



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**Programa de Maestría y Doctorado en Ciencias Médicas,**  
**Odontológicas y de la Salud**

**ICES-Bioquímica Clínica**

IDENTIFICACIÓN DE GENES CANDIDATO EN MIELOMA  
MÚLTIPLE MEDIANTE ENFOQUES DE ESTRUCTURA-  
FUNCIÓN DE PROTEÍNAS E INTEGRACIÓN DE ÓMICAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRA EN CIENCIAS

PRESENTA

SHIRLEY BIKEL

Director de tesis:

**Dr. Adrián Ochoa Leyva**

Programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud

MEXICO, D.F.

Diciembre 2014



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# ÍNDICE

---

<b>1. INDICE</b>	<b>1</b>
<b>2. INDICE DE TABLAS, FIGURAS Y GRÁFICOS</b>	<b>2</b>
<b>3. RESUMEN</b>	<b>5</b>
<b>4. HOJA DE ABREVIATURAS</b>	<b>6</b>
<b>5. INTRODUCCIÓN</b>	<b>7</b>
<b>6. ANTECEDENTES</b>	<b>10</b>
<b>7. JUSTIFICACIÓN</b>	<b>13</b>
<b>8. HIPÓTESIS</b>	<b>14</b>
<b>9. OBJETIVO GENERAL</b>	<b>14</b>
<b>10. OBJETIVOS ESPECÍFICOS</b>	<b>14</b>
<b>11. MATERIALES Y MÉTODOS</b>	<b>16</b>
11.1. Identificación de las proteínas que unen elementos esenciales	18
11.2. Construcción de interactomas	19
11.3. Análisis estadístico	19
11.4. Análisis funcional	20
11.5. Análisis de estructura-función	20
11.6. Estimación del valor global de afectación para el análisis de estructura-función	23
11.7. Determinación de genes candidato	24
11.8. Análisis de los genes candidato clínicamente accionables	24
11.9. Planteamiento del método para definir la expresión absoluta de genes	24
11.10. Verificación del método de expresión absoluta	25
11.11. Aplicación del método de expresión absoluta	26
<b>12. RESULTADOS</b>	<b>27</b>
<b>12.1. Capítulo 1. Determinación de genes candidato en Mieloma Múltiple</b>	
12.1.1. Proteínas que unen elementos químicos esenciales (EE)	27
12.1.2. Interactomas en Mieloma Múltiple y análisis estadístico	30
12.1.3. Análisis funcional del interactoma	36
12.1.4. Análisis funcional de cada módulo del interactoma	39
12.1.5. Identificación de las proteínas afectadas en el interactoma	49
12.1.6. Integración de los resultados de los predictores de E-F en el valor global de afectación	50
12.1.7. Determinación de genes candidato	56
12.1.8. Caracterización de los nuevos genes candidato y su relación con Mieloma Múltiple	61
12.1.9. Genes candidato clínicamente accionables	65
<b>12.2. Capítulo 2. Análisis de la expresión absoluta de genes</b>	
12.2.1. Genes del cromosoma Y expresados en forma diferencial entre sexos	70
12.2.2. Expresión absoluta de genes constitutivos “Housekeeping”	74
12.2.3. Expresión absoluta de genes de levadura	76
12.2.4. Expresión absoluta de los genes candidato	78
<b>13. DISCUSIÓN</b>	<b>81</b>
<b>14. CONCLUSIONES</b>	<b>85</b>
<b>15. PERSPECTIVAS</b>	<b>87</b>
<b>16. BIBLIOGRAFÍA</b>	<b>88</b>
<b>17. ANEXOS</b>	<b>93</b>

## 2. ÍNDICE DE TABLAS, FIGURAS Y GRÁFICOS

---

### TABLAS

<b>Tabla 1.</b> Criterios de cada programa para predecir la afectación de la proteína	<b>22</b>
<b>Tabla 2.</b> Frecuencia de proteínas que unen EE en Cancer Gene Census	<b>28</b>
<b>Tabla 3.</b> Frecuencia de proteínas que unen EE en el genoma mutado de MM	<b>28</b>
<b>Tabla 4.</b> Lista de genes candidato obtenidos por la secuenciación de los 38 genomas de MM	<b>29</b>
<b>Tabla 5.</b> Distribución de los EE dentro del set de genes candidato en MM	<b>29</b>
<b>Tabla 6.</b> Propiedades de los interactomas de los diferentes elementos esenciales	<b>30</b>
<b>Tabla 7.</b> Funciones biológicas del interactoma del Zn	<b>38</b>
<b>Tabla 8.</b> Funciones biológicas del módulo 0	<b>39</b>
<b>Tabla 9.</b> Funciones biológicas del módulo 1	<b>41</b>
<b>Tabla 10.</b> Funciones biológicas del módulo 2	<b>42</b>
<b>Tabla 11.</b> Funciones biológicas del módulo 3	<b>42</b>
<b>Tabla 12.</b> Funciones biológicas del módulo 5	<b>43</b>
<b>Tabla 13.</b> Funciones biológicas del módulo 7	<b>44</b>
<b>Tabla 14.</b> Funciones biológicas del módulo 8	<b>45</b>
<b>Tabla 15.</b> Funciones biológicas del módulo 9	<b>46</b>
<b>Tabla 16.</b> Funciones biológicas del módulo 10	<b>46</b>
<b>Tabla 17.</b> Funciones biológicas del módulo 11	<b>47</b>
<b>Tabla 18.</b> Funciones biológicas del módulo 12	<b>48</b>
<b>Tabla 19.</b> Funciones biológicas del módulo 15	<b>49</b>
<b>Tabla 20.</b> Proteínas del Zinc con mutaciones de cambio de aminoácido y sus valores de predicción global de afectación correspondientes	<b>51-54</b>
<b>Tabla 21.</b> Nuevos genes candidato en Mieloma Múltiple	<b>57</b>
<b>Tabla 22.</b> Los nuevos genes candidato ordenados en base a los criterios para ser gen candidato	<b>60</b>
<b>Tabla 23.</b> Genes del cromosoma Y	<b>70</b>

<b>Tabla 24.</b> Número de sondas expresadas y no expresadas en el microarreglo humano hibridado con levadura	<b>77</b>
<b>Tabla 25.</b> Expresión de los genes candidato	<b>80</b>

---

## **FIGURAS**

<b>Figura 1.</b> Resumen de la metodología	<b>17</b>
<b>Figura 2.</b> Distribución de las proteínas que unen EE en cada set de datos genómicos	<b>27</b>
<b>Figura 3.</b> Interactoma de las proteínas mutadas con dominios de unión de Zinc	<b>31</b>
<b>Figura 4.</b> Resumen del análisis de estructura-función	<b>49</b>
<b>Figura 5.</b> Genes candidato en Mieloma Múltiple	<b>56</b>
<b>Figura 6.</b> Estructura de la proteína ESR1	<b>66</b>
<b>Figura 7.</b> Estructura de la proteína IDE	<b>68</b>

---

## **GRÁFICOS**

<b>Gráfico 1.</b> Conectividad global del componente más grande del interactoma del Zinc	<b>32</b>
<b>Gráfico 2.</b> Conectividad global del interactoma del Zinc Vs interactomas de 103 proteínas generados de la lista de proteínas de Mieloma Múltiple	<b>33</b>
<b>Gráfico 3.</b> Conectividad global del interactoma del Zinc comparado con otras listas de genes	<b>34</b>
<b>Gráfico 4.</b> Análisis de correlación entre el tamaño de muestra y la conectividad global	<b>35</b>
<b>Gráfico 5.</b> Conectividad global de los interactomas de proteínas Zn no mutadas Vs el interactoma de las proteínas Zn mutadas en MM	<b>36</b>
<b>Gráfico 6.</b> Correlación entre los valores globales de afectación y número de predictores	<b>50</b>
<b>Gráfico 7.</b> Comparación del valor global de afectación de las proteínas afectadas y no afectadas	<b>55</b>
<b>Gráfico 8.</b> Comparación de los valores globales de afectación	<b>59</b>
<b>Gráfico 9.</b> Comparación de los valores de expresión de los genes del cromosoma Y entre sexos	<b>71</b>
<b>Gráfico 10.</b> Expresión de genes del cromosoma Y en hombres y mujeres con Mieloma Múltiple	<b>72</b>

<b>Gráfico 11.</b> Histograma de la fluorescencia de los genes del cromosoma Y en microarreglos de mujeres con Mieloma Múltiple	<b>73</b>
<b>Gráfico 12.</b> Histograma de la fluorescencia de los genes del cromosoma Y en microarreglos de hombres con Mieloma Múltiple	<b>73</b>
<b>Gráfico 13.</b> Expresión de los genes constitutivos “Housekeeping” y del SRY según el punto de corte calculado	<b>74</b>
<b>Gráfico 14.</b> Expresión de genes del cromosoma Y, Housekeeping y SRY según el punto de corte	<b>75</b>
<b>Gráfico 15.</b> Determinación de la expresión de las sondas de microarreglo humano (HGU133 Plus2) en 5 muestras de levadura según el punto de corte	<b>76</b>
<b>Gráfico 16.</b> Comparación de las fluorescencias entre el set de sondas no expresadas, sondas expresadas ortólogas y sondas expresadas no ortólogas	<b>77</b>
<b>Gráfico 17.</b> Cálculo del punto de corte a partir de los microarreglos de mujeres con Mieloma Múltiple del MMRC	<b>79</b>

### 3. RESUMEN

**Introducción:** La aplicación de las tecnologías ómicas en la investigación del cáncer ha generado millones de datos que pueden ser analizados de una manera integrada para responder a una multitud de preguntas sobre las alteraciones moleculares asociadas con dicha enfermedad. Sin embargo, el poder distinguir las mutaciones candidato de las mutaciones pasajeras sigue siendo un importante reto en la genómica del cáncer. El estudio de las ómicas de manera integrada y las aproximaciones de sistemas pueden ayudarnos a entender aspectos claves dentro de la investigación del cáncer en diferentes niveles biológicos de información pero de una manera integrada. Estos puntos claves incluyen mutaciones a nivel de ADN, perfiles de expresión de ARN, modificaciones en la estructura-función de proteínas, asociación entre genotipo y fenotipo, contenido metabólico y redes de interacción intermoleculares. En este trabajo, se presenta un análisis basado en la integración de datos provenientes de diferentes plataformas ómicas y de estructura-función de proteínas mutadas en el genoma de Mieloma Múltiple. **Objetivo:** Identificar genes candidato en Mieloma Múltiple mediante un análisis integrado de ómicas con estructura-función de proteínas. **Métodos:** De los genes mutados en 38 genomas de Mieloma Múltiple obtenidos del Consorcio de investigación de Mieloma Múltiple (MMRC), seleccionamos aquellos cuyas proteínas unen elementos químicos esenciales (Zn, Ca, Cl, Cu, Fe, K, Mg, Mn y Na). Con estos genes se construyó una red de interacción proteína-proteína para cada uno de los elementos y analizamos su significancia estadística y biológica. A partir del interactoma que resultó ser estadística y biológicamente significativo, determinamos el impacto funcional de las mutaciones en las proteínas que lo conforman a través de diferentes herramientas bioinformáticas. Finalmente, definimos como candidato a aquellos genes de proteínas del interactoma que están afectadas en su estructura-función en dos o más individuos con Mieloma y/o que tienen una gran cantidad de interacciones proteína-proteína (proteínas HUB). **Resultados:** Sólo obtuvimos un interactoma estadísticamente significativo con las proteínas mutadas que unen Zinc. A través del análisis integral de esta red encontramos 28 genes candidato de los cuales 18 no han sido reportados como tales en Mieloma Múltiple, 5 ya han sido reportados como candidato en otros tipos de cánceres y otros 5 como candidato en mieloma. **Conclusiones:** El 35% de nuestros genes candidato ya han sido reportados como tales en Mieloma sugiriendo que nuestra estrategia es viable para descubrir genes candidato nuevos. Tres de los nuevos genes candidato son clínicamente accionables lo que abre la posibilidad de explorar nuevas alternativas terapéuticas.

## 4. HOJA DE ABREVIATURAS

**Ca:** Calcio

**CGC:** Cancer Gene Census

**Cl:** Cloro

**Cu:** Cobre

**EE:** Elementos esenciales

**Fe:** Hierro

**IgA:** Inmunoglobulina A

**IgD:** Inmunoglobulina D

**IgG:** Inmunoglobulina G

**IgM:** Inmunoglobulina M

**K:** Potasio

**Mg:** Magnesio

**Mn:** Manganeseo

**MMRC:** Multiple Myeloma Research Consortium

**MM:** Mieloma Múltiple

**Mts:** Metalotioneínas

**MO:** Médula ósea

**Na:** Sodio

**OMIM:** Online Mendelian Inheritance in Man

**Zn:** Zinc



## 5. INTRODUCCIÓN

El Mieloma Múltiple (MM) es una neoplasia de células plasmáticas que inicia y cursa con proliferación de plasmocitos tumorales en médula ósea (MO). Está caracterizada por la infiltración de dichas células malignas a MO y otros tejidos, causando destrucción ósea y/o daño a órganos blanco. Los síntomas comunes a la presentación de la enfermedad son: fatiga, dolor óseo e infecciones recurrentes [1-a, 2-a]. La frecuencia de esta enfermedad de acuerdo con la OMS corresponde a aproximadamente el 1% de todas las neoplasias malignas y del 10 al 15% de las neoplasias hematológicas [1-a]. De acuerdo a las Guías Mexicanas de Mieloma Múltiple, en nuestro país corresponde del 4.2 al 7.7% de las enfermedades oncohematológicas y el 90% de los casos se presenta en pacientes mayores de 50 años [1-a]. Esta enfermedad es más frecuente en hombres presentando una mayor incidencia en raza negra [1-a]. La edad media de presentación es de 62 años y suele ser precedida por una condición pre-maligna progresiva con la edad definida como Gammapatía Monoclonal de Significado Incierto (siglas en inglés “MGUS”), la cual se presenta en nuestro país de manera más frecuente en personas mayores de 50 años, con un riesgo de transformación a Mieloma Múltiple del 1% por año [1-a]. El diagnóstico de Mieloma Múltiple se realiza por medio de aspirado de médula ósea, siendo positivo cuando se observan más del 10% de células plasmáticas anormales, cuando en sangre se detecta elevación de la proteína monoclonal (proteína M) (en el 55-60% de los casos pertenece al subtipo IgG, 20% al IgA, 1% al IgD y 1% al IgM) y por la presencia de la proteína de Bence Jones en orina junto a datos clínicos característicos de daño a órganos blanco, los cuales involucran anemia, lesiones líticas en la serie ósea, calcio sérico elevado y falla renal [2-a]. Por muchos años los recursos terapéuticos disponibles han sido muy limitados, considerándose hasta el momento una enfermedad incurable a pesar de los tratamientos farmacológicos más novedosos [1-a, 2-a]. Generalmente, la sobrevida global en los pacientes va de 29 a 62 meses dependiendo del grupo pronóstico al que pertenezcan [1-a]. Tomando en cuenta las tendencias nacionales del incremento de la población adulta mayor, se podría inferir que la incidencia de Mieloma Múltiple se incrementará durante los próximos años en México [1-a].

Esfuerzos significativos han sido dirigidos hacia la identificación de eventos genéticos moleculares dentro de esta enfermedad con el objetivo de lograr una detección temprana y proveer nuevos blancos terapéuticos, sin embargo, su patogénesis sólo está parcialmente entendida [2] [3]. El proceso de carcinogénesis en el humano es un conjunto de múltiples eventos que involucran alteraciones genéticas que dirigen la transformación de células humanas normales hacia eventos altamente malignos [4]. Lamentablemente, el diagnóstico de cáncer generalmente es realizado en estadios avanzados de la

enfermedad, otorgando pocas posibilidades de aplicar una terapia curativa [5] [6]. Esto resalta la necesidad de la integración de las ciencias biomédicas con una atención médica multidisciplinaria para descubrir nuevos blancos moleculares del cáncer, los cuales puedan contribuir en la predicción y/o prevención de su desarrollo y también para definir blancos terapéuticos enfocados a cada paciente de manera personalizada [7]. Generalmente, las células del cáncer exhiben mutaciones genómicas y perfiles de expresión de ARN alterados de cientos de genes [9] [10]. Las alteraciones en la expresión del ARN y las mutaciones en las proteínas pueden perturbar las interacciones proteína-ADN, proteína-ARN y proteína-proteína. Como consecuencia, las interacciones afectadas entre biomoléculas resultan en aberraciones de las vías de señalización y regulación celular que pueden conllevar al fenotipo de la enfermedad [11]. Por lo tanto, el cáncer no solo es una enfermedad de células sino también una enfermedad de varios sistemas que interactúan a nivel molecular y celular y que conllevan al inicio y progresión de la enfermedad, por lo que los cambios en las interacciones de estos sistemas en un ambiente dinámico destacan su complejidad [12]. En este sentido, las alteraciones en cáncer a nivel de ADN, ARN, proteínas y metabolitos requieren del desarrollo de metodologías novedosas tanto experimentales como bioinformáticas para el análisis de los cambios biológicos en todas estas biomoléculas de una manera integrada.

Con el advenimiento de las tecnologías de secuenciación masiva, una gran cantidad de información ha sido generada, pero el estudio e integración de los datos generados es todavía un reto [13]. Varias aproximaciones han sido desarrolladas para distinguir los genes candidato, que son los que le dan cierta ventaja adaptativa a la enfermedad, de los genes pasajeros (“passengers”) [14]. Una de las más comunes es la búsqueda de genes que han acumulado un mayor número de mutaciones con respecto a los demás genes del genoma secuenciado [14] [15] [16]. Recientemente, los esfuerzos para identificar genes candidato también se han centrado en identificar a aquellos genes que están enriquecidos de mutaciones y que pertenecen a redes de genes construidas de estudios de co-expresión, los cuales involucran ciertas vías metabólicas y/o de señalización, y vías de interacción proteína-proteína ya conocidas [14] [15] [16] [17] [11] [18]. Actualmente, se ha sugerido que la mayoría de las mutaciones candidato conocidas en el cáncer afectan ampliamente la estructura y/o función de las proteínas involucradas [19] y que la desestabilización de la estructura tridimensional es uno de los principales mecanismos moleculares llevados a cabo por las mutaciones de genes candidato [19]. Por lo tanto, se sugiere que la priorización de variantes funcionales de acuerdo a sus efectos deletéreos sobre la estructura-función de una proteína es esencial para el proceso de búsqueda de genes candidato.

