexo y enfermedad.

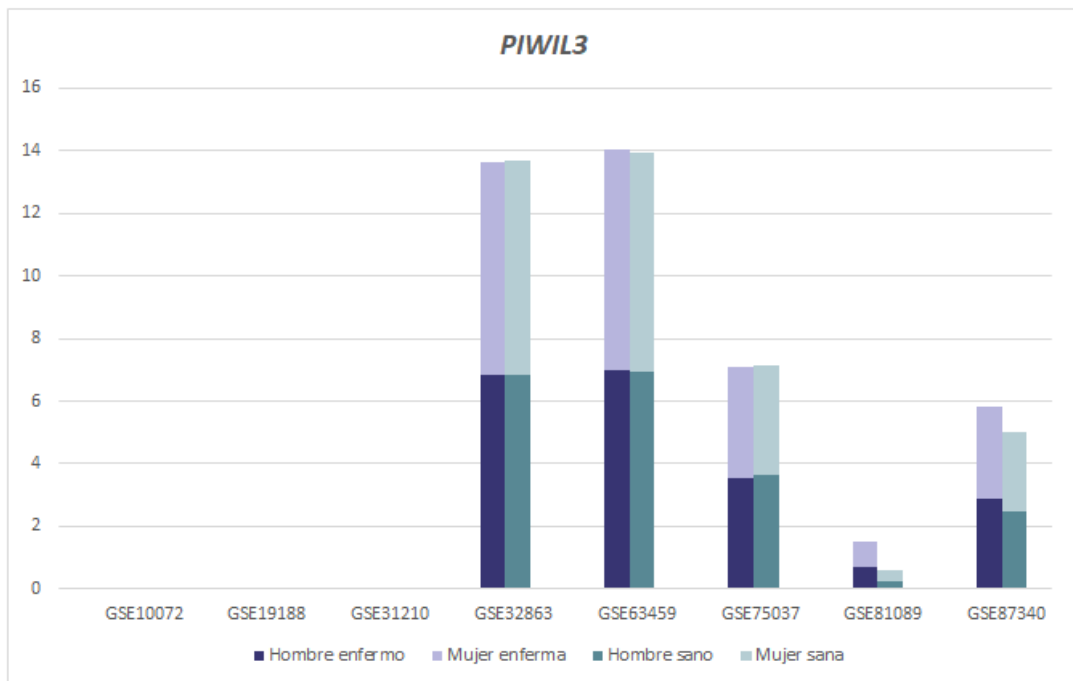


Figura 4.36. Niveles de expresión de PIWIL3 según la distribución de individuos por sexo y enfermedad.

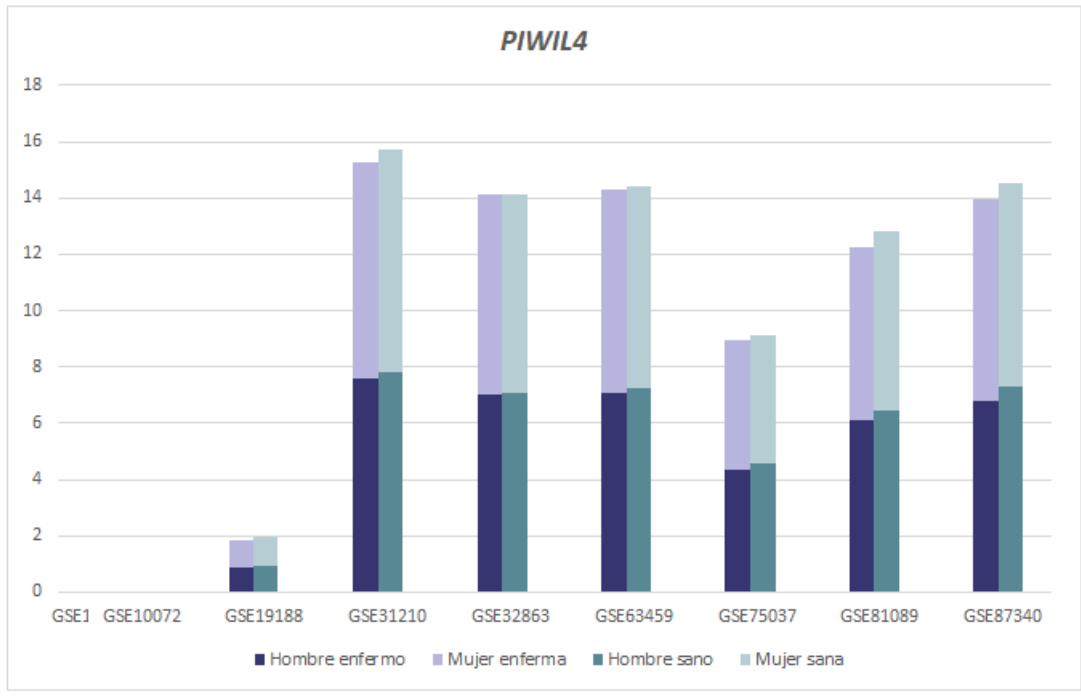


Figura 4.37. Niveles de expresión de PIWIL4 según la distribución de individuos por sexo y enfermedad.

Este análisis de los niveles de expresión de la familia de proteínas PIWI, atendiendo a la distribución de los individuos según el sexo y la enfermedad, será discutida en el siguiente punto del presente trabajo.

5. DISCUSIÓN

Uno de los principales problemas en el desarrollo y evolución del cáncer es su tardía detección, que implica una falta de eficacia de los tratamientos terapéuticos aplicados y, como consecuencia, un aumento de mortalidad. Por tanto, una detección temprana del tumor es determinante para una mejor predicción del curso de la enfermedad y para reducir la tasa de mortalidad y de incidencia avanzada (Wender et al., 2019). Para que esto sea posible, es necesario el descubrimiento de nuevas estrategias de detección, como el reconocimiento de nuevos biomarcadores para un mejor diagnóstico y pronóstico del cáncer. Es en este punto donde la familia de proteínas PIWI juega un papel importante, pues muchos estudios recientes las muestran como un potencial biomarcador en el pronóstico del cáncer (Liu et al., 2019).

Para abordar la caracterización del papel pronóstico de esta familia de proteínas, se ha optado por una estrategia *in silico* basada en la revisión de datos ómicos en cáncer.

En primer lugar, con los resultados proporcionados por GTEx Portal y THPA, contemplados en las Figuras 4.13 a 4.26, se observa una mayor expresión de la familia PIWI en los testículos. Estos resultados tienen una concordancia biológica, ya que estos genes cuentan con importantes funciones en el control de la línea germinal masculina y la espermatogénesis. Además, en los casos particulares de *PIWIL2* y *PIWIL4* (Fig. 4.14, 4.16, 4.20, 4.24), se observa que su expresión se distribuye en otros tejidos humanos, teniendo unos niveles más altos *PIWIL4*.

Atendiendo a los resultados de GEPIA de las Figuras 4.1 a 4.4, se presentan los tipos tumorales en los que este grupo de genes tienen una mayor o menor expresión estadísticamente significativa. En el caso de *PIWIL1* (Figura 4.1), se observa una mayor expresión en cáncer de tiroides (THCA) y una baja expresión en cáncer de testículo (TGCT), *PIWIL2* (Figura 4.2) presenta una baja expresión en cáncer de testículo (TGCT), cáncer epitelial de ovario (OV), cáncer de endometrio (UCEC) y carcinosarcoma uterino (UCS), *PIWIL3* (Figura 4.3) no tiene valores de expresión que sean estadísticamente significativos y, por último, *PIWIL4* (Figura 4.4) está sobreexpresado en leucemia mieloide aguda (LAMIL) y presenta baja expresión en cáncer de mama (BRCA), cáncer de células escamosas del cérvix (CESC), adenocarcinoma de pulmón (LUAD), carcinoma de células escamosas de pulmón (LUSC), cáncer epitelial de ovario (OV), cáncer de testículo (TGCT), cáncer de endometrio (UCEC) y carcinosarcoma uterino (UCS). Por tanto, podemos afirmar que estos tres genes tienen un papel diagnóstico en cada uno de los tumores mencionados.

En el heatmap de la Figura 4.5, se comparan los niveles de expresión de los cuatro genes en los distintos grupos tumorales, obteniendo unos resultados afines a los mencionados anteriormente. Los genes *PIWIL2* y *PIWIL4* tienen una expresión que se distribuye a lo largo de los diferentes tumores. Además, comparando esta información con la obtenida de GTEx Portal y THPA, estos genes también tienen una mayor expresión en diversos tejidos libres de tumor con respecto a los otros dos genes de esta misma familia. En cuanto a *PIWIL1*, tiene una mayor expresión en varios tipos de cáncer como adenocarcinoma de colon (COAD), adenocarcinoma de recto (READ), TGCT y THCA, pero en la Figura 4.1 se comprobó que esos niveles sólo son significativos en TGCT y THCA. En cambio, *PIWIL3* no presenta ninguna expresión diferencial significativa en los distintos tumores, lo que nos lleva a pensar que no está involucrado con el desarrollo del cáncer.

Por otro lado, con GEPIA hemos generado unas curvas de supervivencia para cada gen, en aquellos tipos de cáncer con niveles de expresión diferencial significativos. En las Figuras 4.6 a 4.8, se reflejan estas curvas, en las que se compara la supervivencia entre los grupos de paciente con alta y baja expresión de cada gen de interés. Para la evaluación de estos resultados, hay que determinar que sólo aquellas curvas que cuenten con un valor de p (designado como $p(\text{HR})$ en los gráficos) menor del 0.05, las diferencias de tiempo de supervivencia serán estadísticamente

significativas. De modo que, ni la expresión diferencial de *PIWIL1* (Figura 4.6) ni la de *PIWIL2* (Figura 4.7) están involucradas en la tasa de supervivencia de los pacientes. En el caso de *PIWIL4* (Figura 4.8), la curva de supervivencia de LAMIL, con un valor de p de 0.048, nos revela que un alto nivel de expresión de este gen se asocia con una mayor supervivencia, mientras que la curva de supervivencia de UCEC con un valor de p de 0.013, nos muestra que un bajo nivel de expresión de este gen provoca una disminución en la supervivencia.

Con el servidor *cBioPortal*, se han estudiado las alteraciones más frecuentes que tienen lugar en esta familia de genes en los diferentes tumores, información contemplada en las Tablas 4.1 a 4.4. Se observa que el melanoma y el UCEC son los tipos de cáncer con mayor porcentaje de alteración en los cuatro genes de las proteínas PIWI. Además, las alteraciones que tienen lugar en estos genes son principalmente mutaciones y, con menor porcentaje de aparición, amplificaciones y deleciones.

También se ha evaluado si estas alteraciones influyen en la tasa de supervivencia de los pacientes, información mostrada en las Figuras 4.9 a 4.12. Al igual que con los resultados obtenidos en GEPIA, sólo aquellas curvas con un valor de p menor que 0.05 serán consideradas estadísticamente significativas. Por ello, la curva de supervivencia del gen *PIWIL1* (Figura 4.9), con un valor de p de $3.717e-3$, indica que las alteraciones en este gen están relacionadas con una mayor tasa de supervivencia, mientras que la curva de *PIWIL2* (Figura 4.10), con un valor de p de 0.0275, asocia las alteraciones del gen con una menor tasa de supervivencia.

THPA también nos ha permitido valorar la expresión de RNA y de cada proteína de interés en los diferentes tipos de cáncer, tal y como se observa en las Figuras 4.27 a 4.33. En este repositorio se detectan diferencias en los perfiles de expresión entre estas moléculas, lo que nos lleva a pensar que se producen cambios desde la expresión de RNA hasta la expresión de proteína. Esto se traduce en que, dependiendo del tipo de abordaje ómico, proteómica o transcriptómica, empleado se obtendrán diferentes resultados. Atendiendo a los niveles de expresión de RNA obtenidos para los diferentes grupos tumorales, se observa una alta expresión de los genes *PIWIL1* y *PIWIL2* en TGCT (Figuras 4.29 y 4.30). En cuanto a los niveles de expresión de la proteína, se contempla una alta expresión de *PIWIL1* y *PIWIL3* en cáncer colorrectal (Figuras 4.28 y 4.31), mientras que en cáncer de próstata hay una alta expresión de *PIWIL3* y *PIWIL4* (Figuras 4.31 y 4.33). Contrastando esta información con los resultados en GEPIA, en el TGCT no coincide el patrón de expresión, ya que, para este tumor, los niveles de expresión de *PIWIL1* y *PIWIL2* son inferiores a los controles.

En la segunda parte de esta estrategia *in silico*, los resultados obtenidos del análisis estadístico descriptivo sobre los niveles de expresión de las proteínas PIWI, apuntan a que no están involucradas en el adenocarcinoma de pulmón., puesto que en ninguno de los gráficos del Anexo I, se observan diferencias entre individuos enfermos y sanos. Por otro lado, sí que se contempla que *PIWIL4* tiene una mayor expresión con respecto a los otros tres genes, lo que concuerda con la revisión realizada previamente por las bases de datos GTEx Portal y THPA. En GEPIA, *PIWIL4* es la única proteína en la que se muestran resultados significativos en LUAD, concretamente una infraexpresión. Sin embargo, en este análisis no se aprecia que haya diferencias entre enfermos y sanos, lo que puede deberse a un sesgo tecnológico, puesto que GEPIA emplea datos secuenciados por RNA-Seq y la selección de estudios empleada ha sido secuenciada mediante microarrays. Por lo tanto, merece la pena considerar la información de la plataforma tecnológica tanto en las bases de datos como en nuestros estudios, para evitar discrepancias. Además, también existe un desbalanceo en el tamaño muestral ya que en algunos estudios la distribución de individuos en las diferentes categorías no es homogénea. Tampoco se observan diferencias ligadas al sexo, por lo que podemos afirmar que la familia PIWI no es un factor genético ligado al sexo.

Con nuestro estudio, hemos sido capaces de determinar el papel diagnóstico de tres de los cuatro miembros de la familia PIWI y observar que las alteraciones en sus secuencias genéticas son un factor pronóstico de la enfermedad.

6. CONCLUSIONES

La detección temprana en cáncer es la estrategia más efectiva para reducir la mortalidad de esta enfermedad. Por ello, en este estudio se ha evaluado el papel como biomarcador de las proteínas PIWI.