Una de las aproximaciones que está tratando de descifrar la red compleja de relaciones biológicas es la interactómica, la cual consiste en el estudio del conjunto de interacciones entre distintas biomoléculas en un entorno determinado, centrada principalmente en el estudio de interacciones entre proteínas, pero también basada en el estudio de las interacciones proteína-ADN, proteína-ARN, proteína-metabolitos, ARN-ADN, entre otras [20]. Al conjunto de estas interacciones se le denomina “interactoma” y su análisis permite, por un lado, identificar listas globales de biomoléculas que participan de forma conjunta en determinados procesos celulares [21] y, por el otro, generar una comprensión detallada de las diferentes interacciones involucradas para traducirlas a procesos biológicos y fisiopatológicos. Debido a que el cáncer puede alterar el funcionamiento normal de las proteínas que participan en la correcta interconexión de este tipo de redes funcionales, resulta de gran interés su estudio desde un análisis multidimensional el cual integre la información proveniente de las distintas tecnologías “ómicas” y con ello poder abordar el problema desde una perspectiva de sistemas [7] [22] [17] [11] [23] [18] [24]. En la actualidad, las ciencias biomédicas se están enfrentando a la necesidad de nuevas aproximaciones, una transformación del reduccionismo hacia un paradigma holístico, desde una molécula hacia los niveles poblacionales [7]. Las aproximaciones ómicas integradas usan métodos experimentales y computacionales para caracterizar los sistemas biológicos a niveles moleculares, celulares, de tejido y de población [25] [22]. En este sentido, el estudio a nivel de metales, dentro de los cuales se encuentran la metalómica y la metaloproteómica, ha comenzado a tomar un lugar dentro de las aproximaciones ómicas para entender procesos biológicos complejos como el cáncer desde una perspectiva de sistemas [25-a] [25-b]. Esto debido a que tanto los metales como las proteínas que tienen motivos de unión a metales (metaloproteínas) están involucrados en muchas funciones celulares importantes [25-a] [25-b]. Además, la unión a elementos esenciales (Ca, Na, Zn, Fe, etc) por estos motivos proteicos es una de las funciones más antiguas que ha sido conservada durante la evolución de las proteínas, lo que destaca su importancia funcional para la viabilidad celular [25-a] [25-b]. Por lo tanto, abordar el estudio del Mieloma Múltiple desde un enfoque que incluya este tipo de análisis integrados nos permitirá estudiar su complejidad desde otra perspectiva, la cual implica el definir los componentes del sistema, determinar cómo esos componentes interactúan entre sí y delinear la dinámica de los mismos en el espacio y en el tiempo, los cuales enmarcan sus necesidades biológicas. Para ello, planteamos por un lado el análisis e integración de datos experimentales provenientes del genoma y del transcriptoma de personas con Mieloma Múltiple, y por otro lado el análisis de estructura y/o función de las proteínas con motivos de unión a elementos esenciales y que están mutadas como consecuencia de este cáncer, así como el análisis del interactoma

(proteína-proteína) que ellas conforman mediante su inferencia a través de diversos algoritmos computacionales validados experimentalmente.

## 6. ANTECEDENTES

En los últimos años, las tecnologías de última generación han expandido el número de moléculas candidato en diferentes tipos de cánceres [26] [27] [28] [29]. Sin embargo, el descubrimiento de estas listas de nuevas moléculas ha estado limitado al análisis de las ciencias ómicas (la genómica, transcriptómica, proteómica, metabolómica o interactómica) de manera independiente. El desarrollo del cáncer resulta de la perturbación de diferentes vías de señalización, de redes de interacción proteína-proteína, y de vías de expresión, entre otras, por lo que el estudio de los componentes celulares alterados y de las vías afectadas relacionadas podría ser un factor clave para el entendimiento de la fisiopatología de esta enfermedad[7]. En este sentido, la investigación médica se ha transformado con el desarrollo de los estudios basados en los análisis de las tecnologías ómicas y en las aproximaciones de sistemas, los cuales han permitido el descubrimiento de vías, redes de interacción y procesos celulares involucrados en células normales y cancerosas, a través del análisis simultáneo de miles de componentes moleculares a diferentes niveles de organización biológica [22]. Bajo este contexto, varios trabajos de investigación han realizado análisis integrados que involucran datos de las diferentes ómicas para finalmente detectar genes candidato. Por ejemplo, se han secuenciado genomas completos a partir de los cuales fueron comparados el tejido enfermo con respecto al sano para luego obtener la lista de genes mutados, los cuales han sido estudiados en modelos de redes de interacción (interactomas) para finalmente obtener las vías biológicas implicadas en el desarrollo del cáncer [30]. Otros estudios, como el de Zhou D. et al [31] realizaron un análisis similar a partir del transcriptoma, a través del cual fueron obtenidas las listas de genes que se encontraban sub o sobre expresados en el cáncer con respecto al tejido sano, con los cuales formaron redes de co-expresión y evaluaron cuáles funciones o vías celulares se estaban activando o desactivando en esta enfermedad. Con la finalidad de diferenciar a los genes candidato de los pasajeros, muchos estudios se han basado en diferentes aproximaciones computacionales enfocadas principalmente en el descubrimiento de los genes que presentan una mayor frecuencia de mutaciones con respecto a la distribución de todas las mutaciones en el genoma [2, 14, 29, 32]. El estudio de Chapman et al [2] es un ejemplo de ello, en donde se utilizó este tipo de análisis para descubrir genes candidato en Mieloma Múltiple. El razonamiento de esta aproximación para seleccionar determinados genes como candidato es que dichos genes pueden estar siendo seleccionados positivamente durante la evolución del tumor, esto debido a la alta frecuencia de

mutaciones que ellos presentan [33, 34]. Otra aproximación diseñada para identificar mutaciones biológicamente relevantes en cáncer ha sido el análisis de la frecuencia de mutaciones en determinados grupos de genes. Esta aproximación involucra el enriquecimiento en mutaciones en grupos de genes que integran vías o redes génicas y/o procesos moleculares específicos los cuales son relevantes para el desarrollo del cáncer [2] [35]. Otras aproximaciones han involucrado el uso de predicciones del efecto de la mutación sobre la estructura-función de la proteína mutada [19] [14]. Estos enfoques han sido ampliamente utilizados para clasificar las mutaciones somáticas en cáncer por su probabilidad de ser candidato en base al nivel de daño que tienen las proteínas con el supuesto de que el conjunto de las mutaciones de alto impacto funcional resalta a los posibles genes candidato [19]. El propósito de esta clasificación ha sido dar prioridad a las mutaciones para su posterior validación experimental. Por ejemplo, Ding et al utilizó los predictores SIFT [36] y PolyPhen [37] [38] para evaluar el impacto de 811 mutaciones de cambio de aminoácido sobre las proteínas. En conjunto, estos programas identificaron 579 mutaciones con consecuencias funcionales, las cuales fueron priorizadas para posteriores estudios funcionales [39]. Los estudios basados en la priorización de la estructura-función de los genes mutados no solo han permitido conocer cuáles mutaciones podrían estar afectando la funcionalidad de la proteína sino también entender la consecuencia de dicha alteración sobre redes de interacción proteína-proteína, proteína-ADN, y las vías bioquímicas y de señalización involucradas [18]. Respecto a ello, Won et al. desarrolló una nueva aproximación que integra múltiples predictores para el análisis de estructura-función para priorizar variantes de acuerdo a su potencial deletéreo [40]. Esta aproximación también integró información de las redes de interacción de proteínas y perfiles de expresión de tejidos para priorizar genes [40]. Ellos encontraron que el análisis de un grupo de genes con mutaciones predichas como deletéreas a través del análisis de estructura-función mostraron tener un enriquecimiento significativo en vías asociadas al desarrollo del cáncer. Esto sugiere que los cambios genéticos deletéreos a nivel de vías celulares podrían ser necesarios para el desarrollo del cáncer y para entender la heterogeneidad inter-tumoral a nivel del gen mutado [40]. De manera similar, el estudio de Nishi et al, mapeó y analizó todas las proteínas con mutaciones de cambio de aminoácido del glioblastoma sobre un interactoma de casi todas las proteínas humanas, demostrando que las mutaciones que se encontraban en las interfaces de interacción ocasionan cambios más drásticos sobre las propiedades fisicoquímicas de los aminoácidos con respecto a las mutaciones encontradas fuera de las interfaces de interacción [43]. En el mismo sentido, se ha observado que la mayoría de las mutaciones de los genes candidato en distintos tipos de cáncer causan un fuerte impacto en la estructura de las correspondientes proteínas [19], sugiriendo que la desestabilización de la estructura

tridimensional de las proteínas puede ser uno de los mecanismos moleculares mayormente relacionados con las mutaciones funcionales en los genes candidato [19]. Recientemente, los estudios de las interacciones proteína-proteína desde el punto de vista del análisis de los interactomas han abierto la oportunidad para la identificación de nuevos genes candidato en enfermedades complejas como el cáncer [44] [45] [18] [21]. Para ello, se han construido diversas redes de interacción proteína-proteína a partir de la lista de genes mutados, de la lista de los transcritos sobre/sub-expresados y de la lista de las proteínas que presentan variaciones en el número de copias [46] [27] [11] [10]. Estos análisis han mostrado que las proteínas que forman los nodos centrales de las redes de interacción (HUBs) tienden a estar codificadas por genes esenciales para la fisiología celular [47]. Además, se ha observado que la eliminación de los genes que codifican HUBs conduce a una gran cantidad de características fenotípicas a diferencia de los no HUBs [48]. Esto ha sugerido que los HUBs podrían estar representando a genes posiblemente candidato para distintas enfermedades [44] [45] [21]. Por ejemplo, en el estudio de Jonsson et al, se construyó un interactoma a partir de todo el genoma humano con el cual se demostró que proteínas humanas codificadas por genes conocidos como candidato en cáncer tienen una mayor cantidad de interacciones dentro del modelo de interactoma que aquellas proteínas que todavía no han sido documentadas como mutadas en cáncer [44]. También ha sido reportado que algunas de las proteínas implicadas en enfermedades mendelianas poseen un mayor número de interacciones proteína-proteína que las que no están involucradas en dichas enfermedades [46]. Estas observaciones sugieren que mediante la construcción del interactoma de una lista de genes de interés para una enfermedad, se puede encontrar a otros genes que están relacionados con la misma enfermedad por estar involucrados dentro de los mismos módulos funcionales o patogénicos construidos a partir de redes de interacción proteína-proteína.

## 7. JUSTIFICACIÓN

Uno de los retos en el estudio de la biología de cualquier tipo de cáncer lo constituye el análisis e integración de la gran cantidad de datos obtenidos por las tecnologías “ómicas” [7] [25] [22] y con ello poder identificar y diferenciar a los genes candidato de los genes pasajeros. Sin embargo, la mayoría de las mutaciones suelen estar en pocos genes y existe una gran cantidad de genes en los cuales las mutaciones se han encontrado de manera ocasional o sólo en una muestra, complicando aún más el análisis del papel de las mutaciones en el cáncer. A partir de los perfiles de mutación es complicado conocer qué genes están implicados en la patogenia de la enfermedad, además de que cada paciente con el mismo tipo de cáncer suele tener un espectro de mutaciones muy diferente [2]. Debido a esta complejidad biológica, el descubrimiento de nuevos genes candidato implica la necesidad del desarrollo de nuevas herramientas y estrategias tanto experimentales como de análisis masivo de datos que permitan extraer la significancia biológica y traducirla a una interpretación molecular aplicable a la patogenia del cáncer [7] [25] [17] [23]. Por lo tanto, resulta de gran interés el estudio del Mieloma Múltiple, así como de otros tipos de cáncer, desde un análisis multidimensional el cual integre la información proveniente de distintas tecnologías “ómicas” y con ello poder abordar el problema desde una perspectiva de sistemas que pueda ofrecer nuevas alternativas para el estudio y entendimiento de las enfermedades comunes.

## **8. HIPÓTESIS**

Las mutaciones en las proteínas que unen elementos esenciales y/o la modificación de su expresión afectan su función conllevando a la alteración de sus redes de interacción proteína-proteína y al descubrimiento de nuevos genes candidato en Mieloma Múltiple.

## **9. OBJETIVO GENERAL**

Identificar nuevos genes candidato en la patogenia del Mieloma Múltiple mediante la integración de un análisis basado en las interacciones proteína-proteína, la estructura-función de proteínas y la expresión de genes utilizando aproximaciones de integración de datos provenientes de las tecnologías ómicas.

## **10. OBJETIVOS ESPECÍFICOS**

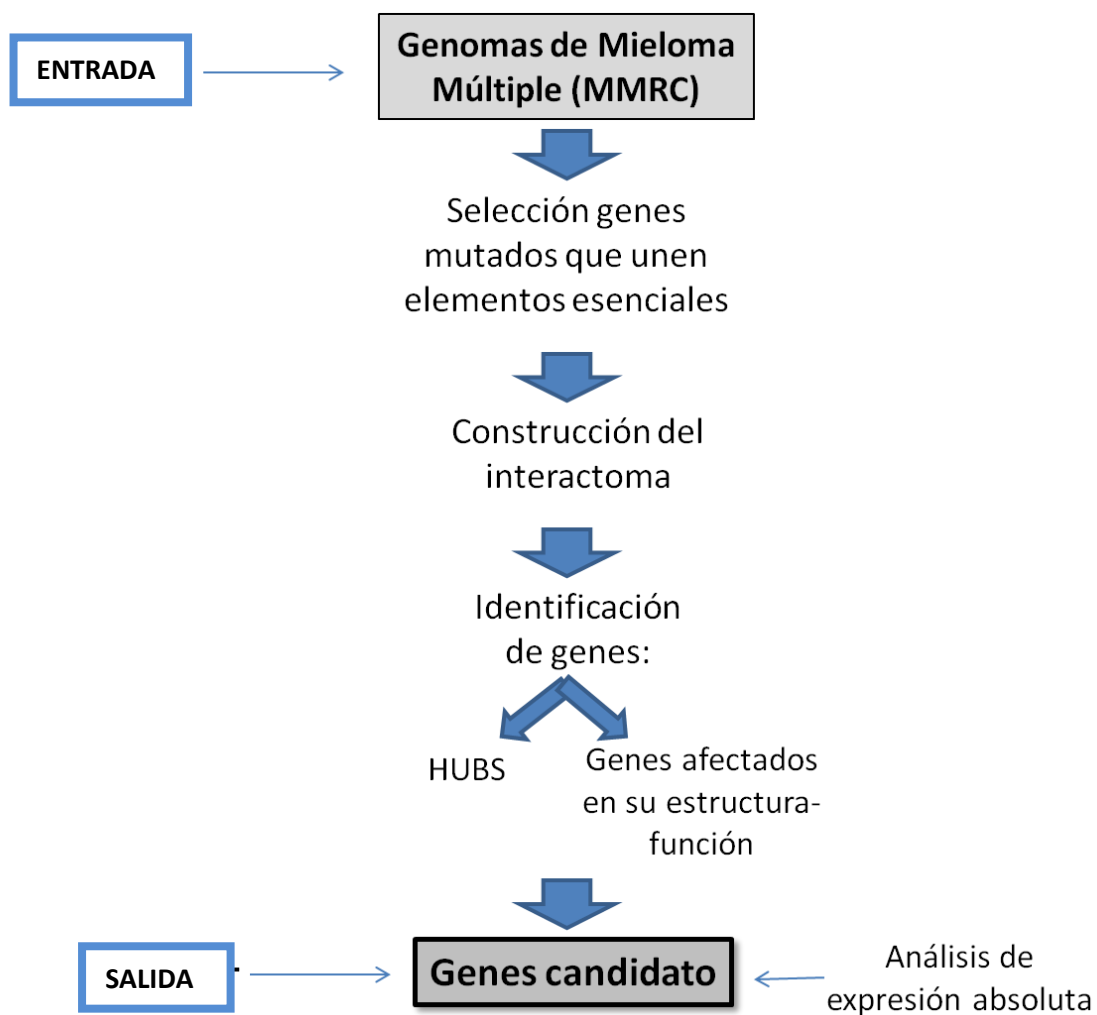
- a) De los genes ya identificados como mutados en los 38 genomas disponibles de pacientes con Mieloma Múltiple del MMRC, seleccionar aquellos cuyas proteínas posean una anotación funcional por interactuar físicamente con los elementos químicos esenciales: Ca, Zn, Cl, Cu, Fe, K, Mg, Mn y Na.
- b) Seleccionar las proteínas que tienen la capacidad de unir elementos esenciales y construir con ellas las redes de interacción proteína-proteína que pudieran estar formando (interactomas de cada elemento esencial).
- c) Analizar la significancia estadística de los interactomas.
- d) Analizar cuáles son las funciones celulares, vías bioquímicas y de señalización identificadas por las proteínas que componen el interactoma y los módulos del mismo.
- e) Predecir mediante programas de biología estructural y biología evolutiva el efecto de las mutaciones en la estructura y/o función de las proteínas que conforman el interactoma.
- f) Desarrollar un método que nos permita identificar la expresión absoluta de cada uno de los genes que representan las proteínas del interactoma.
- g) Integrar la información para cada proteína procedente de cada uno de los análisis de estructura-función y de la expresión absoluta de los respectivos genes que representan las proteínas del interactoma del Mieloma Múltiple.
- h) Identificar los módulos funcionales dentro del interactoma que pudieran estar siendo alterados por las mutaciones y/o cambios de su expresión absoluta para cada paciente.