1. Los recursos y bases de datos en cáncer permiten establecer estrategias *in silico* para revisar y evaluar el papel de potenciales biomarcadores en diferentes grupos tumorales.
2. PIWIL4 es el gen con mayor expresión en numerosos tejidos.
3. PIWIL1, PIWIL2 y PIWIL4 tiene un papel diagnóstico en diversos tumores.
4. PIWIL4 tiene un papel pronóstico en UCEC y LAMIL.
5. Las alteraciones en PIWIL1 se asocian con una mayor tasa de supervivencia mientras que las alteraciones en PIWIL2 se relacionan con una menor supervivencia.
6. PIWIL3 no está involucrado en el desarrollo del cáncer.

7. BIBLIOGRAFÍA

- ALANDES, S. (2017). *Abordajes Bioinformáticos para la Integración Funcional de Datos Transcriptómicos en Biomedicina*.
- ARANGO, S. S. (2012). *Biomarkers for the evaluation of human health risks*. 75–82.
- ARAVIN, A. A., SACHIDANANDAM, R., GIRARD, A., FEJES-TOTH, K., & HANNON, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, 316(5825), 744–747.
- BAUMGARTNER, C., OSL, M., NETZER, M., & BAUMGARTNER, D. (2011). Bioinformatic-driven search for metabolic biomarkers in disease. *Journal of Clinical Bioinformatics*, 1(1), 1–10.
- BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A., & JEMAL, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424.
- CERAMI, E., GAO, J., DOGRUSOZ, U., GROSS, B. E., SUMER, S. O., AKSOY, B. A., JACOBSEN, A., BYRNE, C. J., HEUER, M. L., LARSSON, E., ANTIPIN, Y., REVA, B., GOLDBERG, A. P., SANDER, C., & SCHULTZ, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery*, 2(5), 401–404.
- COX, D. N., CHAO, A., & LIN, H. (2000). *PIWI, a nuclear stem cell factor whose activity modulates the number and division rate of germline stem cells*.
- DAYEH, D. M., KRUITHOFF, B. C., & NAKANISHI, K. (2018). Structural and functional analyses reveal the contributions of the C- and N-lobes of Argonaute protein to selectivity of RNA target cleavage. *Journal of Biological Chemistry*, 293(17), 6308–6325. <https://doi.org/10.1074/jbc.RA117.001051>
- EDGAR, R., DOMRACHEV, M., & LASH, A. E. (2002). *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. 30(1), 207–210.
- GAO, J., AKSOY, B. A., DOGRUSOZ, U., DRESDNER, G., GROSS, B., SUMER, S. O., SUN, Y., JACOBSEN, A., SINHA, R., LARSSON, E., CERAMI, E., SANDER, C., & SCHULTZ, N. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling*, 6(269), p11–p11.
- GROSSMAN, R. L., HEATH, A. P., FERRETTI, V., VARMUS, H. E., LOWY, D. R., KIBBE, W. A., & STAUDT, L. M. (2016). Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine*, 375(12), 1109–1112.
- HAN, Y. N., LI, Y., XIA, S. Q., ZHANG, Y. Y., ZHENG, J. H., & LI, W. (2017). PIWI Proteins and PIWI-Interacting RNA: Emerging Roles in Cancer. *Cellular Physiology and Biochemistry*, 44(1), 1–20.
- JENSEN, M. A., FERRETTI, V., GROSSMAN, R. L., & STAUDT, L. M. (2017). The NCI Genomic Data Commons as an engine for precision medicine. *Blood*, 130(4), 453–459.
- KINOSHITA, F. L., ITO, Y., MORISHIMA, T., MIYASHIRO, I., & NAKAYAMA, T. (2017). Sex differences in lung cancer survival: long-term trends using population-based cancer registry data in Osaka, Japan. *JJCO Japanese Journal of Clinical Oncology Japanese Journal of Clinical Oncology*, 47(9), 863–869.
- LI, L., YU, C., GAO, H., & LI, Y. (2010). Argonaute proteins: Potential biomarkers for human colon cancer. *BMC Cancer*, 10(1), 1–8.

- LIBERATI, A., ALTMAN, D. G., TETZLAFF, J., MULROW, C., GÖTZSCHE, P. C., IOANNIDIS, J. P., CLARKE, M., DEVEREAUX, P., KLEIJNEN, J., & MOHER, D. (2009). *The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration*.
- LIU, Y., DOU, M., SONG, X., DONG, Y., LIU, S., LIU, H., TAO, J., LI, W., YIN, X., & XU, W. (2019). The emerging role of the piRNA/piwi complex in cancer. *Molecular Cancer*, 18(1), 1–15.
- LU, Y., ZHANG, K., LI, C., YAO, Y., TAO, D., LIU, Y., ZHANG, S., & MA, Y. (2012). Piwil2 suppresses P53 by inducing phosphorylation of signal transducer and activator of transcription 3 in tumor cells. *PLoS ONE*, 7(1).
- NATIONAL CANCER INSTITUTE. (n.d.). *¿Qué es el cáncer?* Last Update February 2015. <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>
- NATIONAL CANCER INSTITUTE. (2006). *The Cancer Genome Atlas Program*. <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>
- OZATA, D. M., GAINETDINOV, I., ZOCH, A., O'CARROLL, D., & ZAMORE, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics*, 20(2), 89–108.
- PENTTINEN, J., PUJANTO, D. A., SIPILÄ, P., HUHTANIEMI, I., & POUTANEN, M. (2003). Discovery in Silico and Characterization in Vitro of Novel Genes Exclusively Expressed in the Mouse Epididymis. *Molecular Endocrinology*, 17(11), 2138–2151.
- PÉREZ, I. (2019). *Metaanálisis funcional de las diferencias de sexo en estudios ómicos de adenocarcinoma de pulmón*.
- PHAN, J. H., MOFFITT, R. A., STOKES, T. H., LIU, J., YOUNG, A. N., NIE, S., & WANG, M. D. (2009). Convergence of biomarkers, bioinformatics and nanotechnology for individualized cancer treatment. *Trends in Biotechnology*, 27(6), 350–358.
- R CORE TEAM. (2018). *R: A language and environment for statistical computing*.
- RADKIEWICZ, C., DICKMAN, P. W., JOHANSSON, A. L. V., WAGENIUS, G., EDGREN, G., & LAMBE, M. (2019). Sex and survival in non-small cell lung cancer: A nationwide cohort study. *PLoS ONE*, 14(6).
- SCULLY, O. J., BAY, B.-H., YIP, G., & YU, Y. (2012). Breast Cancer Metastasis. *Cancer Genomics - Proteomics*, 9(5), 311–320.
- SEOM: SOCIEDAD ESPAÑOLA DE ONCOLOGÍA MÉDICA. (2019). *¿Qué es el cáncer y cómo se desarrolla?* <https://seom.org/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla?showall=1>
- SETO, A. G., KINGSTON, R. E., & LAU, N. C. (2007). The Coming of Age for Piwi Proteins. *Molecular Cell*, 26(5), 603–609.
- STRIMBU, K., & TAVEL, J. A. (2010). What are biomarkers? *Current Opinion in HIV and AIDS*, 5(6), 463–466.
- TANG, Z., LI, C., KANG, B., GAO, G., LI, C., & ZHANG, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Web Server Issue Published Online*, 45.
- TERAMOTO, A., TSUKAMOTO, T., KIRIYAMA, Y., & FUJITA, H. (2017). *Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks*.