- i) Definir a los genes candidato por su afectación en estructura-función y/o por su condición de HUB en el interactoma.
- j) Identificar potenciales fármacos capaces de regular la función de los genes candidato.

## 11. MATERIALES Y MÉTODOS

Se inició con la identificación *in silico* de los genes cuyas proteínas interactúan con los siguientes elementos esenciales: Zn, Ca, Cl, Cu, Fe, K, Mg, Mn y Na. Esto debido a que la unión a elementos esenciales por motivos proteicos es una de las funciones más antiguas que ha sido conservada durante la evolución de las proteínas y muchas funciones celulares importantes están reguladas por proteínas que poseen este tipo de dominios de unión. Por lo tanto, nuestro análisis comprende una primera sección basada en la selección de genes mutados que codifican proteínas con motivos de unión a elementos esenciales. Dichos genes provienen de 38 pacientes con Mieloma Múltiple del Consorcio de investigación del Mieloma Múltiple (MMRC) [2]. Además, esta sección involucra la construcción de una red de interacciones proteína-proteína (interactoma) a partir de la lista de proteínas mutadas, y la determinación de la significancia estadística de esta red, así como la caracterización de las funciones biológicas de los genes que componen dicho interactoma. A partir de los análisis de estructura-función de las proteínas mutadas que conforman el interactoma, se predice cuáles se encuentran afectadas y se identifican los módulos funcionales que pudieran estar alterados por dichas mutaciones. Por otro lado, también se determinó qué proteínas del interactoma son HUB y se identificaron los posibles genes candidato de la enfermedad a partir de la integración de dichos análisis. Finalmente, la primera sección concluye con la identificación de posibles fármacos potenciales que sean capaces de regular la función de los genes candidato. La segunda sección de nuestro trabajo consiste en el desarrollo de un método que nos permita determinar la expresión absoluta de los genes candidato, es decir, que sea capaz de discriminar los genes candidato expresados de los no expresados, utilizando la información proveniente de los microarreglos de expresión de los pacientes con Mieloma Múltiple del MMRC. La finalidad del análisis de expresión absoluta es, por un lado, poder predecir si las proteínas candidato afectadas se encuentran o no expresadas, considerando que si una proteína afectada se expresa podría estar afectando la función biológica en la que está involucrada y, por el otro, conocer si las proteínas candidato que no se encuentran afectadas en su estructura-función, pudieran estar afectando las interacciones proteína-proteína en las que participan debido a cambios en su expresión (figura 1).



**Figura 1. Resumen de la metodología.** Análisis realizado a partir de los genomas de Mieloma Múltiple para descubrir los genes candidato de la enfermedad.

### 11.1. Identificación de las proteínas que unen elementos esenciales

Cada una de las proteínas producto de los genes mutados en los pacientes con MM se anotó funcionalmente con UniProtKB [49]. En base a estas anotaciones, se seleccionaron todas aquellas proteínas que tienen una anotación funcional de interacción física con los siguientes elementos esenciales: Zn, Ca, Cl, Cu, Fe, K, Mg, Mn y Na. Para poder identificar dichas proteínas, dentro de la lista de las anotaciones de UniProtKB se realizó la búsqueda de elementos esenciales a través de las siguientes descripciones: las que unen Zinc se identificaron mediante la siguiente búsqueda de términos: "(organism:9606 keyword:1185) AND reviewed:yes AND (go:"Zinc ion binding [0008270]" or keyword:"Zinc [KW-0862])"; las que unen calcio: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"calcium ion binding [0005509]" or keyword:"Calcium [KW-0106]"); las que unen cloro: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"chloride ion binding [0031404]" or keyword:"Chloride [KW-0868]"); las que unen cobre: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"copper binding [0005507]" or keyword:"Copper [KW-0186]"); las que unen hierro: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"iron binding [0005506]" or keyword:"Iron [KW-0408]"); las que unen potasio: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"potassium ion binding [0030955]" or keyword:"Potassium [KW-0630]"); las que unen magnesio: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"magnesium binding [0000287]" or keyword:"Magnesium [KW-0460]"); las que unen manganeso: (organism:9606 keyword:1185) AND reviewed:yes AND (go:"manganese binding [0030145]" or keyword:"Manganese [KW-0464])" y las que unen sodio: (organism:9606 keyword:1185) AND reviewed:yes AND keyword:"Sodium [KW-0915]".

Para poder definir si el genoma de Mieloma Múltiple se encuentra enriquecido en proteínas mutadas que unen elementos esenciales, se analizó qué proporción de las proteínas de todo el proteoma humano tienen dominios de unión a dichos elementos esenciales. Para poder identificar las proteínas que unen elementos esenciales en el proteoma humano, se realizó la búsqueda de elementos esenciales en UniProtKB usando los mismos términos que fueron descritos en el párrafo anterior. En base a la distribución de proteínas que unen elementos esenciales en el proteoma humano se calculó la distribución de cuántas proteínas mutadas se esperaba encontrar en el genoma de Mieloma Múltiple para cada uno de los elementos esenciales basados en su distribución normal en el proteoma humano. La cantidad esperada de proteínas mutadas para cada elemento esencial fue comparada con respecto a la cantidad de proteínas mutadas observadas para su respectivo elemento esencial. Finalmente se calculó la diferencia que hay entre la cantidad de proteínas observadas y esperadas en el genoma de

Mieloma Múltiple por cada elemento esencial para determinar si se encuentran sub o sobre-enriquecidas. Además, se analizó si existe un enriquecimiento de proteínas que unen elementos esenciales dentro del set de proteínas que ya se han identificado como candidatas en diferentes tipo de canceres (Cancer Gene Census) [50]. Para ello, dichas proteínas fueron anotadas funcionalmente con UniProtKB, y se identificaron cuáles y cuántas unen elementos esenciales usando la estrategia descrita anteriormente.

## **11.2. Construcción de interactomas**

Se construyeron redes de interacción (interactomas) a través del mapeo de las proteínas mutadas que unen elementos esenciales contra el Human Interaction Network (HIN), el cual consiste en un set de genes del genoma humano codificadores de proteínas derivados de cuatro fuentes de datos que incluyen el Human Protein Reference Database (HPRD), Reactome, NCI-Nature Pathway Interaction (PID) Database, y el MSKCC Cancer Cell Map, cuyas interacciones involucran interacciones proteína-proteína y vías de señalización basadas en resultados experimentales. A través del programa NetBox (<http://cbio.mskcc.org/netbox>), se construyeron los diferentes interactomas.

## **11.3. Análisis estadístico**

La significancia estadística de cada interactoma fue analizada a través de los modelos estadísticos “global null” y “local null”. El modelo “global null” permite evaluar el nivel de conectividad global de la red, para lo cual se compara el tamaño del componente más grande de la red con respecto al tamaño de cada uno de los componentes más grandes generados a partir de listas al azar de genes presentes en el interactoma humano (HIN por sus siglas en inglés: Human Interaction Network) [18]. El tamaño del componente más grande es determinado en base al número de proteínas totales (nodos) más el número de interacciones entre proteínas (edges). Todos los análisis de interactomas del presente trabajo están basados en el componente más grande de cada red. La significancia estadística del interactoma se obtuvo a través del valor de la  $p$  empírica, el cual se calculó mediante el conteo del número de veces en que el tamaño del componente más grande de las redes generadas de genes al azar sobrepasó al tamaño del componente más grande de la red de interés y el resultado se dividió entre la cantidad de interactomas al azar que fueron generados. El modelo “local null” evalúa la significancia estadística de la modularidad de la red. Para ello, la modularidad del interactoma analizado se compara con la modularidad de cada interactoma construido con listas de genes seleccionados al azar. Con base en ello, la significancia estadística de la modularidad es calculada automáticamente por Netbox a partir

del cual se obtiene un z-score [18]. Para determinar la conectividad global de los interactomas contruidos a partir de las proteínas de unión a los elementos esenciales, se comparó el tamaño del componente más grande (número de nodos + número de edges) de cada uno de los interactomas con respecto al componente más grande de los interactomas generados por listas de genes seleccionados al azar de los genes que componen el HIN. Para obtener el valor de la conectividad global generada al azar, se generaron 1000 listas al azar del número de proteínas de interés a comparar. El valor de la p empírica fue calculado para cada interactoma por elemento esencial, como se mencionó anteriormente y también fueron obtenidos el Z score de la modularidad de dichos interactomas. Cuanto más significativa es la modularidad, es más probable que la red represente algún proceso biológico[11][18][20][21]. Sin embargo, cuando no es significativo, es más probable que la interacción entre proteínas sea debida al azar y no pertenezca a un proceso celular real [11][18][20][21].

#### **11.4. Análisis funcional**

Mediante el mapeo a Gene Ontology (GO) y a KEGG a través de la herramienta de análisis bioinformático WebGestalt [50-a], se determinaron las funciones moleculares, procesos biológicos, y componentes celulares y vías del KEGG de cada una de las proteínas que componen el interactoma del Zn y con ello se obtuvo el posible rol biológico del interactoma y de cada módulo que lo compone. Finalmente, utilizando esta herramienta bioinformática, determinamos cuáles enfermedades se encontraron enriquecidas por las proteínas que componen al interactoma del Zn y a cada uno de sus módulos a través de su mapeo al término “Disease Association Analysis” y usando como referencia al genoma humano. El punto de corte establecido para obtener las funciones más significativas del interactoma y de cada módulo fue de  $p \leq 0.05$ . De las funciones moleculares, procesos biológicos y componentes celulares obtenidos por GO y que resultaron ser significativos, solo registramos en las tablas los términos más específicos. Para resumir y facilitar el entendimiento de la información obtenida por KEGG y las enfermedades asociadas, solo fueron seleccionados los primeros diez datos más significativos. Los resultados fueron presentados en inglés, tal como los obtuvimos de cada análisis, ya que existen términos que no tienen una traducción idónea al español.

#### **11.5. Análisis de estructura-función**

La predicción del efecto de las mutaciones de cambio de aminoácido (missense) en la estructura-función de las proteínas del interactoma se realizó mediante una integración y curación manual de los resultados de 6 herramientas bioinformáticas frecuentemente utilizadas para predecir las consecuencias

de las mutaciones en las proteínas, de las cuales 5 están basadas en análisis de secuencia: FIS, SIFT, SNAP, PolyPhen-2 y ConSurf, y una está basada en el análisis de la estructura: FoldX. Como resultado, se obtuvieron valores del grado de afectación de la proteína mutada de cada uno de los predictores utilizados. Los análisis basados en secuencia comparan la secuencia de la respectiva proteína mutada con un alineamiento de secuencias de proteínas ortólogas. El análisis basado en estructura se realizó a través de FoldX, el cual es un programa de diseño de proteínas y modelaje molecular, ampliamente utilizado para predecir el efecto de las mutaciones sobre la estabilidad de la proteína, el efecto sobre la interacción proteína-proteína y también proteína ligando [51] [52]. Las estructuras de las proteínas fueron obtenidas del Protein Data Bank (PDB). Si la proteína analizada no tenía estructura disponible, se construyó un modelo estructural y solo aquellas que tuvieron más del 50% de homología de secuencia con el molde utilizado para el modelaje fueron analizadas por FoldX. De esta manera, cuando la mutación cambia el  $\Delta\Delta G > 0.5$  kcal/mol se determinó que dicha mutación afecta la estructura de la proteína.

A continuación se describe brevemente la función de cada uno de los predictores usados. FIS predice el impacto funcional de las sustituciones de aminoácidos en proteínas evaluando la conservación evolutiva de los aminoácidos afectados en proteínas homólogas [53]. SIFT predice si la sustitución de un aminoácido afecta la función de la proteína, basándose en el grado de conservación de los residuos de aminoácidos mediante alineaciones de secuencias estrechamente relacionadas, recogidas a través de PSI-BLAST [54] [55]. SNAP utiliza algoritmos de redes neurales (modelos computacionales o matemáticos basados en enfoques de conexiones o redes de cálculos capaces de aprender en forma automática las propiedades deseadas a través de sistemas adaptativos que cambian su estructura según la información que fluye a través de la red y posteriormente generar resultados) para la predicción de los efectos funcionales de las mutaciones no sinónimas pero también su algoritmo incluye predicciones de estructura secundaria, y accesibilidad al solvente, [56]. PolyPhen-2 predice el posible impacto de una sustitución de aminoácido sobre la estructura y función de la proteína a través del mapeo de la mutación a la estructura tridimensional de la misma y combina una variedad de características tales como secuencia, propiedades evolutivas e información estructural [37]. ConSurf estima la conservación evolutiva de la posición de un aminoácido o ácido nucleico en moléculas de proteína, ADN o ARN basándose en las relaciones filogenéticas entre secuencias homólogas a través de un alineamiento [57]. Los criterios de cada predictor para considerar que la mutación afecta a la proteína se presentan en la tabla 1.

Predictor	Proteína no afectada	Proteína afectada
FIS	Neutral/low	Medium/high
SIFT	Tolerant	Damaging
SNAP	Neutral	Non-neutral
PolyPhen2	Neutral	Deleterious
FoldX	$\Delta\Delta G < 0.5$ kcal/mol	$\Delta\Delta G \geq 0.5$ kcal/mol
ConSurf	< 80% de conservación	$\geq 80\%$ de conservación

**Tabla 1. Criterios de cada programa para predecir la afectación de la proteína.** El  $\Delta\Delta G$ , determinado por FoldX, es la variación de energía que existe entre la proteína silvestre y la mutante. Cuando el cambio de energía entre la proteína mutada con respecto a la silvestre es mayor a 0.5 kcal/mol se predice que la estructura de la proteína está afectada ya que ha perdido su estabilidad estructural. ConSurf estima la conservación evolutiva del aminoácido que ha sido cambiado por otro, de esta forma, cuanto más conservado esté el aminoácido, es decir, cuanto más cerca esté del 100% de conservación, más probable es que su cambio por otro altere la función de la proteína mutada.

Para determinar en primera instancia si una proteína está o no afectada tomamos en cuenta las siguientes opciones:

1. Si la proteína mutada cuenta con el análisis de predicción por todos los programas, consideramos que su estructura-función está alterada cuando al menos 4 de 6 programas la predicen como afectada o cuando FoldX estima que se encuentra inestable aunque menos de 4 programas hagan la predicción de afectada.
2. Si la proteína mutada cuenta con el análisis de predicción de solo 5 programas (es decir cuando no tenga estructura para ser analizada por FoldX), consideramos que su estructura-función se encuentra alterada cuando al menos 3 de 5 programas predijeron que está afectada.
3. Las proteínas con mutaciones de tipo codón de paro (nonsense) que no contaron con estructura o modelo para la región de la mutación fueron directamente consideradas como afectadas.

En el genoma del Mieloma Múltiple ya se reportaron las mutaciones de tipo deleciones, duplicaciones en tándem, en el sitio de corte y empalme, inversiones, translocaciones e inserciones que afectan la estructura de diversos genes y por lo tanto afectan a las proteínas codificadas provocando la pérdida de su función [2]. Todas estas proteínas fueron entonces consideradas como afectadas en su estructura y/o función a lo largo de nuestro análisis.



### 11.6. Estimación del valor global de afectación para el análisis de estructura-función

Para poder integrar los resultados de las seis herramientas de predicción de manera cuantitativa, se analizaron diferentes opciones de fórmulas que permitieran combinar e integrar los puntajes obtenidos por cada predictor en un solo valor de predicción (ANEXO 1). De esta manera cuanto mayor sea el número de programas que predican que la proteína está afectada, más alto debería ser el valor global de afectación obtenido por dicha fórmula. La fórmula que mejor integró el resultado de todas las predicciones en un solo valor fue la siguiente:

$$\text{Valor global de afectación} = \text{Predicción (FIS)} + \text{predicción (SIFT)} + \text{predicción (PolyPhen2)} + \text{predicción (SNAP)} + \text{predicción (ConSurf)} + \text{predicción (FoldX)}$$

Pero para poder hacer comparables los valores de predicción de todos los programas tuvimos que normalizarlos de la siguiente manera y acorde a los datos de predicción obtenidos para nuestro set de proteínas mutadas. FIS manejó valores que van de -5.76 a 5.37. Cuanto más alto es este valor de predicción, más afectada se encuentra la proteína. Dichos valores fueron normalizados a una escala de 0 a 1. De esta manera, cuanto más cercano a 1 es el valor de predicción existe mayor probabilidad de que la proteína esté afectada. Por otro lado, los valores de predicción de SIFT van de 0 a 1. Cuando el valor de predicción es más cercano a 0, es más probable que la proteína se encuentre afectada. Para normalizarlo, dichos valores le fueron restados a 1, de manera tal que cuanto más afectada está la proteína, los valores tienden a acercarse a 1 y cuanto menos afectada esté la proteína, los valores tiendan a acercarse a 0, como los demás predictores. En el caso de PolyPhen2, los valores de predicción van de 0 a 1. Cuanto más afectada está la proteína, dichos valores tienden a acercarse a 1. Por lo tanto, no fue necesario normalizarlo. Por otro lado, debido a que SNAP reporta los resultados de una mutación como “neutral” o “non-neutral”, tomamos a “neutral” como cero y a “non-neutral” como 1. Por lo tanto, tampoco fue necesario normalizar los valores de este programa. ConSurf reporta el porcentaje de conservación para cada aminoácido por lo que dicho valor de conservación fue cambiado a valores decimales. Por ejemplo, si el residuo se encuentra conservado en un 60%, el valor se cambió a 0.6. De esta forma, la mutación en aminoácidos más conservados tendría valores cercanos a 1 y los menos conservados tendrían valores cercanos a 0. Los resultados de FoldX fueron normalizados de la siguiente manera: si el  $\Delta\Delta G$  es mayor a 0.5 kcal/mol, el valor normalizado correspondiente es 1. Cuando el  $\Delta\Delta G$  es menor a 0.5 kcal/mol, la estructura de la proteína es estable y el valor de normalización es de -1. Si la proteína no pudo ser analizada por FoldX, le fue colocado el valor cero para no involucrar a FoldX en el resultado final de la fórmula.