- THE GTEx CONSORTIUM. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235), 648–660.
- UHLÉN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJOSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A.-K., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ... PONTEN, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419–1260419.
- UHLÉN, MATHIAS, ZHANG, C., LEE, S., SJÖSTEDT, E., FAGERBERG, L., BIDKHORI, G., BENFEITAS, R., ARIF, M., LIU, Z., EDFORS, F., SANLI, K., VON FEILITZEN, K., OKSVOLD, P., LUNDBERG, E., HOBER, S., NILSSON, P., MATTSSON, J., SCHWENK, J. M., BRUNNSTRÖM, H., ... PONTEN, F. (2017). A pathology atlas of the human cancer transcriptome. *Science*, 357(6352).
- WANG, Q. E., HAN, C., MILUM, K., & WANI, A. A. (2011). Stem cell protein Piwil2 modulates chromatin modifications upon cisplatin treatment. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 708(1–2), 59–68.
- WANG, X. (2011). Role of clinical bioinformatics in the development of network-based Biomarkers. *Journal of Clinical Bioinformatics*, 1(1), 1–3.
- WENDER, R. C., BRAWLEY, O. W., FEDEWA, S. A., GANSLER, T., & SMITH, R. A. (2019). A blueprint for cancer screening and early detection: Advancing screening’s contribution to cancer control. *CA: A Cancer Journal for Clinicians*, 69(1), 50–79.
- WILHITE, S. E., & BARRETT, T. (2012). Strategies to explore functional genomics data sets in NCBI’s GEO database. *Methods in Molecular Biology*, 802, 41–53.
- WILKINSON, M. D. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Nature Publishing Group*.
- WORLD HEALTH ORGANIZATION. (2001). *Biomarkers In Risk Assessment: Validity And Validation. International Programme on Chemical Safety*, 1–21.
- WORLD HEALTH ORGANIZATION. (2018). *Cáncer*. <https://www.who.int/es/news-room/fact-sheets/detail/cancer>

A. ANEXO

I. Figuras

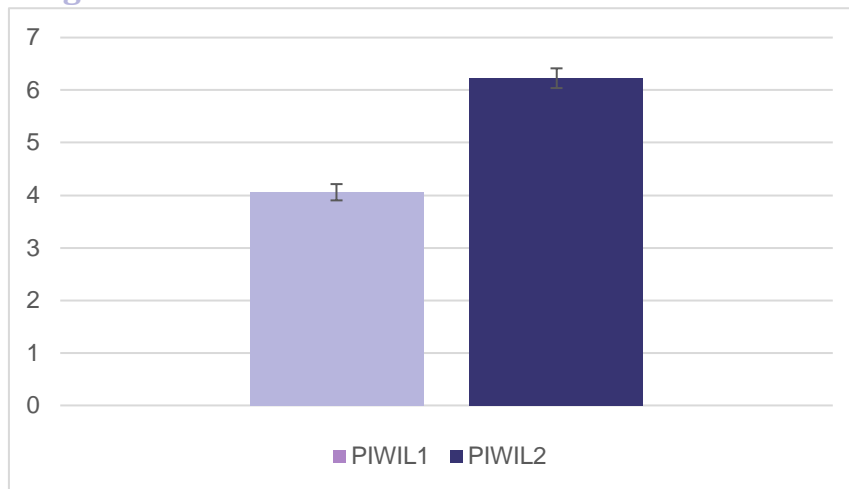


Figura A.1. Nivel de expresión de las proteínas PIWI en GSE10072.

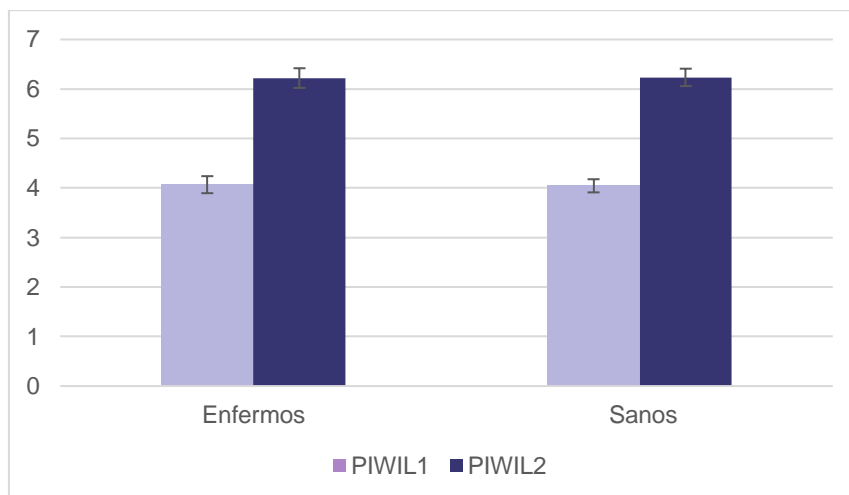


Figura A.2. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE10072.

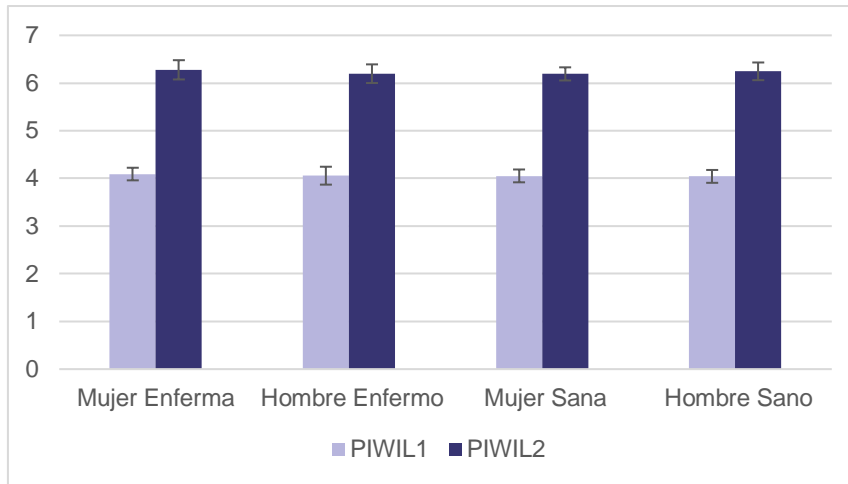


Figura A.3. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE10072.

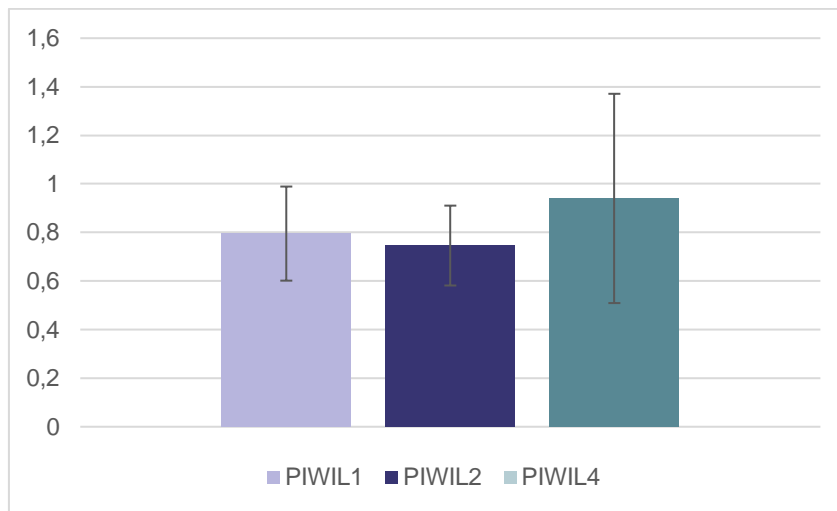


Figura A.4. Nivel de expresión de las proteínas PIWI en GSE19188.

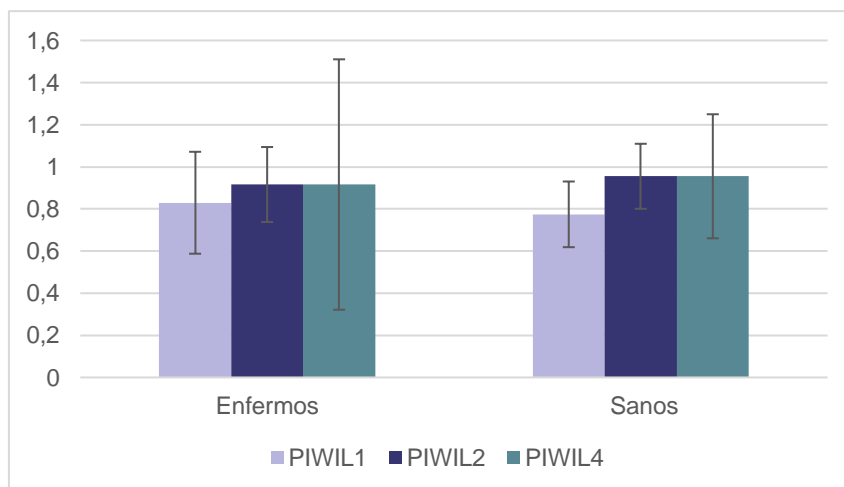


Figura A.5. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE19188.

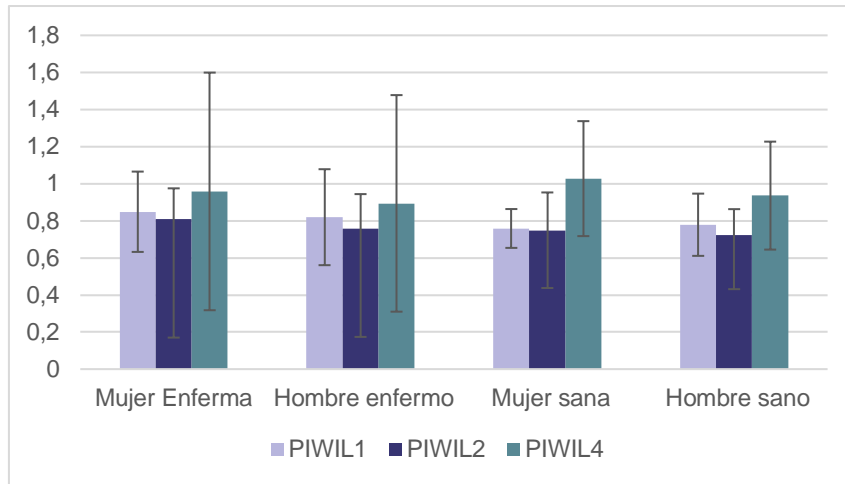


Figura A.6. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE19188.

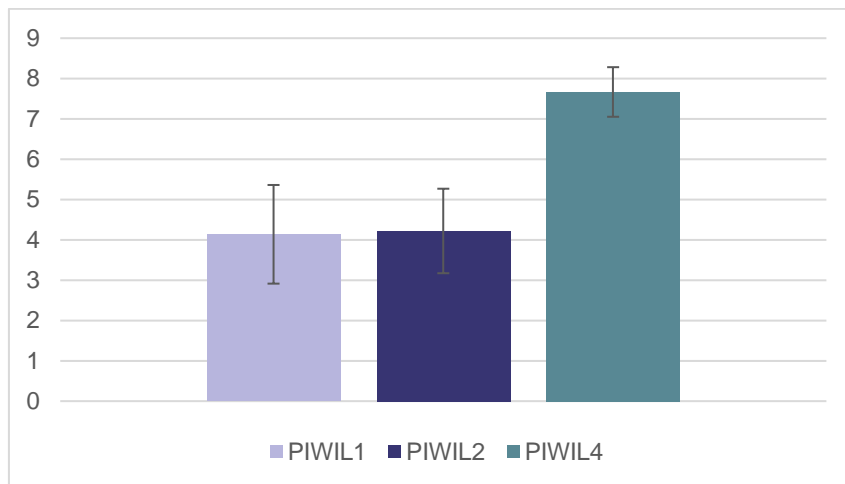


Figura A.7. Nivel de expresión de las proteínas PIWI en GSE31210.

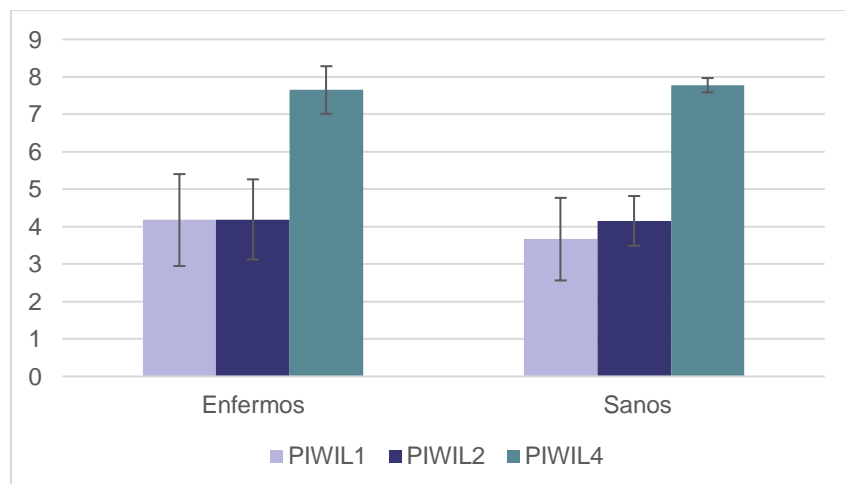


Figura A.8. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE31210.

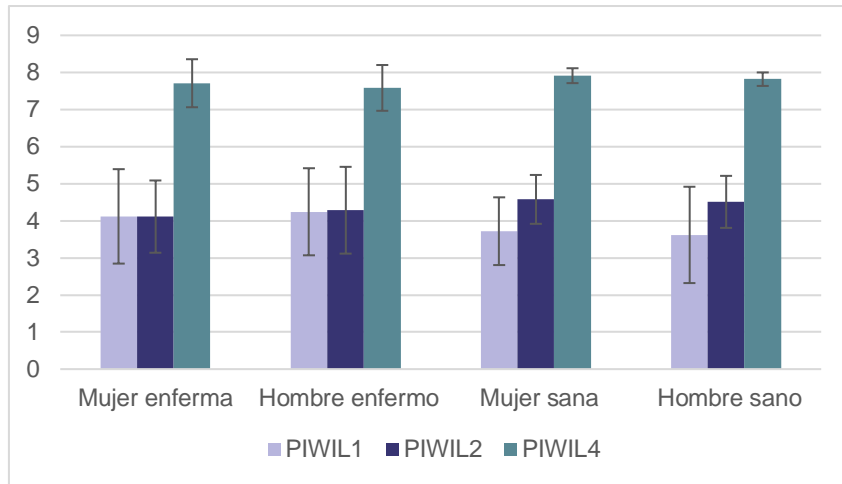


Figura A.9. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE31210.