### **11.7. Determinación de genes candidato**

Debido a que los genes candidato suelen estar enriquecidos en mutaciones que afectan su función y que tienen la característica de estar altamente conectados en los interactomas [17][27][44], se seleccionaron los genes candidato acorde a sus propiedades determinadas por el análisis de estructura-función y el análisis de interactomas de la siguiente manera:

1. Genes que forman parte del interactoma y que están afectados por estructura-función en al menos dos pacientes (genes Zn mutados) o en al menos un paciente (gen linker).
2. Genes que son HUB en el interactoma (proteína que tiene muchas interacciones proteína-proteína). En este caso, consideramos a un gen como HUB cuando está mutado en por lo menos dos pacientes y tiene más de 5 interacciones [45], o está mutado en 1 paciente y tiene más de 12 interacciones [45]. En caso de un gen linker, fue considerado como candidato cuando tiene más de 5 interacciones y está mutado en al menos un paciente.
3. Genes que están afectados en su estructura-función y que además son HUB.

### **11.8. Análisis de los genes candidato clínicamente accionables**

Buscamos cuales proteínas definidas como candidato pueden ser blanco de diferentes drogas ya aprobadas por la FDA a través del “Drug Bank database” (<http://www.drugbank.ca/>). En esta página se buscó el nombre del gen bajo el término “targets” y se filtró la búsqueda de drogas por “drogas aprobadas”.

### **11.9. Planteamiento del método para definir la expresión absoluta de genes**

Desarrollamos un método de análisis de las fluorescencias de los microarreglos que permitiera establecer un punto de corte “threshold” para determinar la expresión absoluta de cada gen, mediante un análisis funcional de la expresión de los genes del cromosoma Y. Este método para determinar la expresión absoluta se basa en que los genes del cromosoma Y no deben estar expresados en muestras de mujeres, por lo tanto, la fluorescencia obtenida para las sondas de estos genes a partir de muestras de ARN procedentes de mujeres podría considerarse como ruido de fondo propio de la tecnología del microarreglo [58]. De esta manera, el valor de fluorescencia que presenten los genes del cromosoma Y en muestras de mujeres, lo consideraremos como ruido de fluorescencia y en base en él estableceremos nuestro punto de corte de expresión, el cual nos servirá para identificar que aquellos genes con valores de fluorescencia mayores a éste punto de corte pueden considerarse como expresados y los que están por debajo pueden considerarse como no expresados. Los valores de fluorescencia para esos genes

fueron analizados inicialmente en un set de muestras de mucosa oral de 5 hombres y 5 mujeres (GSE17913) y posteriormente en un set de 204 hombres y 210 mujeres con Mieloma Múltiple (GSE2658) [59]. Los análisis en estos grupos de pacientes se realizaron con el objetivo de identificar qué genes del cromosoma Y tienen una expresión diferencial dependiente del sexo (dado lo explicado anteriormente) y a partir de ellos poder establecer nuestro punto de corte de expresión absoluta. Una vez seleccionados los genes con diferencias de expresión dependientes del sexo en los sets de pacientes mencionados, calculamos el punto de corte de expresión para los genes del set de individuos con Mieloma Múltiple del MMRC correspondiente a los genomas secuenciados que analizamos (GSE26760).

#### **11.10. Verificación del método de expresión absoluta**

Para validar nuestro punto de corte de expresión analizamos la fluorescencia de un pequeño set de genes constitutivos o “Housekeeping” en relación al punto de corte establecido en los individuos con Mieloma Múltiple (GSE2658) [59]. Dado que estos genes codifican para proteínas que son esenciales para el funcionamiento general de todas las células, deben expresarse sin importar el género, es decir, deben de estar expresados tanto en hombres como en mujeres. Con este análisis pudimos definir la expresión de los genes constitutivos a través del punto de corte de expresión y verificamos si éste es capaz de demostrar que este set de genes se encuentra expresado en todos los individuos independientemente del sexo. Analizamos también, en base al punto de corte de expresión, la expresión absoluta del gen SRY en la misma población con Mieloma Múltiple (GSE2658). Éste es un gen que a pesar de pertenecer al cromosoma Y y ser determinante del sexo, no debe estar expresado ni en hombres ni en mujeres en etapa adulta ya que solo se activa durante el desarrollo embrionario [60]. Finalmente, analizamos las fluorescencias reportadas de un estudio (GSE22974) en donde se hibridó ARN de levadura en 5 microarreglos de humano (Affymetrix U133plus2). Este análisis nos permitió observar el ruido de fluorescencia de las sondas del microarreglo dado que el microarreglo está diseñado para hibridar ARN de genes de humano y no de levadura. Además, comprobamos si los genes que fueron seleccionados para establecer el punto de corte de expresión absoluta son capaces de determinar la no expresión del ARN de levadura en microarreglos de humano, dado que los genes de levadura no deberían hibridar con sondas de genes de humanos.

### **11.11. Aplicación del método de expresión absoluta**

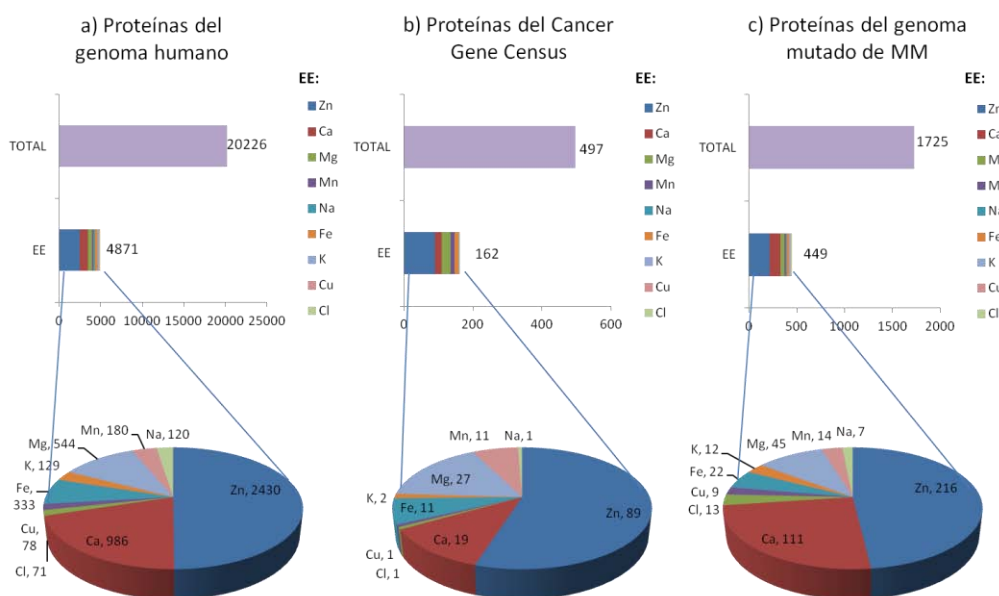
Una vez ya validado el algoritmo de expresión absoluta, éste fue aplicado a los microarreglos de expresión correspondientes a los pacientes secuenciados con Mieloma Múltiple del MMRC (GSE26760). De esta forma, se analizó la posibilidad de que cada uno de los módulos que conforman el interactoma se esté expresando y de que las proteínas afectadas también lo estén.

## 12. RESULTADOS

### 12.1. CAPÍTULO 1. Determinación de genes candidato en Mieloma Múltiple

#### 12.1.1. Proteínas que unen elementos químicos esenciales (EE)

Por la secuenciación del genoma del Mieloma Múltiple se reportaron en total 1725 mutaciones en 1429 genes. Para poder evaluar si el genoma de MM posee una sobre-representación de proteínas que unen EE necesitamos primero analizar la distribución de las proteínas del genoma humano que unen EE. La distribución encontrada se muestra en la figura 2a, donde encontramos que el 24% (4871 de 20226) de las proteínas totales en el genoma humano unen EE. También, evaluamos cómo se encuentran distribuidas las proteínas que unen EE dentro del grupo de proteínas previamente asociadas con cáncer (“Cancer Gene Census” (CGC por sus siglas en inglés)). Como se muestra en la figura 2b, encontramos que el 33% (162 de 497) de las proteínas del CGC unen EE. Además, observamos que el 26% (449 de 1725) de las proteínas mutadas en el genoma de MM unen EE (figura 2c).



**Figura 2. Distribución de las proteínas que unen elementos esenciales (EE) en cada set de datos genómicos. a)** Distribución de EE dentro de las 20226 proteínas del genoma humano. **b)** Distribución de EE en los 497 genes codificadores de proteínas del Cancer Gene Census (CGC). **c)** Distribución de EE dentro de los 1725 genes mutados codificadores de proteínas en Mieloma Múltiple (MM).

Para analizar si existe un enriquecimiento de las proteínas que unen EE en los Cancer Gene Census y en el genoma de Mieloma Múltiple con respecto al genoma humano se analizó su frecuencia esperada y observada con base a las distribuciones previamente descritas (tablas 2 y 3).

EE	$f_{obs}$	$f_{esp}$	Razón de cambio	Nº de proteínas enriquecidas
Zn	89	60	1.5	29
Ca	19	24	0.8	-5
Cl	1	2	0.6	-1
Cu	1	2	0.5	-1
Fe	11	8	1.3	3
K	2	3	0.6	-1
Mg	27	13	2.0	14
Mn	11	4	2.5	7
Na	1	3	0.3	-2

**Tabla 2. Frecuencia de proteínas que unen EE en Cancer Gene Census.** El número de proteínas que unen EE en CGC fue comparado con respecto a la distribución de las proteínas que unen EE en el genoma humano. Cancer Gene Census: 497 genes codificadores de proteínas. La fracción observada ( $f_{Obs}$ ) fue calculada acorde al número de proteínas que unen EE en Cancer Gene Census. La fracción esperada ( $f_{Esp}$ ) fue calculada acorde a la distribución de las proteínas que unen EE para el genoma humano. Razón de cambio= ( $f_{Obs}/f_{Exp}$ ). Nº de proteínas enriquecidas= ( $f_{Obs} - f_{Esp}$ ).

En la tabla 2 se puede observar que algunas proteínas del CGC que unen EE están sub o sobre-enriquecidas, sugiriendo que este tipo de proteínas tienen un rol importante en el cáncer. También, analizamos si existe un sobre-enriquecimiento de las proteínas que unen EE dentro del genoma de Mieloma Múltiple (tabla 3).

EE	<sup>a</sup> $f_{obs}$	<sup>b</sup> $f_{Esp}$	<sup>c</sup> Razón de cambio	<sup>d</sup> Nº de proteínas enriquecidas
Ca	111	84.52	1.31	26.48
Cl	13	6.9	1.88	6.1
Cu	9	6.9	1.30	2.1
Fe	22	27.6	0.80	-5.6
K	12	10.35	1.16	1.65
Mg	45	46.57	0.97	-1.57
Mn	14	15.52	0.90	-1.52
Na	7	10.35	0.68	-3.35
Zn	216	207	1.04	9

**Tabla 3. Frecuencia de proteínas que unen EE en el genoma mutado de Mieloma Múltiple.** Genoma mutado de MM: 1725 genes codificadores de proteínas. <sup>a</sup>La fracción observada ( $f_{Obs}$ ) fue calculada acorde al número de proteínas que unen EE en el genoma de MM. <sup>b</sup>La fracción esperada ( $f_{Esp}$ ) fue calculada acorde a la distribución de las proteínas que unen EE en el genoma humano. <sup>c</sup>Razón de cambio= ( $f_{Obs}/f_{Esp}$ ). <sup>d</sup>Nº de proteínas enriquecidas= ( $f_{Obs} - f_{Esp}$ ).

Es interesante notar que las proteínas sobre-representadas en CGC unen diferentes EE en comparación con las proteínas que están sobre-representadas en el genoma de Mieloma Múltiple (tabla 3). Probablemente esto se deba a que el set de Cancer Gene Census se encuentra conformado por genes ya reportados como candidato en cáncer, en cambio los genes mutados en Mieloma Múltiple contienen tanto mutaciones candidato como mutaciones pasajeras. Debido a estos resultados decidimos analizar si los 27 genes candidato obtenidos por el MMRC (tabla 4) a través de la secuenciación de los 38 genomas con Mieloma Múltiple se encuentran enriquecidos en elementos esenciales (tabla 5) y encontramos que las proteínas que unen Cu, Fe y Zn se encuentran enriquecidas en este set (7.14 veces el Cu, 2.94 veces el Fe y 2.36 veces el Zinc). Esto sugiere al igual que en las proteínas del CGC que podría existir un proceso de selección hacia un enriquecimiento de mutaciones para este tipo de genes.

ALOX12B	IKBKB	RIPK4
BTRC	LRRK2	TLR4
CARD11	MAGED1	TNFRSF1A
CCND1	MAP3K1	TP53
CYLD	MAP3K14	TRAF3
DIS3	MLL	UTX
HLA-A	MLL2	WHSC1
HOXA9	MLL3	WHSC1L1
IKBIP	PNRC1	XBPI

**Tabla 4. Lista de genes candidato obtenidos por la secuenciación de los 38 genomas de Mieloma Múltiple**

Motivo de unión a elemento esencial	<sup>a</sup> fObs	<sup>b</sup> fExp	<sup>c</sup> Razón de cambio
<b>Ca</b>	0	--	--
<b>Cl</b>	0	--	--
<b>Cu</b>	1	0.14	7.14
<b>Fe</b>	1	0.34	2.94
<b>K</b>	0	--	--
<b>Mg</b>	0	--	--
<b>Mn</b>	0	--	--
<b>Na</b>	0	--	--
<b>Zn</b>	8	3.38	2.36

**Tabla 5. Distribución de los elementos esenciales dentro del set de genes candidato en Mieloma Múltiple.**

<sup>a</sup>La fracción observada ( $f_{Obs}$ ) fue calculada acorde al número de genes candidato que unen elementos esenciales.

<sup>b</sup>La fracción esperada ( $f_{Exp}$ ) fue calculada acorde la distribución de las proteínas que unen elementos esenciales en el genoma de Mieloma Múltiple. <sup>c</sup>Razón de cambio = ( $f_{Obs}/ f_{Exp}$ ).

Es interesante señalar que las proteínas con dominios de unión a Zinc y hierro se encuentran sobre-representadas tanto en el set de los genes candidato del Cancer Gene Census como en el de los genes candidato ya reportados en Mieloma Múltiple. Probablemente, este tipo de proteínas puedan estar involucradas de manera importante en el proceso del desarrollo del cáncer.

### 12.1.2. Interactomas en Mieloma Múltiple y análisis estadístico

A partir de las proteínas mutadas en Mieloma Múltiple construimos la red de interacción proteína-proteína para cada elemento esencial. Para cada interactoma fue calculada su significancia estadística con respecto a sus modularidades locales (“local null model”) y conectividades globales (“global null model”) generadas por listas del mismo número de genes pero seleccionados al azar del Human Interaction Network (HIN) (ver materiales y métodos). En la tabla 6 se presentan los resultados.

<i>Modelo “local null”</i>		<i>Modelo “global null”</i>			
<sup>a</sup> EE	<sup>b</sup> Modularidad (Z score)	<i>Nodos</i>	<sup>c</sup> p-value	<i>Edges</i>	<sup>c</sup> p-value
<b>Zn</b>	<b>11.64</b>	180	<b>0.00</b>	585	<b>0.04</b>
<b>Ca</b>	6.04	75	0.05	151	0.09
<b>Cl</b>	-0.30	7	0.08	8	0.09
<b>Cu</b>	0	0	1	0	1
<b>Fe</b>	0	0	1	0	1
<b>K</b>	2.21	8	0.10	19	0.07
<b>Mg</b>	1.53	13	0.35	18	0.36
<b>Mn</b>	0	0	1	0	1
<b>Na</b>	0	0	1	0	1

**Tabla 6. Propiedades de los interactomas de los diferentes elementos esenciales.**

<sup>a</sup>Interactoma de Zn: 73 proteínas mutadas y 110 linkers, interactoma Ca: 33 proteínas mutadas y 42 linkers, interactoma Cl: 4 proteínas mutadas y 3 linkers, interactoma Cu: 0 proteínas mutadas y 0 linkers, interactoma Fe: 0 proteínas mutadas y 0 linkers, interactoma K: 6 proteínas mutadas y 6 linkers, interactoma Mg: 8 proteínas mutadas y 8 linkers. Interactoma Mn: 2 proteínas mutadas y 1 linker, interactoma Na: 2 proteínas mutadas y 0 linkers. Las proteínas linker son proteínas que interactúan con las proteínas mutadas en MM en forma significativa.

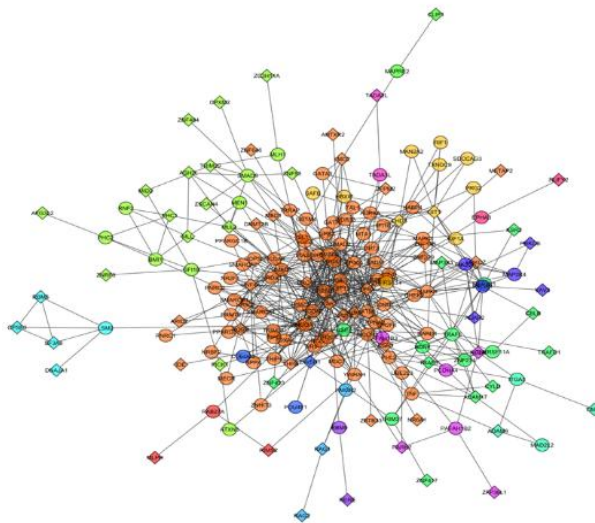
<sup>b</sup>Para cada interactoma fue calculada la modularidad. Dicha modularidad fue comparada con la modularidad obtenida por la interacción de genes del HIN seleccionados al azar. Con base en ello, el valor de la modularidad fue obtenida como Z-score para cada interactoma.

<sup>c</sup>El valor de la p empírica (p-value) fue determinado tomando en cuenta el número de veces que el componente más grande de los 1000 interactomas al azar igualó o excedió el componente más grande de los interactomas formados por proteínas que unen EE.