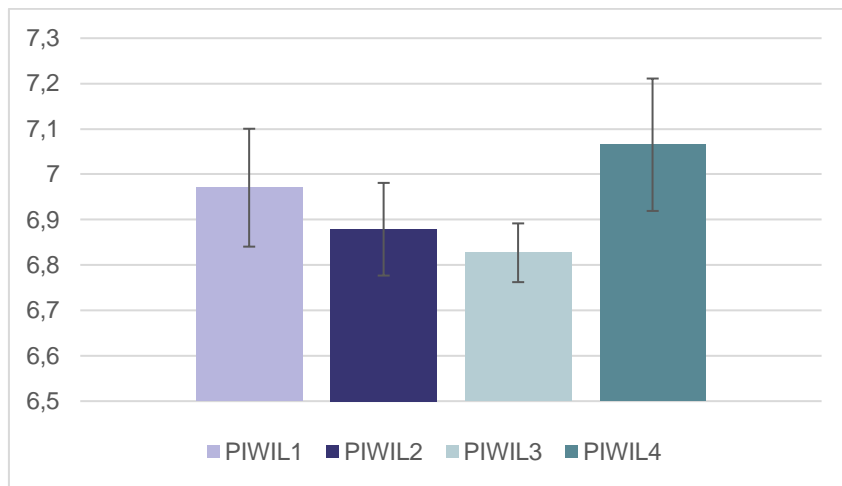


Figura A.10. Nivel de expresión de las proteínas PIWI en GSE32863.

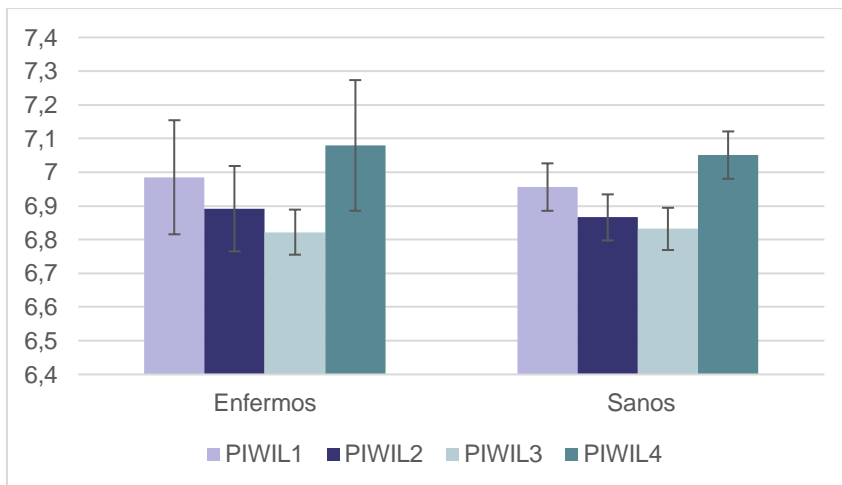


Figura A.11. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE32863.

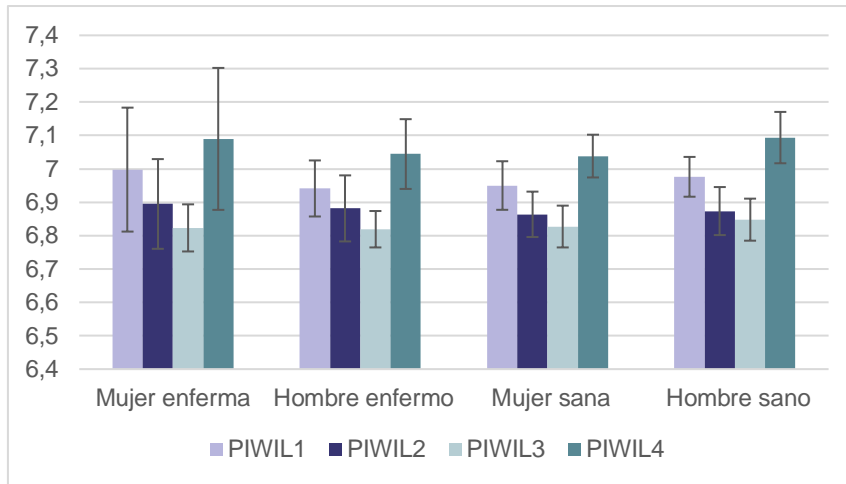


Figura A.12. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE32863.

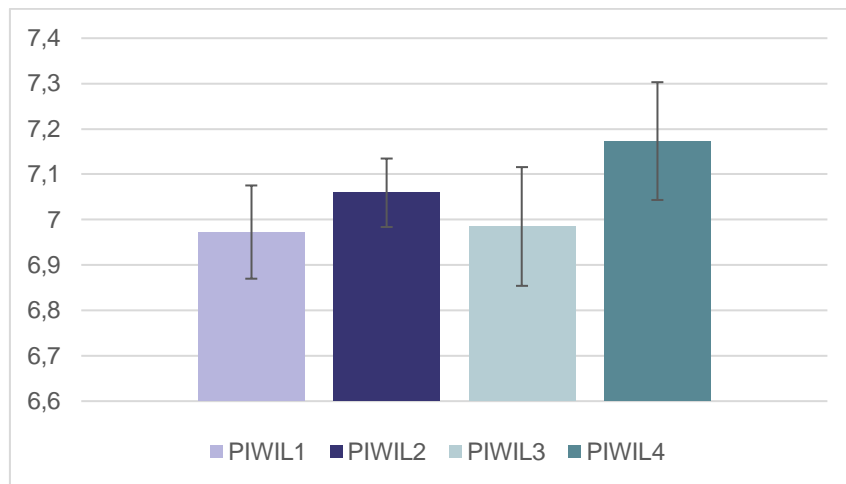


Figura A.13. Nivel de expresión de las proteínas PIWI en GSE63459.

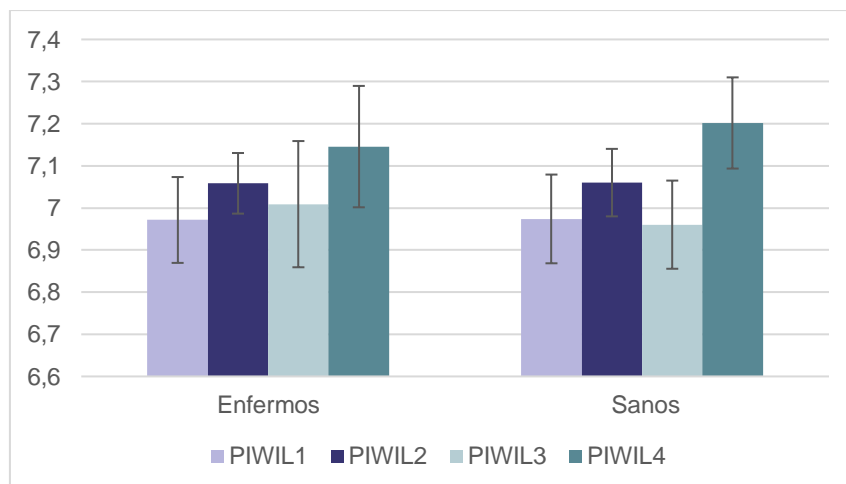


Figura A.14. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE63459.