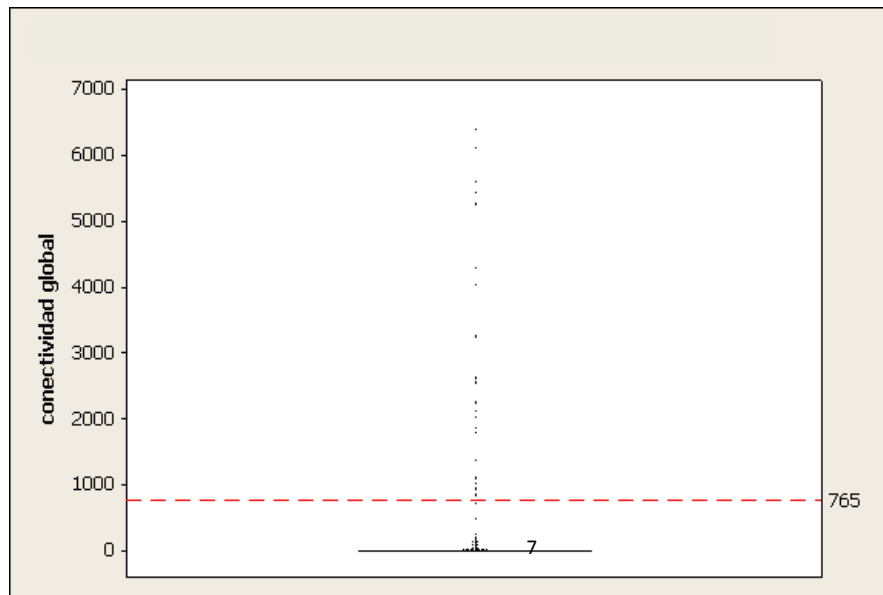
El único interactoma que resultó ser significativo con respecto al azar fue el construido a partir de las proteínas de unión a Zinc con un Z score de modularidad de 11.64 y una  $p < 0.05$  para la conectividad global. Es decir, las proteínas mutadas en Mieloma Múltiple que unen Zinc son las únicas capaces de formar redes de interacción proteína-proteína que tienen una probabilidad menor del 5% de ser formadas por el azar, a diferencia de las proteínas que unen otros elementos esenciales, cuyas redes no fueron significativas con respecto al azar. Los análisis de los interactomas de EE y de sobre-representación sugieren la importancia del estudio del set de proteínas con dominios de unión a Zinc en Mieloma Múltiple. Por lo tanto, este set de proteínas se seleccionó para el desarrollo del resto de las estrategias presentadas a continuación en la búsqueda de nuevos genes candidato.

El interactoma del Zinc, descrito en la tabla 6, fue obtenido mediante el mapeo de las 182 proteínas mutadas que unen Zinc en el MM al Human Interaction Network (HIN). De estas 182 proteínas el HIN solo reconoció en su base de datos a 103, por lo tanto, el interactoma se formó por estas 103 proteínas iniciales. Obtuvimos un interactoma cuyo componente más grande está conformado por 180 nodos (número de proteínas totales) de los cuales 72 pertenecen a las proteínas mutadas de unión a Zn y 108 a proteínas linkers (proteínas que interactúan con las del Zinc pero que no necesariamente se encuentran mutadas). Los nodos se encuentran conectados a través de 588 edges (número de interacciones). En total el interactoma posee 16 módulos (figura 3).



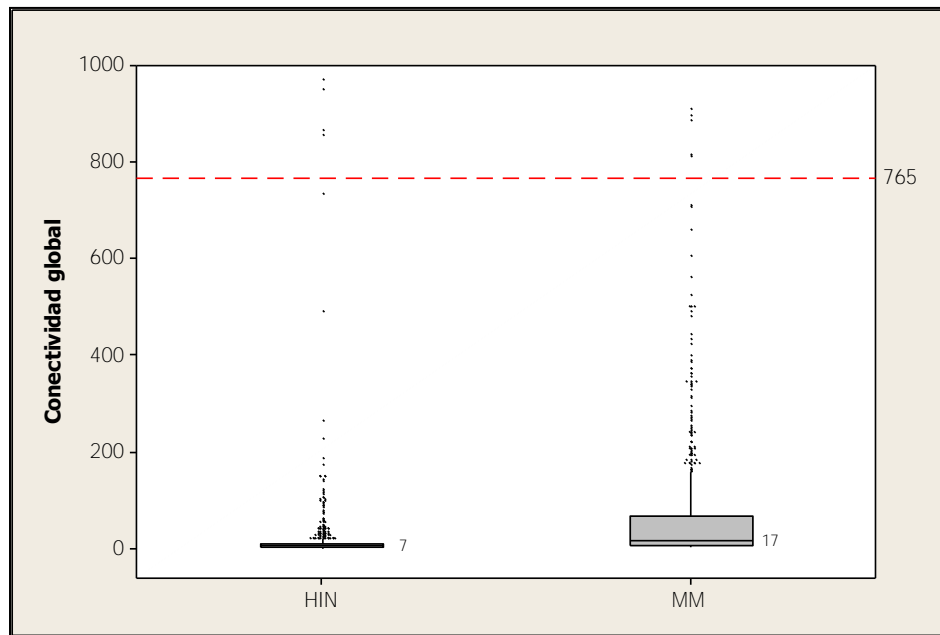
**Figura 3. Interactoma de las proteínas mutadas con dominios de unión de Zinc.** El interactoma del Zinc está formado por 180 nodos de los cuales 72 corresponden a 72 proteínas mutadas que unen Zinc y 108 corresponden a proteínas linker (proteínas que interactúan con las proteínas mutadas del Zinc). Los nodos están conectados a través de 588 interacciones. Un total de 16 módulos fueron identificados como significativos con un z-score de 11.64. Los diferentes colores representan cada módulo. Los módulos son grupos de genes densamente conectados que podrían representar diferentes funciones celulares. Cada color corresponde a un módulo, los rombos corresponden a las proteínas mutadas que unen Zn y los círculos a las proteínas linker (las cuales no necesariamente unen Zn).

La conectividad global del componente más grande del interactoma del Zinc fue comparada con respecto al azar. Para ello se hicieron 1000 listas de 103 proteínas, cada una de ellas obtenidas al azar del genoma humano y que estén presentes en la base de datos del HIN. Una vez obtenidas las conectividades globales de los componentes más grandes de los interactomas al azar, se calculó la p empírica (ver métodos) y se obtuvieron los siguientes resultados (gráfico 1).



**Gráfico 1. Conectividad global del componente más grande del interactoma del Zinc.** Gráfico de cajas donde se presenta la conectividad global del interactoma del Zinc comparada con 1000 listas al azar del genoma humano. La línea roja punteada representa la conectividad global del interactoma del Zinc.

La conectividad global del componente más grande del interactoma del Zinc es de 765 y resultó ser estadísticamente significativa comparada con la obtenida por el azar con un valor de  $p= 0.03$  (mediana de la conectividad global de las listas al azar = 7). La modularidad también fue significativa con respecto al azar con un valor de  $z= 11.64$ . Dada la alta conectividad global observada por el interactoma formado a partir de las proteínas mutadas que unen Zn en MM, decidimos comparar la conectividad global del interactoma del Zinc con respecto a las conectividades globales de 1000 interactomas formados por todos los genes mutados en Mieloma Múltiple (gráfico 2). Para ello, generamos 1000 listas al azar de 103 genes mutados en Mieloma Múltiple.

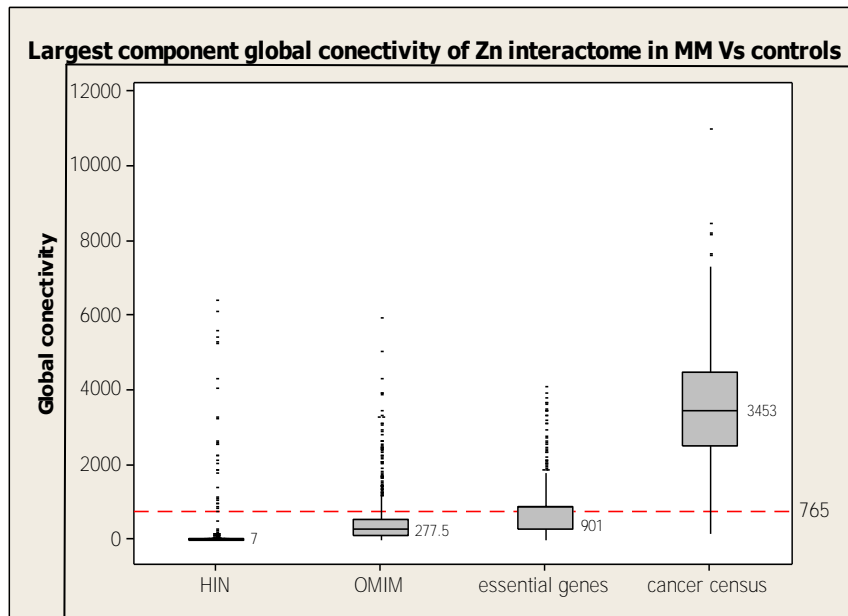


**Gráfico 2. Conectividad global del interactoma del Zn Vs interactomas de 103 proteínas generados de la lista de proteínas mutadas de Mieloma Múltiple (1429 proteínas).** La conectividad global del Zinc (CG=765) resultó ser mayor a las conectividades obtenidas por todas las proteínas mutadas de Mieloma Múltiple en forma significativa con una  $p= 0.036$ . Los grupos de 103 proteínas mutadas en mieloma tienden a formar interactomas con conectividades (mediana CG= 17) similares al azar (mediana CG= 7). La línea punteada roja corresponde a la conectividad global del interactoma del Zinc.

Las conectividades globales de los interactomas formados a partir de listas de 103 proteínas mutadas en Mieloma Múltiple no son significativas con respecto al azar. Probablemente esto se deba a que dentro de este grupo de proteínas mutadas se encuentran muchas con mutaciones pasajeras. Contrario a ello, la conectividad global del interactoma del Zinc es significativa con respecto al azar y su valor de conectividad es alta. Debido a que se ha observado que las proteínas candidato tienden a formar interactomas con altas conectividades globales en diferentes tipos de cánceres [44], probablemente, las proteínas que unen Zinc en Mieloma Múltiple se estén enriqueciendo de mutaciones candidato.

Dado que encontramos que las proteínas mutadas que unen Zinc forman un interactoma de conectividad global alta y significativo con respecto al azar, decidimos analizar si a partir de genes que son miembros del HIN, OMIM (Online Mendelian Inheritance in Man), genes esenciales y Cancer Gene Census (CGC) se construyen interactomas con las mismas características que las observadas en el interactoma del Zn. Esto debido a que dichos grupos de genes juegan papeles importantes dentro de la biología celular y muchos de ellos están relacionados con enfermedades monogénicas y el cáncer, entre otras. Por lo tanto, probablemente estos grupos de genes también formen interactomas significativos con respecto al azar y presenten altas conectividades globales. Por ejemplo, OMIM involucra a un grupo de genes cuya afectación está relacionada con el desarrollo de alguna enfermedad mendeliana, los genes esenciales son genes importantes para la funcionalidad de las células en humanos, y Cancer

Gene Census son un grupo de genes que ha sido reportado como candidato en diferentes tipos de cánceres. Por lo tanto, comparamos la conectividad global del interactoma del Zinc con respecto a listas de 103 genes seleccionados al azar provenientes del HIN, OMIM (Online Mendelian Inheritance in Man), genes esenciales y Cancer Gene Census (gráfico 3). Para realizar el análisis estadístico, 1000 listas de los 103 genes seleccionados al azar fueron analizadas por su capacidad de formar interactomas.

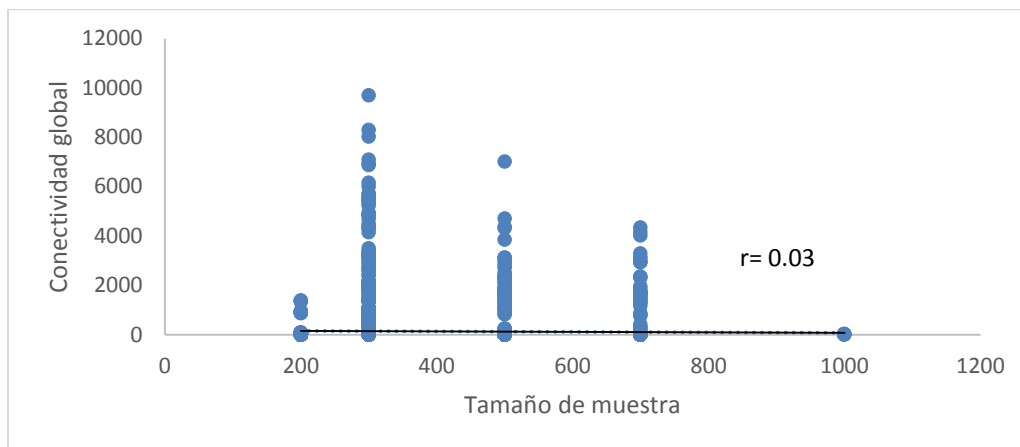


**Gráfico 3. Conectividad global del interactoma del Zinc comparado con otras listas de genes.** La línea punteada es la conectividad global del interactoma del Zinc. Las medianas de las conectividades globales de cada lista de genes se encuentran al lado de cada caja. Para cada grupo de genes, 1000 listas de los 103 genes seleccionados al azar fueron analizadas por su capacidad de formar interactomas.

La conectividad global del interactoma del Zinc resultó ser significativa con respecto a los genes del HIN, como fue descrito anteriormente, sin embargo, no fue significativa con respecto al OMIM, a los genes esenciales, y a los Cancer Gene Census, con los cuales se obtuvieron los valores de p iguales a 0.15, 0.55 y 0.99, respectivamente. Con estos resultados puede observarse que las proteínas mutadas del Zinc forman interactomas con conectividades diferentes al azar pero similares a la de genes que participan dentro de funciones celulares importantes y en procesos relacionados con el desarrollo de enfermedades monogénicas y cáncer. De esta manera, es posible validar estadísticamente la relevancia biológica de este interactoma.

Debido a que los interactomas obtenidos del HIN, OMIM, genes esenciales, Cáncer Gene Census y Mieloma Múltiple provienen de diferentes tamaños de muestra (HIN= 9254 genes, OMIM= 1442, genes esenciales= 1952 genes, Cancer Gene Census= 497 genes, Mieloma Múltiple= 1725 genes), decidimos evaluar si el número de muestra de donde se obtienen las listas al azar de genes influye en la

conectividad global. Para ello, decidimos tomar del HIN subgrupos de 200, 300, 500, 700 y 1000 genes, y a partir de cada uno de ellos generamos 1000 listas de 103 genes al azar (gráfico 4).

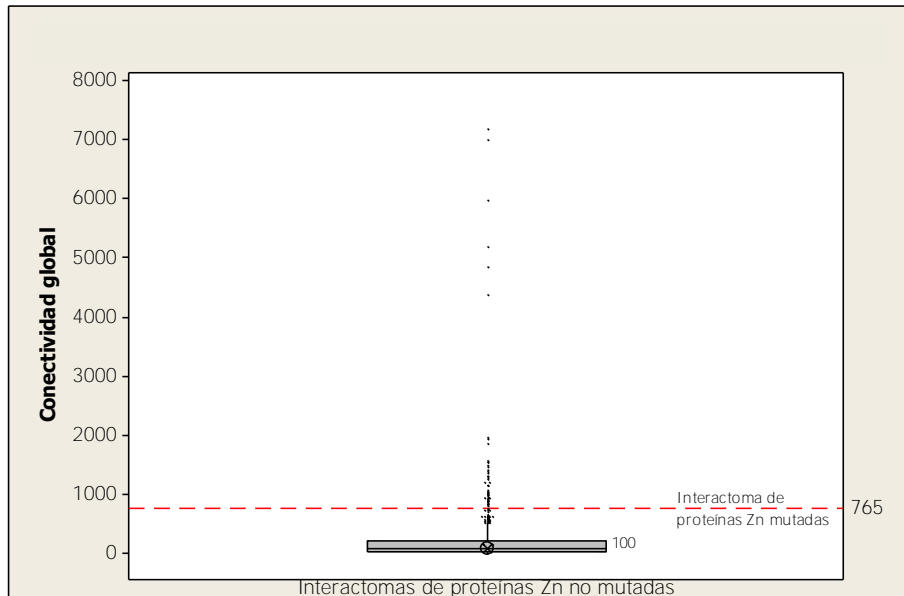


**Gráfico 4. Análisis de correlación entre el tamaño de muestra y la conectividad global.** Se presentan las conectividades globales obtenidas de 1000 listas al azar de 103 genes provenientes de muestras de 200, 300, 500, 700 y 1000 genes del HIN.

A través de este análisis pudimos observar que el tamaño de la muestra de donde se generan las listas de genes al azar no influye en las conectividades globales dado que no hubo correlación (gráfico 4). Esto sugiere que las conectividades globales obtenidas para cada set de genes analizado (HIN, OMIM, genes esenciales, Cancer Gene Census y Mieloma Múltiple) no están influenciadas por la cantidad de donde salen los subgrupos de genes que los conforman sino por la importancia biológica que presentan.

Finalmente, evaluamos qué pasa si generamos interactomas a partir de proteínas que unen Zinc pero que no están mutadas en Mieloma. Esto con la finalidad de ver si cualquier proteína que unen Zinc tiene la característica de formar interactomas más significativos que el azar o si solo la selección de ciertas proteínas Zinc tiende a construir interactomas más significativos. Para ello, obtuvimos una lista con todas las proteínas del genoma humano que tienen dominios de unión a Zinc a través de UniProtKB. A la lista obtenida le eliminamos las proteínas que están mutadas en mieloma quedándonos con una lista de 2233 proteínas no mutadas que unen Zinc. Con ellas, realizamos 1000 listas al azar de 103 proteínas y obtuvimos sus conectividades globales para cada interactoma. Como resultado no hubo diferencia significativa entre la conectividad de las proteínas de Zn mutadas y no mutadas con una  $p = 0.054$ . Este resultado sugiere que las proteínas que unen Zinc, aunque no estén necesariamente mutadas, tienden a formar interactomas con conectividades globales altas. Sin embargo, al analizar los datos a través del gráfico de cajas, es posible observar que la mediana de las conectividades globales (100) de los

interactomas formados por las proteínas de unión a Zn que no están mutadas en MM es mucho menor que el valor de conectividad global (765) del interactoma del Zn (proteínas Zn mutadas en MM) y que, además, solo valores atípicos caen a la altura de la conectividad global del interactoma del Zn, lo que sugiere que las proteínas de unión a Zn mutadas en MM tienden a tener conectividades globales más altas que las proteínas Zn no mutadas aunque no es estadísticamente significativa (Gráfico 5).