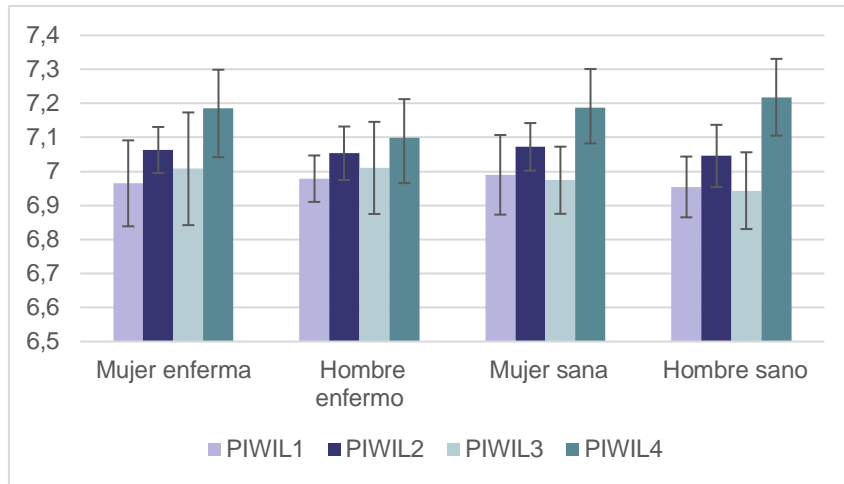


Figura A.15. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE63459.

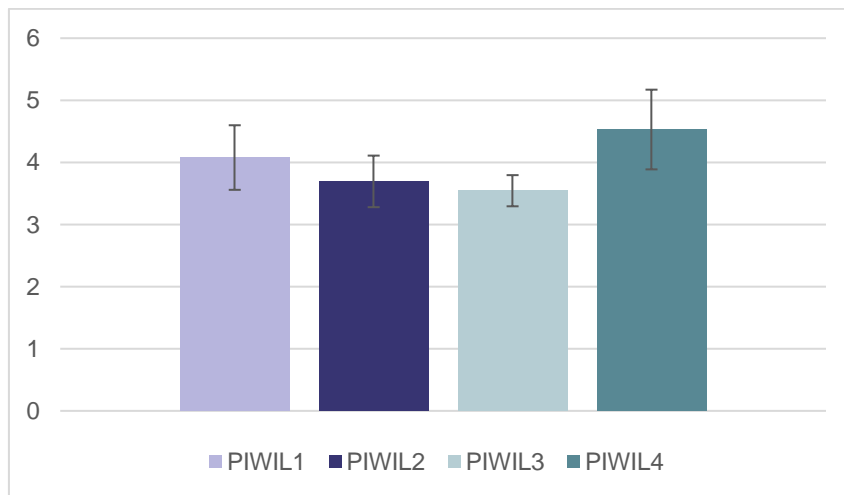


Figura A.16. Nivel de expresión de las proteínas PIWI en GSE75037.

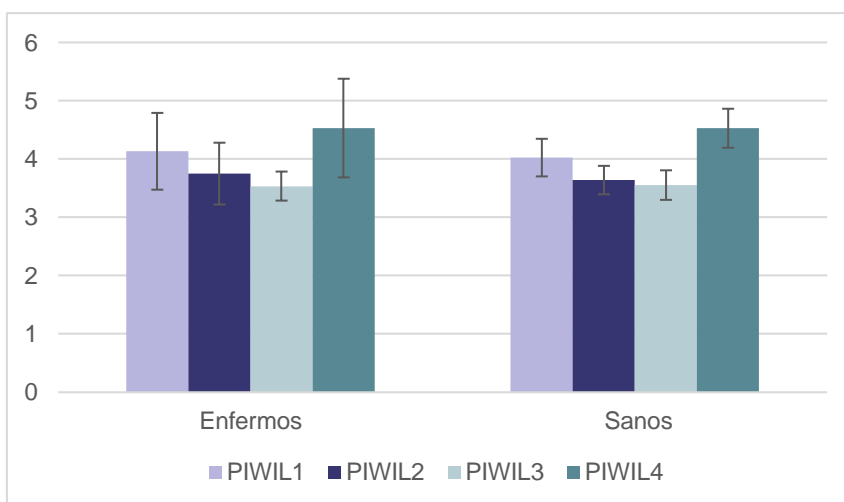


Figura A.17. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE75037.

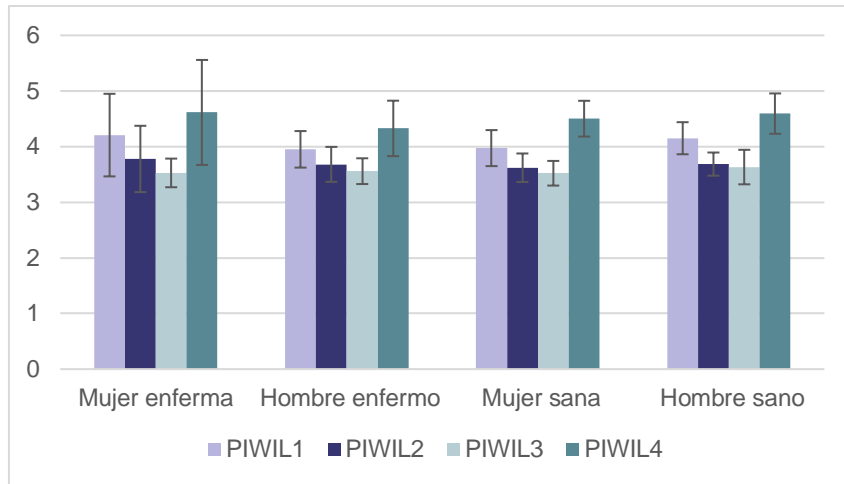


Figura A.18. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE75037.

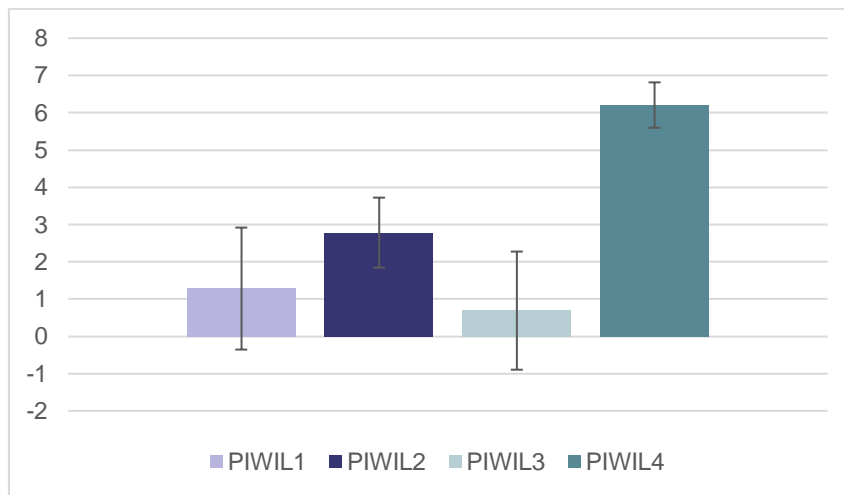


Figura A.19. Nivel de expresión de las proteínas PIWI en GSE81089.