**Gráfico 5. Conectividad global de los interactomas de proteínas Zn no mutadas vs el interactoma de las proteínas Zn mutadas en MM.** La línea punteada es la conectividad global del interactoma del Zinc. La mediana de las conectividades globales de los interactomas conformados a partir de proteínas de unión a Zn no mutadas en MM se encuentra al lado de la caja.

### 12.1.3. Análisis funcional del interactoma

A través de la herramienta de análisis bioinformático WebGestalt (<http://bioinfo.vanderbilt.edu/webgestalt/>) mapeamos al Gene Ontology y al KEGG, por un lado, a los genes que componen a todo el interactoma del Zn y, por el otro, a los genes que componen a cada uno de los 16 módulos del interactoma, a partir de los cuales obtuvimos las funciones biológicas, procesos biológicos y las vías que se encuentran enriquecidas para cada uno de ellos. Finalmente, determinamos cuáles son las posibles enfermedades asociadas al interactoma y a cada módulo a través del mapeo de las proteínas que los componen al WebGestalt bajo el término “Diseases Association Analysis” donde se obtuvieron a las enfermedades que se encontraron enriquecidas con respecto a los genes que están involucrados en ellas y usando como referencia el genoma humano. En la tabla 7, se muestran los resultados obtenidos de estos análisis para el interactoma del Zinc. Las funciones moleculares que se encontraron enriquecidas fueron la transcripción, la unión al ion Zinc, a la ubiquitina ligasa y a los

receptores de andrógenos. Los componentes celulares involucrados estuvieron relacionados al núcleo. Además, dicho interactoma resultó estar enriquecido en vías relacionadas al cáncer y al ciclo celular, cuya afectación suele conducir a la supervivencia de las células tumorales. También se vio enriquecida la vía de señalización TGF-beta, la cual juega un papel importante en el control del sistema inmune, inhibe la proliferación y la apoptosis de linfocitos B, y juega un papel crucial en el control de la expresión de anticuerpos. Por lo tanto, su alteración posiblemente contribuya de manera importante al desarrollo del Mieloma Múltiple. La diferenciación osteoclástica, otra vía enriquecida por el interactoma del Zn, hace referencia a la maduración y actividad de los osteoclastos, cuya prolongada vida útil permite la expansión de la masa osteoclástica que, en su forma activa, es capaz de formar sitios de resorción ósea, la cual se ha visto aumentada en el MM y ser la causante de las lesiones líticas que son características de esta enfermedad. Dentro de las enfermedades que se encontraron asociadas al interactoma del Zinc de manera más significativa son el cáncer, las infecciones virales y algunas neoplasias. La asociación con las infecciones virales probablemente se relacione con la activación de la vía NF-kappa B (factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas), el cual es un complejo proteico que controla la transcripción del ADN y está implicado en la respuesta celular frente a estímulos como los antígenos bacterianos o virales. El NF-κB juega un papel clave en la regulación de la respuesta inmune debida a la infección ya que las cadenas ligeras kappa son componentes cruciales de las inmunoglobulinas. La regulación defectuosa del NF-kB está relacionada con el cáncer, enfermedades inflamatorias y autoinmunes, shock séptico, infecciones virales o un desarrollo inmune inadecuado. Aunque la vía de NF-kB solo se encuentra enriquecida en el módulo 5 (ver tabla 12), las proteínas que la conforman también son parte del interactoma del Zn. Interesantemente, también se ha asociado al interactoma del Zn, el Síndrome de Li-Fraumeni, que a pesar de ser una enfermedad hereditaria que se transmite de forma autosómica dominante, su principal característica es que los pacientes afectados presentan una mutación que afectan al gen supresor tumoral TP53 situado en el cromosoma 17p13.1, lo que les hace muy propensos a desarrollar diferentes tipos de cáncer a edades tempranas, entre ellos osteosarcomas y otros sarcomas, cáncer de mama, leucemias, linfomas y tumores cerebrales, entre otros. Por lo tanto, esto puede sugerir que dicha enfermedad podría estar relacionada con el desarrollo de MM en pacientes jóvenes.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		ubiquitin protein ligase binding	1.32E-10
transcription from RNA polymerase II promoter	5.17E-33	RNA polymerase II transcription factor binding	2.5E-10
regulation of transcription from RNA polymerase II promoter	5.83E-29	steroid hormone receptor activity	2.46E-09
positive regulation of transcription, DNA-dependent	1.07E-28	transcription corepressor activity	3.14E-09
negative regulation of transcription, DNA-dependent	2.36E-25		
		<b>Componente celular</b>	
<b>Función molecular</b>		nuclear chromatin	2.38E-10
Zinc ion binding	8.18E-28	nucleolus	1.05E-08
transcription coactivator activity	6.76E-17	histone deacetylase complex	1.79E-07
sequence-specific DNA binding RNA polymerase II transcription factor activity	5.44E-14	methyltransferase complex	1.88E-05
double-stranded DNA binding	1.18E-13	PRC1 complex	2.67E-05
ligand-dependent nuclear receptor binding	2.1E-12	PML body	6.42E-05
androgen receptor binding	3.63E-12	NuRD complex	6.83E-05
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
	<b>P ajustada</b>		<b>P ajustada</b>
Pathways in cancer	1.26E-19	cancer or viral infections	1.37E-24
Cell cycle	2.31E-13	Drug interaction with drug	1.97E-22
Pancreatic cancer	2.31E-13	Li-Fraumeni syndrome	1.72E-19
Chronic myeloid leukemia	2.82E-13	Cell Transformation, Neoplastic	4.91E-17
Colorectal cancer	1.81E-12	Breast Diseases	2.77E-16
Neurotrophin signaling pathway	3.67E-12	Breast Neoplasms	1.04E-15
Osteoclast differentiation	9.07E-11	Neoplasms	2.60E-15
Fc epsilon RI signaling pathway	5.15E-10	Skin Diseases, Genetic	8.92E-14
Non-small cell lung cancer	6.35E-10	Lymphoma	4.34E-13
TGF-beta signaling pathway	7.23E-10	Osteosarcoma	5.35E-13

**Tabla 7. Funciones biológicas del interactoma del Zn**



#### 12.1.4. Análisis funcional de cada módulo del interactoma

El módulo 0 (tabla 8) involucra procesos biológicos basados en la metilación de histonas, en el desarrollo del sistema inmune, y en la regulación de la transcripción. Además, comprende funciones moleculares con actividad de unión al ADN, a la cromatina y con actividad de ubiquitina-ligasa, y los componentes celulares enriquecidos están relacionados con el núcleo celular. Este módulo no presentó enriquecimiento en ninguna vía metabólica o de señalización, pero sí se asoció de manera significativa con leucemias de diferentes tipos y otras neoplasias.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		transcription regulatory region DNA binding	0.0037
histone H3-K4 methylation	2.99E-06	ubiquitin-protein ligase activity	0.0153
positive regulation of histone methylation	0.0013	exopeptidase activity	0.0218
immune system development	0.0013	sequence-specific DNA binding transcription factor activity	0.0272
regulation of transcription from RNA polymerase II promoter	0.002	double-stranded DNA binding	0.0408
<b>Función molecular</b>		<b>Componente celular</b>	
Zinc ion binding	1.62E-08	sex chromatin	0.0001
RING-like Zinc finger domain binding	3.36E-05	MLL1 complex	0.0013
histone methyltransferase activity (H3-K4 specific)	0.0004	nuclear body	0.0088
chromatin binding	0.003		
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
			<b>P ajustada</b>
Ninguna vía enriquecida		Leukemia	3.42E-06
		Lymphoid leukemia NOS	3.42E-06
		Leukemia, Lymphoid	3.42E-06
		Cell Transformation, Neoplastic	3.42E-06
		Leukemia, Myeloid	3.42E-06
		Multiple Endocrine Neoplasia	7.39E-06
		Neoplasms, Second Primary	4.16E-05
		Spinocerebellar Ataxias	5.01E-05
		Leukemia, Biphenotypic, Acute	2.00E-04
		Cerebellar Ataxia	2.00E-04

**Tabla 8. Funciones biológicas del módulo 0**

El módulo 1 (tabla 9) también está implicado en la regulación de la transcripción y sus funciones moleculares estuvieron relacionadas con la unión a la cromatina, al ion Zinc, a la actividad de receptor de hormonas esteroideas y al mecanismo de transcripción. Es sabido que la desregulación de estas

funciones están relacionadas con el cáncer, sin embargo, es importante destacar la actividad del receptor de hormonas esteroideas dado que los receptores de estrógenos, específicamente, suelen estar sobre-expresados en diferentes tipos de cáncer [68]. Interesantemente, las células de mieloma sobreexpresan receptores de estrógenos lo cual se ha relacionado con el desarrollo de MM [66]. Uno de los componentes celulares enriquecidos son los cuerpos nucleares PML que tienen la función de inhibir el crecimiento celular, está relacionado con la regulación de genes y está implicado en las vías de supresión tumoral. Otro de los componentes celulares relacionados es el complejo enzimático Sin3 que regula la transcripción por la desacetilación de los residuos de histona. El complejo NuRD, también enriquecido en este módulo, está relacionado con la represión de la transcripción y con el remodelamiento de la cromatina y el complejo Rb-E2F está involucrado en la regulación del ciclo celular. Las vías metabólicas y de señalización más significativas resultaron ser aquellas que involucran el ciclo celular, vías en cáncer, la vía de señalización TGF-beta anteriormente descrita, la vía de señalización Wnt que está implicada en la regulación de la transcripción génica en donde se ha visto que ciertas mutaciones en ella conducen al desarrollo del cáncer, y la vía de señalización del p53 implicada en la apoptosis y control del ciclo celular. Dentro de las enfermedades asociadas se obtuvo al Síndrome de Li-Fraumeni, leucemias y otras neoplasias.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		RNA polymerase II transcription factor binding	4.61E-12
positive regulation of transcription, DNA-dependent	2.75E-31	androgen receptor binding	5.51E-12
regulation of transcription from RNA polymerase II promoter	7.04E-28	activating transcription factor binding	1.35E-11
negative regulation of transcription, DNA-dependent	3.59E-26	transcription corepressor activity	1.66E-11
		ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity	1.44E-09
<b>Función molecular</b>		<b>Componente celular</b>	
transcription coactivator activity	1.47E-19	PML body	1.04E-05
chromatin binding	2.85E-18	Sin3 complex	0.0001
ligand-dependent nuclear receptor binding	4.76E-13	NuRD complex	0.0002
Zinc ion binding	7.10E-13	Rb-E2F complex	0.0006
steroid hormone receptor activity	2.16E-12		
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
	<b>P ajustada</b>		<b>P ajustada</b>
Cell cycle	2.39E-15	Drug interaction with drug	5.21E-24
Pathways in cancer	2.31E-13	cancer or viral infections	3.96E-21
Pancreatic cancer	4.22E-11	Li-Fraumeni syndrome	1.96E-18
TGF-beta signaling pathway	1.42E-10	Breast Diseases	8.75E-17
Chronic myeloid leukemia	2.26E-09	Breast Neoplasms	2.45E-16
Wnt signaling pathway	1.03E-08	Skin Diseases, Genetic	7.81E-15
Colorectal cancer	3.18E-08	Ataxia Telangiectasia	7.75E-12
p53 signaling pathway	4.90E-08	Neoplasms	7.75E-12
Adherens junction	6.71E-08	Retinoblastoma	5.32E-11
Prostate cancer	2.01E-07	Leukemia, Promyelocytic, Acute	2E-10

**Tabla 9. Funciones biológicas del módulo 1**

El módulo 2 (tabla 10) involucra a la regulación de vías metabólicas, sobre todo de lípidos, regulación de la comunicación celular, regulación de la respuesta a estímulos, y espermatogénesis. Además, se encuentra enriquecido en proteínas que están relacionadas con la predisposición genética a enfermedades. Probablemente, la alteración de este módulo esté implicado en el desarrollo de la dislipidemia, sobre todo en el aumento de las concentraciones plasmáticas de LDL que ya han sido reportadas en pacientes con MM.

Gene Ontology			
	P ajustada		P ajustada
<b>Proceso biológico</b>		spermatogenesis	0.0452
regulation of autophagy	0.0031		
brain development	0.0093	<b>Función molecular</b>	
positive regulation of cell communication	0.0173	1-alkyl-2-acetylglycerophosphocholine esterase activity	1.20E-05
lipid catabolic process	0.0254	protein heterodimerization activity	0.0271
positive regulation of response to stimulus	0.0279		
KEGG		Enfermedades asociadas	
Nombre de la vía	P ajustada		P ajustada
Ether lipid metabolism	1.35E-05	Genetic Predisposition to Disease	0.0035
Metabolic pathways	0.0065	Adhesion	0.0035
		Disease Susceptibility	0.0035

**Tabla 10. Funciones biológicas del módulo 2**

El módulo 3 (tabla 11) involucró procesos biológicos relacionados con la diferenciación de melanocitos y se asoció a enfermedades tales como desórdenes de la pigmentación, melanoma y tumores neuroendócrinos, entre otras.

Gene Ontology			
	P ajustada		P ajustada
<b>Proceso biológico</b>		<b>Función molecular</b>	
melanosome localization	7.18E-05	myosin V binding	2.16E-06
melanocyte differentiation	1.00E-04	Rab GTPase binding	1.00E-04
protein targeting	4.30E-03		
KEGG		Enfermedades asociadas	
Ninguna vía enriquecida			P ajustada
		Piebaldism	4.06E-07
		Hypopigmentation	2.20E-05
		Pigmentation Disorders	3.86E-05
		Melanoma	0.0001
		Neuroendocrine Tumors	0.0001
	Translocation, Genetic	0.0003	

**Tabla 11. Funciones biológicas del módulo 3**

Los procesos biológicos enriquecidos por las proteínas del módulo 5 (tabla 12) involucran a la vía de señalización NFkB, comentada anteriormente, a la vía de señalización Toll, constituida por una familia de proteínas que forman parte del sistema inmunitario innato, y a la regulación de la apoptosis. Las funciones moleculares enriquecidas en este módulo estuvieron relacionadas con la unión al Zn, con

la actividad de ubiquitina-ligasa, y con la actividad de MAP cinasa entre otras. Los componentes celulares obtenidos estuvieron relacionados con la membrana plasmática y con la región perinuclear del citoplasma. El módulo 5 participa en la vía de señalización de la MAPK, la cual implica procesos de proliferación, diferenciación, motilidad, apoptosis y metabolismo celular, está relacionado con la diferenciación osteoclástica, con la vía de señalización de la neurotrofina cuya función se relaciona con la supervivencia celular y procesos de crecimiento neuronal, con la vía de señalización RIG-I-like encargada de activar señales antivirales, con la proteólisis mediada por la ubiquitina, entre otras. La necrosis, enfermedades del hueso, la esclerosis múltiple, la agammaglobulinemia y linfomas son enfermedades que se asociaron a este módulo de manera significativa.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		mitogen-activated protein kinase kinase kinase binding	0.0004
negative regulation of NF-kappaB transcription factor activity	1.86E-05	MAP kinase kinase activity	0.0004
Toll signaling pathway	3.87E-05	ubiquitin-specific protease activity	0.0034
toll-like receptor 1 signaling pathway	0.0007	ubiquitin thioesterase activity	0.0082
TRIF-dependent toll-like receptor signaling pathway	0.0007	protein serine/threonine kinase activity	0.0245
positive regulation of signal transduction	0.0008	protein homodimerization activity	0.0417
regulation of programmed cell death	0.0009	cytokine receptor binding	0.0417
membrane protein intracellular domain proteolysis	0.0009		
		<b>Componente celular</b>	
<b>Función molecular</b>		CD40 receptor complex	0.0025
Zinc ion binding	2.90E-05	internal side of plasma membrane	0.0028
ubiquitin-protein ligase activity	4.98E-05	internal side of plasma membrane	0.0028
ubiquitin protein ligase binding	8.26E-05	perinuclear region of cytoplasm	0.0112
thioesterase binding	0.0003		
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
	<b>P ajustada</b>		<b>P ajustada</b>
Neurotrophin signaling pathway	1.31E-08	Necrosis	4.39E-08
RIG-I-like receptor signaling pathway	9.08E-08	Mulibrey nanism syndrome	5.80E-07
Ubiquitin mediated proteolysis	6.08E-07	Multiple Sclerosis	0.0008
Osteoclast differentiation	6.08E-07	Agammaglobulinemia	0.0011
MAPK signaling pathway	7.50E-06	Ectodermal Dysplasia	0.0011
Pathways in cancer	1.36E-05	Bone Diseases, Metabolic	0.0011
Toll-like receptor signaling pathway	1.51E-05	Bone Resorption	0.0013
Chronic myeloid leukemia	0.0006	Lymphoma	0.0013
GnRH signaling pathway	0.0008	Wounds and Injuries	0.0017
ErbB signaling pathway	0.0008	Inflammation	0.0017

**Tabla 12. Funciones biológicas del módulo 5**

La activación de plaquetas, procesos de apoptosis, la activación de células B, la regulación de linfocitos, las vías de señalización de los receptores de insulina, la vía de señalización Fc Epsilon RI, las vías de señalización de receptores de células B y la vía de señalización Erb, cuya sobre-expresión ha sido relacionada con diferentes tipos de cáncer, son algunas de las funciones y vías biológicas enriquecidas en el módulo 7 (tabla 13).