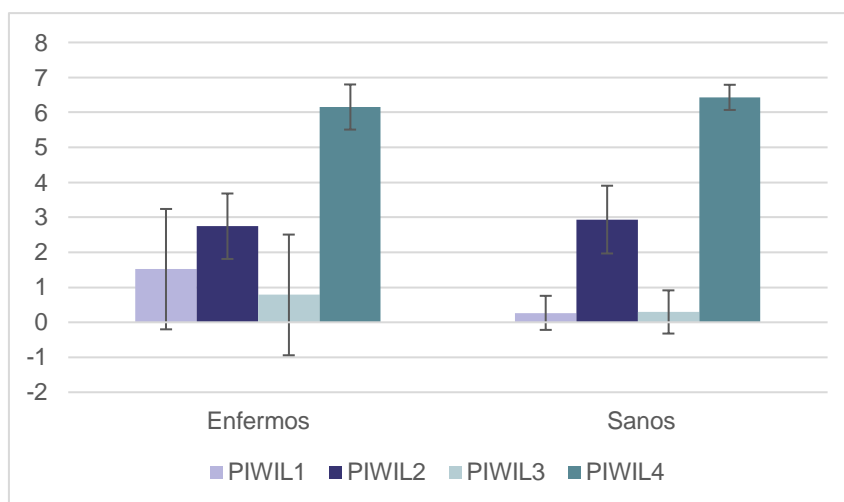


Figura A.20. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE81089.

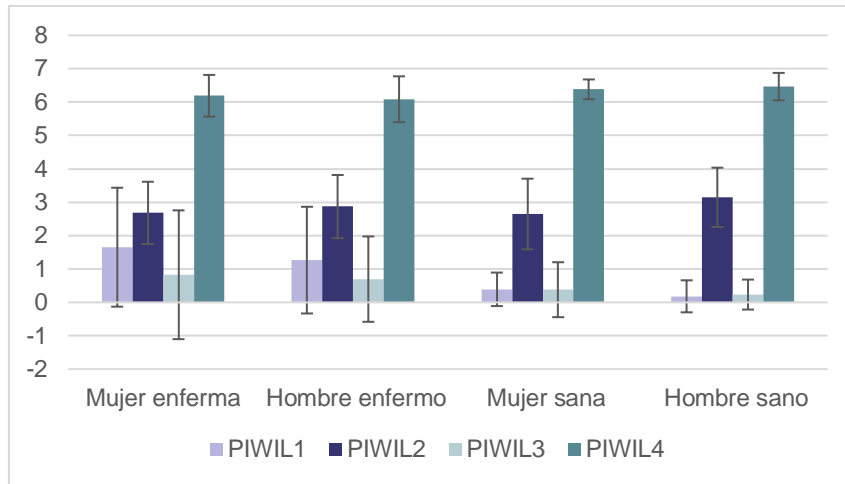


Figura A.21. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE81089.

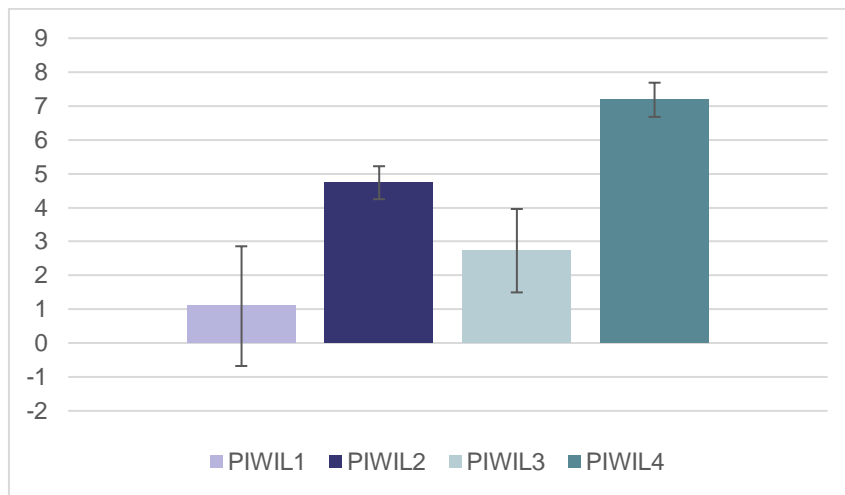


Figura A.22. Nivel de expresión de las proteínas PIWI en GSE87340.

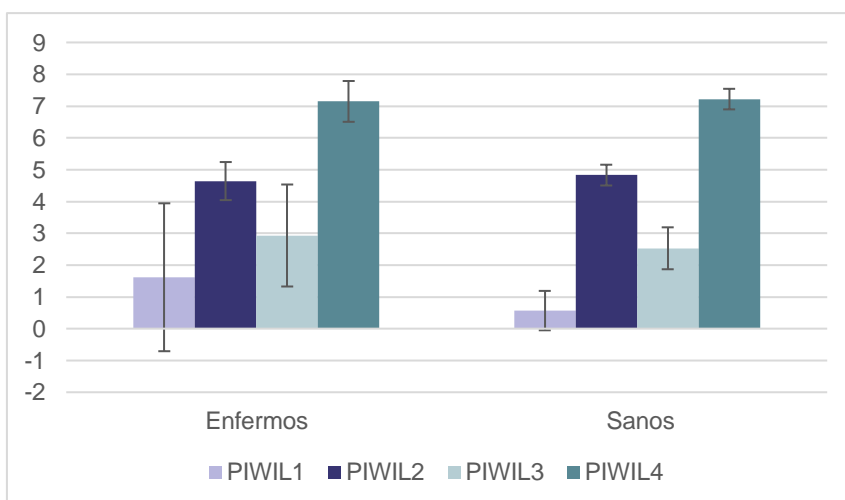


Figura A.23. Nivel de expresión de las proteínas PIWI según la distribución de individuos por enfermedad en GSE87340.

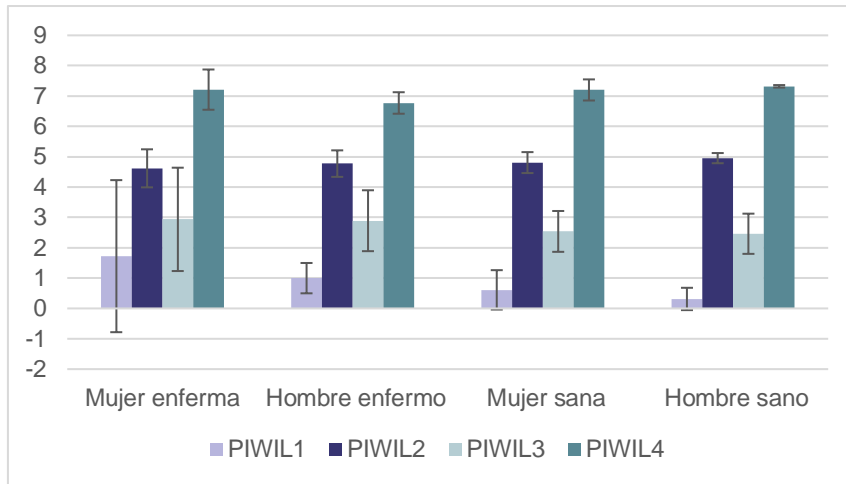


Figura A.24. Figura 59. Nivel de expresión de las proteínas PIWI según la distribución de individuos por sexo y enfermedad en GSE87340.