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		1-phosphatidylinositol-3-kinase activity	0.0002
platelet activation	2.31E-05	steroid hormone receptor binding	0.0016
apoptotic process	3.78E-05	signaling adaptor activity	0.0016
B cell activation	0.0004	protein serine/threonine kinase activity	0.0018
insulin receptor signaling pathway	0.0004	phospholipid binding	0.0024
phosphatidylinositol phosphorylation	0.0004	kinase regulator activity	0.0036
regulation of phosphatidylinositol 3-kinase activity	0.0005	protein binding, bridging	0.0037
positive regulation of lymphocyte activation	0.0006	purine ribonucleoside triphosphate binding	0.0051
B cell receptor signaling pathway	0.0006	purine ribonucleotide binding	0.0051
nerve growth factor receptor signaling pathway	0.0006	protein kinase binding	0.0119
<b>Función molecular</b>		<b>Componente celular</b>	
insulin receptor substrate binding	0.0002	1-phosphatidylinositol-4-phosphate 3-kinase, class IA complex	2.16E-06
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
	<b>P ajustada</b>		<b>P ajustada</b>
Fc epsilon RI signaling pathway	4.58E-12	Urogenital Neoplasms	0.0003
B cell receptor signaling pathway	2.65E-09	Glioblastoma	0.0004
ErbB signaling pathway	3.23E-09	Astrocytoma	0.0004
Fc gamma R-mediated phagocytosis	3.32E-09	HIV	0.0004
Leukocyte transendothelial migration	6.23E-09	Pancreatic Neoplasms	0.0005
Natural killer cell mediated cytotoxicity	9.87E-09	Ovarian Neoplasms	0.0005
Chemokine signaling pathway	3.19E-08	Cell Transformation, Neoplastic	0.0005
Focal adhesion	3.51E-08	Gout	0.0005
Aldosterone-regulated sodium reabsorption	7.98E-08	Ovarian Diseases	0.0005
Carbohydrate digestion and absorption	8.32E-08	Central Nervous System Neoplasms	0.0005

**Tabla 13. Funciones biológicas del módulo 7**

El módulo 8 (tabla 14) estuvo enriquecido en la vía de señalización mediada por integrinas las cuales participan mayoritariamente en la unión de las células con la matriz extracelular. La alteración de este proceso biológico es importante para el desarrollo del cáncer dado que favorece el carácter

invasivo de la enfermedad mediante la metástasis de las células cancerosas. Por otro lado, este módulo también involucró a la función molecular relacionada con la regulación de la proteólisis, entre otras, y dentro de las enfermedades que se le ha asociado fueron el cáncer, infecciones virales, melanoma, metástasis, y procesos neoplásicos.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		<b>Función molecular</b>	
integrin-mediated signaling pathway	0.0115	integrin binding	0.0024
cell-matrix adhesion	0.0173	regulation of proteolysis	0.0173
regulation of protein catabolic process	0.0173		
regulation of proteolysis	0.0173	<b>Componente celular</b>	
regulation of phosphate metabolic process	0.0345	cell surface	0.0018
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
			<b>P ajustada</b>
Ninguna vía enriquecida		cancer or viral infections	0.0004
		Nevi and Melanomas	0.0004
		Neoplasm Metastasis	0.0005
		Neoplasm Invasiveness	0.0005
		Melanoma	0.0005
		Neoplastic Processes	0.0008
		Carcinoma	0.0012
		Adhesion	0.0015
		Neoplasms	0.0023

**Tabla 14. Funciones biológicas del módulo 8**

Uno de los procesos biológicos enriquecidos en el módulo 9 (tabla 15) es la recombinación V (D) J basada en un mecanismo de recombinación genética por el cual se seleccionan y ensamblan al azar segmentos de genes que codifican proteínas específicas con papeles importantes en el sistema inmunitario. Esta recombinación genera un repertorio diverso de receptores de linfocitos T (TCR) y moléculas de inmunoglobulina (Ig) que son necesarios para el reconocimiento de diversos antígenos bacterianos, víricos y de parásitos, así como de células disfuncionales, como son las tumorales. Por lo tanto, la afectación de este proceso podría contribuir a la falta de reconocimiento de las células tumorales de mieloma conllevando al progreso de la enfermedad. Las funciones moleculares involucraron a la unión a histonas, a la actividad endonucleasa y la unión al ADN. La afectación de estas funciones, que están relacionadas con la reparación del ADN, y con la regulación de la expresión génica, también podría estar contribuyendo al desarrollo del MM. Por otro lado, resultó estar

enriquecida la vía asociada a la inmunodeficiencia primaria y este módulo involucró de manera significativa enfermedades como la inmunodeficiencia combinada severa, linfomas y enfermedades hematológicas, entre otras.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		<b>Función molecular</b>	
V(D)J recombination	6.37E-08	histone binding	0.0009
pre-B cell allelic exclusion	1.23E-06	endonuclease activity	0.0009
T cell differentiation in thymus	0.0003	DNA binding	0.0153
T cell differentiation in thymus	0.0003		
regulation of hydrolase activity	0.0181		
Primary immunodeficiency	1.92E-06		
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
	<b>P ajustada</b>		<b>P ajustada</b>
Primary immunodeficiency	1.92E-06	Chickenpox	1.35E-06
		Severe Combined Immunodeficiency	1.70E-05
		Eosinophilia	2.06E-05
		Waldenstrom Macroglobulinemia	2.75E-05
		Hodgkin Disease	9.27E-05
		Lymphoma, Non-Hodgkin	0.0001
		Lymphoma, Low-Grade	0.0002
		Lymphatic Diseases	0.0003
		Lymphoma	0.0003
		Hematologic Diseases	0.0003

**Tabla 15. Funciones biológicas del módulo 9**

El módulo 10 (tabla 16) solo se asoció de manera significativa a enfermedades como tumores neuroectodérmicos, al osteosarcoma, al neuroepitelioma periférico y al sarcoma.

<b>Gene Ontology</b>			
No hubo enriquecimiento			
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
			<b>P ajustada</b>
Ninguna vía enriquecida		Neuroectodermal Tumors, Primitive	1.22E-05
		Osteosarcoma	5.24E-05
		Peripheral neuroepithelioma	5.73E-05
		Sarcoma	6.79E-05

**Tabla 16. Funciones biológicas del módulo 10**



El módulo 11 (tabla 17) estuvo relacionado con el ensamblaje del complejo de corte y empalme (“splicing”), con la unión a ARN y a metales, estuvo asociado con la vía involucrada con el proceso de corte y empalme y no le fue asociada ninguna enfermedad de manera significativa.

<b>Gene Ontology</b>		
	<b>P ajustada</b>	<b>P ajustada</b>
<b>Proceso biológico</b>		metal ion binding
spliceosomal complex assembly	0.0003	0.048
		<b>Componente celular</b>
<b>Función molecular</b>		catalytic step 2 spliceosome
RNA binding	0.024	0.0017
		nucleoplasm part
		0.0496
<b>KEGG</b>		<b>Enfermedades asociadas</b>
	<b>P ajustada</b>	
Spliceosome	8.55E-05	No hay enfermedades asociadas

**Tabla 17. Funciones biológicas del módulo 11**

Algunos de los procesos biológicos que se asociaron al módulo 12 (tabla 18) involucraron a la regulación positiva de la proliferación epitelial que está relacionada con la morfogénesis del pulmón y a la regulación positiva de la proliferación celular del mesénquima. Por lo tanto, este módulo parece estar enriquecido en procesos de regulación que, aunque no involucran de manera directa a la patogénesis del MM, parecen contribuir de manera importante a la generación de procesos que favorecen la proliferación celular cuya afectación podría conducir al desarrollo del cáncer, o particularmente, al MM en los pacientes que tengan este módulo alterado. Otros procesos y funciones asociados de manera significativa son la regulación de la transcripción, la unión al ADN, la heterodimerización y homodimerización de proteínas y la unión al Zn. El Síndrome de Fisher es una de las enfermedades asociadas a este módulo al igual que otros desórdenes cerebrales. El síndrome de Fisher es una enfermedad autoinmune donde el organismo produce anticuerpos contra las neuronas del sistema nervioso periférico. En este sentido, aunque esta enfermedad no se relaciona con el desarrollo tumoral se relaciona con la disfunción de las inmunoglobulinas, las cuales son parte la fisiopatología del MM.

<b>Gene Ontology</b>			
	<b>P ajustada</b>		<b>P ajustada</b>
<b>Proceso biológico</b>		double-stranded DNA binding	8.67E-06
positive regulation of epithelial cell proliferation involved in lung morphogenesis	2.98E-05	protein heterodimerization activity	5.61E-05
positive regulation of mesenchymal cell proliferation	0.0003	protein homodimerization activity	0.0001
regulation of sequence-specific DNA binding transcription factor activity	0.0003	sequence-specific DNA binding	0.0002
skeletal muscle tissue development	0.0025	Zinc ion binding	0.0038
<b>Función molecular</b>		<b>Componente celular</b>	
DNA binding, bending	1.42E-06	transcription factor complex	0.0001
RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity	4.46E-06		
<b>KEGG</b>		<b>Enfermedades asociadas</b>	
No hay vías enriquecidas			<b>P ajustada</b>
		Fisher's syndrome	2.71E-06
		Language Development Disorders	1.17E-05
		Language Disorders	1.32E-05
		Speech Disorders	1.32E-05
		Learning Disorders	1.32E-05
		Communication Disorders	2.51E-05
	Chromosome Aberrations	0.0002	

**Tabla 18. Funciones biológicas del módulo 12**

El módulo 15 (tabla 19) solo se asoció de manera significativa al proceso de mitosis. Los módulos 4, 6, 13 y 14 no estuvieron enriquecidos de manera significativa en ningún proceso biológico, función molecular, componente celular, vías metabólicas o de señalización, ni se les ha asociado ninguna enfermedad. A través de este análisis pudimos observar que el interactoma del Zinc involucra módulos con funciones importantes dentro del proceso de desarrollo del cáncer y que también están relacionados con la disfunción de las inmunoglobulinas. Interesantemente, las funciones y vías obtenidas a través del interactoma del Zn y sus módulos que estuvieron relacionados con el procesamiento de ARN, la homeostasis de las proteínas, la vía NFkB y con las enzimas modificadoras de histonas, ya fueron identificadas a través de los análisis de secuenciación de los genomas de Mieloma Múltiple [2]. Los módulos relacionados con la regulación del sistema inmune y de las células B de manera específica (módulos 7 y 9) no han sido reportados todavía en MM.

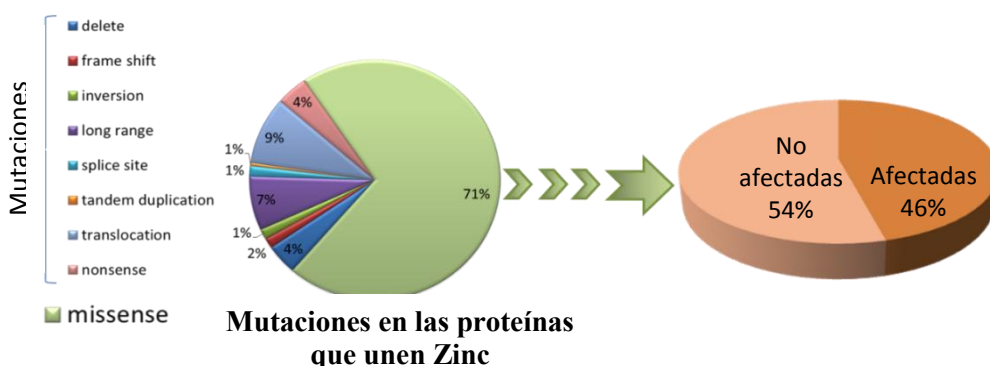
Gene Ontology			
	P ajustada		P ajustada
<b>Proceso biológico</b>		microtubule binding	0.0003
mitosis	0.0033		
		<b>Componente celular</b>	
<b>Función molecular</b>		microtubule	0.0072
KEGG		Enfermedades asociadas	
No hay vías enriquecidas		No hay enfermedades asociadas	

Tabla 19. Funciones biológicas del módulo 15

### 12.1.5. Identificación de las proteínas afectadas en el interactoma

La predicción del efecto de las mutaciones en la estructura y/o función de las proteínas que unen Zinc se llevó a cabo mediante la integración de los resultados obtenidos de los programas FoldX, FIS, SIFT, SNAP, PolyPhen-2 y Consurf basados en estructura y conservación de secuencia (ver la sección de metodología). Los resultados de las predicciones de los 6 programas bioinformáticos se ilustran en el anexo 1.

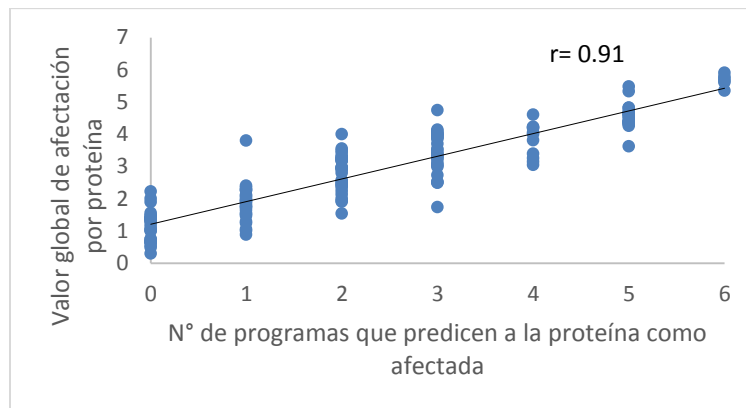
Mediante este análisis pudimos determinar que 71 proteínas con dominios de unión de Zn podrían estar siendo afectadas en su estructura y/o función por mutaciones de cambio de aminoácido “missense”, 9 proteínas afectadas por mutaciones de codón de paro “nonsense” y 52 afectadas por deleciones, duplicaciones en tándem, en el sitio de corte y empalme (“splice site”), inversiones, translocaciones e inserciones (Figura 4) (ver detalles en metodología). Por lo tanto, este análisis sugiere que la actividad original de estas 132 proteínas pudiera estar siendo afectada.



**Figura 4. Resumen del análisis de estructura-función.** De las 216 proteínas de unión a Zn mutadas, el 71.7 % tiene mutaciones de cambio de aminoácido “missense” (155 proteínas), y el 28.3% (61 proteínas) tiene mutaciones de tipo deleción, por desplazamiento de marco de lectura (“frameshift”), inversión, “long range”, en el sitio de corte y empalme (“splice site”), duplicación en tándem, translocación y de codón de paro. Las proteínas con mutaciones de cambio de aminoácido fueron analizadas a través de los diferentes programas de predicción de estructura-función. El 46% (71 proteínas) de las proteínas con mutaciones de cambio de aminoácido “missense” resultaron estar afectadas. En total, el 61% (132 proteínas) de las proteínas de unión a Zn mutadas en MM resultaron estar afectadas.

### 12.1.6. Integración de los resultados de los predictores de E-F en el valor global de afectación

La aplicación de un valor global de afectación para cada proteína con mutaciones de cambio de aminoácido se realizó con la finalidad de integrar los resultados de predicción de las 6 herramientas bioinformáticas y obtener un resultado de manera cuantitativa y, a partir de éste, poder tener un valor global de afectación que integre la información de todos los predictores y estimar de manera cuantitativa el grado de afectación de la proteína. Para ello, se aplicó una fórmula que integra los resultados de los programas de predicción (ver sección de materiales y métodos y anexo 2) de la cual obtuvimos un valor global de afectación. Este valor aumenta a medida que aumenta el número de programas que predicen que la proteína está afectada con una  $r=0.91$ , tal como se presenta en el gráfico 6. El valor global de afectación obtenido para cada proteína mutada se muestra en la tabla 20.



**Gráfico 6. Correlación entre los valores globales de afectación y número de predictores.** Cuanto mayor es la cantidad de programas que predicen que la mutación afecta la E-F de la proteína, el valor global de afectación tiende a ser alto en comparación con las proteínas que no están afectadas. Correlación de Pearson:  $r= 0.91$ .

Este gráfico demuestra que el valor global de afectación permite analizar cuantitativamente el grado de afectación de cada proteína mutada. Por ejemplo, como puede verse en la tabla 20, veinte proteínas se estiman como afectadas por cinco de los seis predictores totales, sin embargo, al aplicarles la fórmula integrada de los predictores, se puede ver que el valor global de afectación de cada proteína es diferente, por lo tanto, es posible diferenciar su grado de afectación de las otras proteínas que tienen el mismo número de predictores que estiman que están afectadas.

Nombre de la proteína	Mutación de cambio de aminoácido	Número de predictores que estiman que está afectada	Valor global de afectación
FNTB	R291C	6	5.90
WHSC1	C1191F	6	5.78
BRCA1	C44F	6	5.72
TIMP2	G114E	6	5.66
TP53	Y205D	6	5.61
TRAF3	D551V	6	5.34
IDE	H134P	5	5.47
ZNF484	G474R	5	5.33
NR3C2	F626C	5	4.82
RNF26	C404F	5	4.80
ADAMTSL3	G497R	5	4.71
ADAM29	M356T	5	4.66
ZNF569	H516Y	5	4.63
ADAMTS1	G594A	5	4.62
USP19	N535Y	5	4.62
MT1A	C36F	5	4.53
RBM5	W547L	5	4.53
RNF168	L71H	5	4.42
MLL3	Y987H	5	4.39
STAC	R169C	5	4.38
AFG3L2	I486V	5	4.37
ZNF416	Y522N	5	4.35
ZNF193	Y310C	5	4.32
MYT1	M1035T	5	4.31
MID2	R105H	5	4.25
TP53	C275S	5	3.62
MMP16	Y459D	4	4.60
RAG1	G954D	4	4.23
ZNF830	W365C	4	4.20
ASH1L	K1355N	4	4.18
PRKCZ	Y126D	4	4.05
MLL	M2022I	4	4.00
RNF213	E2081K	4	3.81
CHD3	V1462I	4	3.39
ZNF492	P420T	4	3.25
ZKSCAN5	E427K	4	3.14
NR3C2	E785K	4	3.04
TRIM10	T41P	3	4.74
ZNF483	R543Q	3	4.13
ADAM18	E457K	3	4.09
CTCF	G510V	3	4.02

Nombre de la proteína	Mutación de cambio de aminoácido	Número de predictores que estiman que está afectada	Valor global de afectación
ZZEF1	D2773H	3	4.01
ASH2L	F254S	3	3.99
ADAMTSL1	W1710C	3	3.98
CLIP1	I466M	3	3.94
ZNF878	S510C	3	3.90
PRICKLE2	R390W	3	3.90
ZBTB3	R503C	3	3.85
WHSC1L1	Y887C	3	3.70
TRIM17	V132L	3	3.54
WHSC1L1	K888N	3	3.50
EHMT1	A988T	3	3.41
ZNF527	Y302N	3	3.38
OSR2	T55M	3	3.33
SRRM2	S510F	3	3.31
WHSC1	A332D	3	3.24
EGR1	Q9H	3	3.24
VAT1L	G262V	3	3.12
ZNRF3	A807P	3	3.05
ADAMTS9	R1431H	3	3.02
METAP2	A2S	3	2.98
ZC3H14	E339K	3	2.72
SF3A3	D144Y	3	2.53
RIMS2	R354W	3	2.49
FOXP1	P407L	3	2.49
HNF4A	R331H	3	1.74
PDZD8	R138H	2	4.00
RORB	C299F	2	3.55
MYO9A	T1974P	2	3.48
SP8	S84F	2	3.42
SOBP	R245G	2	3.30
TRAFD1	N519K	2	3.30
ZNF530	Q398E	2	3.24
AGAP2	G506E	2	3.22
FOXP2	S100T	2	3.20
ADAM9	R53G	2	3.18
ZNF319	H528R	2	3.16
MBD1	G471A	2	2.97
RELN	D207N	2	2.97
PDE8B	V344G	2	2.93
TSHZ3	R1004W	2	2.93
ADAMTS9	N663T	2	2.89

Nombre de la proteína	Mutación de cambio de aminoácido	Número de predictores que estiman que está afectada	Valor global de afectación
UPF1	A771T	2	2.79
ADNP2	L565S	2	2.57
ZNF83	E293V	2	2.53
INTS12	M445I	2	2.46
ZNF559	E326K	2	2.45
ZNF700	V134F	2	2.40
ZZEF1	H1877P	2	2.35
ADAMTS18	S487I	2	2.24
MLPH	V112F	2	2.13
DNAJA1	Q159P	2	2.00
TSHZ3	S58G	2	1.93
TRIM47	G445E	2	1.90
ESR1	Q498P	2	1.53
ZKSCAN3	Q361H	1	3.80
TET1	H1547P	1	2.40
ADAMTS20	D1514G	1	2.34
ST18	E4K	1	2.28
SP140	S474T	1	2.26
ZNF749	E358Q	1	2.25
UBA2	R612K	1	2.11
PHC1	V250L	1	2.08
ZNF667	T213A	1	2.02
MKRN3	D483V	1	1.95
TSHZ2	L757Q	1	1.94
ZBTB33	I395M	1	1.91
SMAP2	N200K	1	1.85
ZZEF1	T2924N	1	1.78
ZBTB45	P384L	1	1.75
USP22	L114V	1	1.73
TET1	Q1504H	1	1.64
ZC3H15	Q264R	1	1.55
ZRANB3	R171S	1	1.52
MDM4	Q262H	1	1.46
ZNF417	C30Y	1	1.29
LMO2	E135D	1	1.23
PDE3A	N681I	1	1.04
MYSM1	Q387H	1	0.99
NR2F2	S223L	1	0.88
ZNF646	Q1397R	0	2.22
RIMS2	P973L	0	2.00
RNF19A	I795V	0	1.89

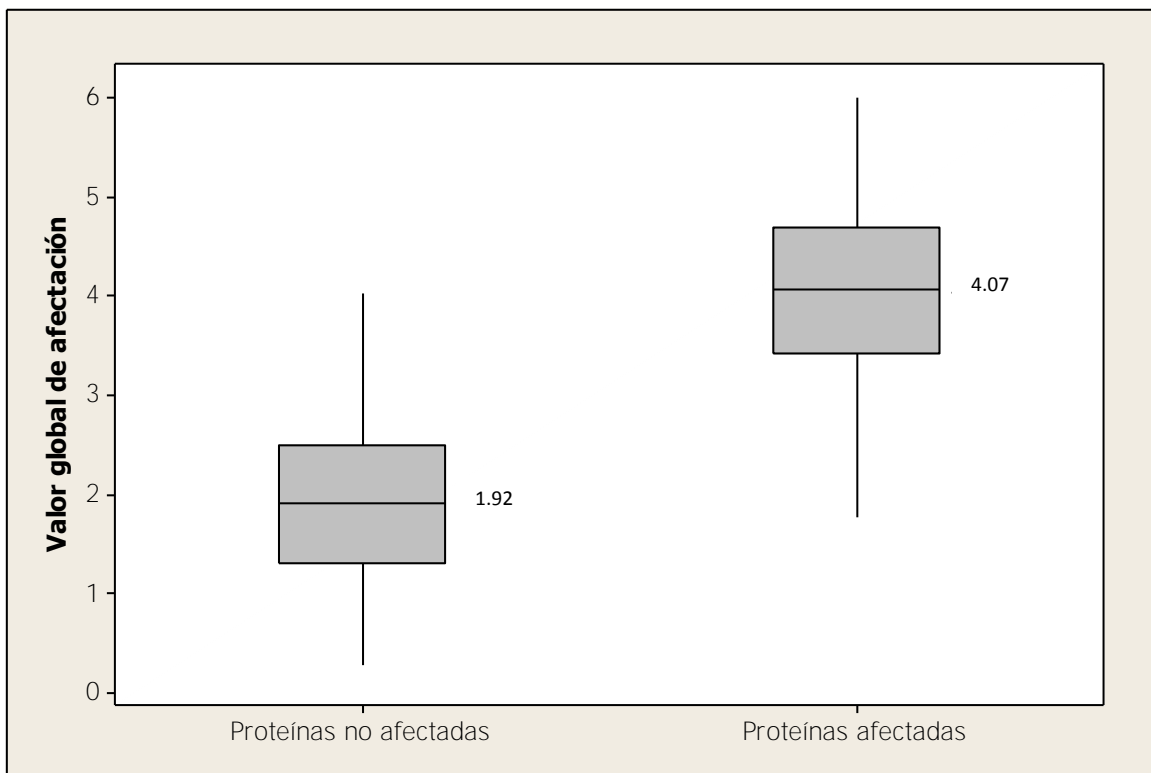
Nombre de la proteína	Mutación de cambio de aminoácido	Número de predictores que estiman que está afectada	Valor global de afectación
ADNP	D814N	0	1.55
POLR3A	D66H	0	1.47
DNMT3B	K421E	0	1.45
CPSF3	S589L	0	1.39
ZNF492	T325N	0	1.38
CHD3	G1755V	0	1.38
MIPEP	L171F	0	1.37
CHD3	P2049S	0	1.32
ADAM17	D161N	0	1.31
ZNF148	S591A	0	1.31
TSHZ2	F634L	0	1.21
ZNF423	E426Q	0	1.10
TADA2A	E33K	0	1.03
ZNF804A	E338K	0	1.02
ARID2	M1625I	0	0.99
ZFPM2	S647P	0	0.74
ZCCHC5	E217K	0	0.74
TRIM29	S301N	0	0.67
ZMYM4	I1182V	0	0.65
ZNF630	Q394K	0	0.65
ZC3H18	S53T	0	0.64
GIT1	S368G	0	0.55
ASH1L	M1356L	0	0.48
ZFYVE1	M152V	0	0.30

**Tabla 20. Proteínas del Zinc con mutaciones de cambio de aminoácido y sus valores de predicción global de afectación correspondientes.** Las proteínas están ordenadas de acuerdo al número de programas que predicen que está afectada su E-F. Es posible ver que cuanto mayor es el número de predictores, mayor es el valor de afectación global para cada proteína.

Posteriormente, se formó un grupo de proteínas afectadas y se comparó con un grupo de proteínas no afectadas. El grupo de proteínas afectadas está compuesto por aquellas que cuentan con al menos 4 de 6 programas que predicen que la proteína está afectada o 3 de 5 programas cuando no se cuenta con estructura 3D, mientras que el grupo de las proteínas no afectadas está integrado por aquellas que cuentan con menos de 4 de 6 programas que predicen que la proteína está afectada o menos de 3 de 5 programas cuando no se cuenta con estructura.

Finalmente, determinamos los valores globales de afectación para cada grupo de proteínas (afectadas y no afectadas) y obtuvimos con un nivel de confianza del 95% que las proteínas afectadas tienen valores de predicción que van de 3.77 a 4.19 y una mediana de 4.074, y que las proteínas no afectadas tienen valores de predicción que van de 1.75 a 2.14 y una mediana de 1.92 (gráfico 7).





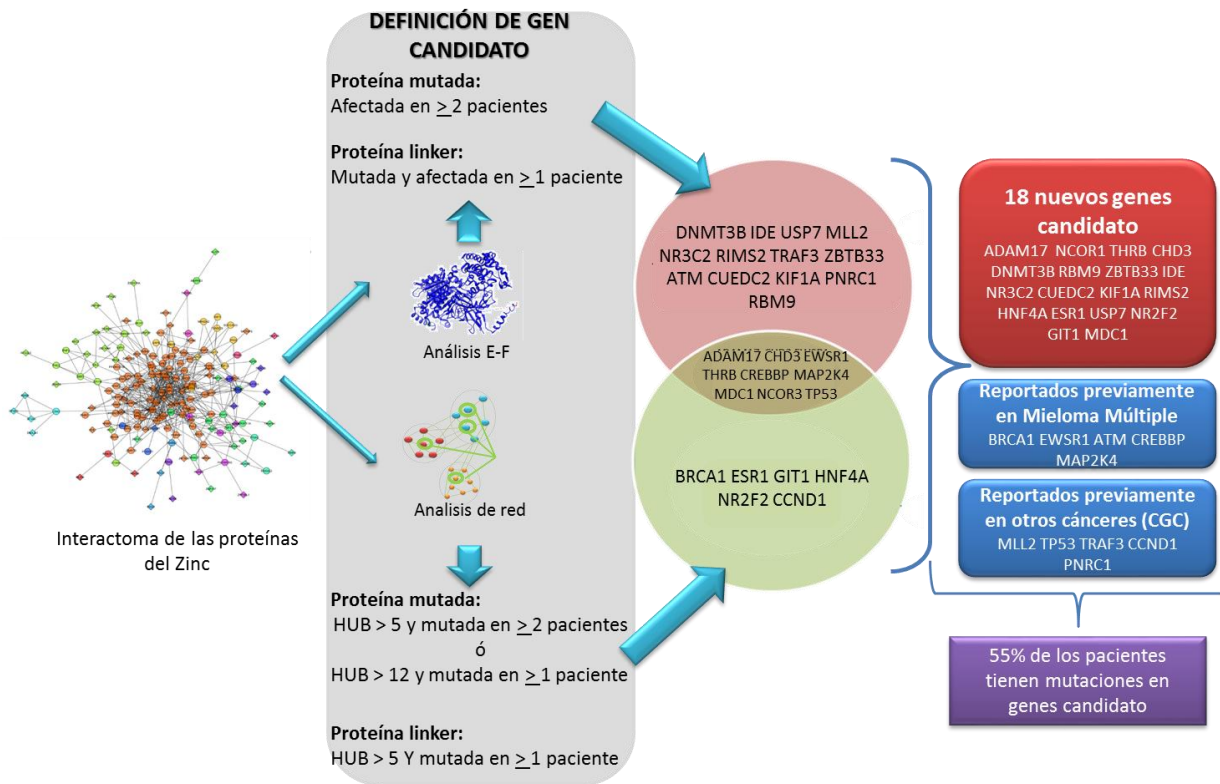
**Gráfico 7. Comparación del valor global de afectación de las proteínas afectadas y no afectadas.** Los valores de predicción de las proteínas afectadas tienden a agruparse en valores más altos que las proteínas no afectadas. Las medianas se presentan al lado de cada caja.

A través de estos análisis fue posible integrar los resultados de los predictores en un valor de predicción para cada proteína mutada. Cuanto mayor es el valor de éste, es posible predecir que la proteína se encuentre más afectada.

### 12.1.7. Determinación de genes candidato

Debido a que los genes candidato suelen estar enriquecidos en mutaciones que afectan su función y/o que tienen la característica de estar altamente conectados en los interactomas [17][27][44], se seleccionaron los genes candidato acorde a sus propiedades determinadas por el análisis de estructura-función y el análisis de interactomas tal como se muestra en la figura 5.

A partir de los análisis de estructura-función y de interactomas sobre las proteínas del interactoma del Zinc, pudimos determinar la existencia de 28 proteínas o genes candidato, de los cuales 13 son candidato por el análisis de estructura-función, 6 son candidato por su condición de HUB y 9 lo son tanto por el análisis de estructura-función como por su condición de HUB (figura 5).



**Figura 5. Genes candidato en Mieloma Múltiple.** A partir de las proteínas presentes en el interactoma del Zinc se realizó el análisis de E-F y HUB a cada una de ellas. Las proteínas o genes que cumplieron las condiciones de candidato por el análisis de E-F se presentan en el círculo rosa. Las que cumplieron las condiciones de gen candidato por el análisis de HUB se encuentran en el círculo verde. Los genes candidato por ambas condiciones (E-F y HUB) se presentan en el círculo café. Finalmente 18 de los 28 genes candidato totales no han sido reportados todavía como tales en Mieloma Múltiple, 5 ya han sido reportados como candidato en MM y otros 5 ya han sido reportados como candidato en otros cánceres. En total, el 55% de los pacientes con mieloma del MMRC tienen mutaciones en los genes candidato.

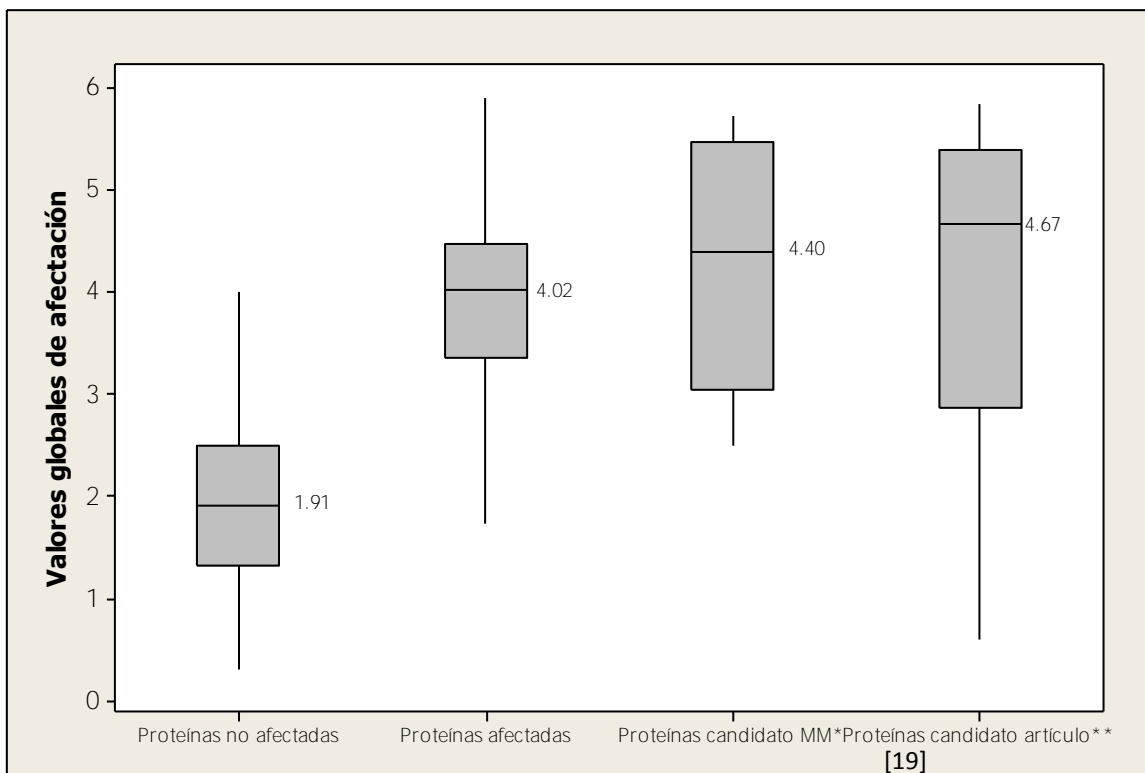
Interesantemente, de los 28 genes candidato, 18 no han sido reportados todavía como tales en Mieloma Múltiple (tabla 21) mientras que 5 si han sido reportados como candidato en esta enfermedad. Además, otros 5 ya fueron reportados como candidato en otros tipos de cáncer. Por lo tanto, nuestra estrategia sugiere ser viable para descubrir nuevos genes candidato mediante la integración de diferentes datos ómicos. Posteriormente, evaluamos qué tan esenciales son los nuevos genes candidato que descubrimos mediante un análisis de viabilidad celular. Dicho análisis también se presenta en la tabla 21.

Gen	<sup>a</sup> Condición	Nº de pacientes con el gen mutado	<sup>b</sup> Esencialidad	<sup>c</sup> OMIM	<sup>d</sup> Fenotipo en ratón
ADAM17	MUTADO	2	Esencial	No	Letal
CHD3	MUTADO	3	No	No	N/A
NCOR1	LINKER	1	Esencial	No	Letal
THRB	MUTADO	1	No	OMIM	Disfunción tiroidea
CUEDC2	LINKER	1	No	No	N/A
NR3C2	MUTADO	2	Esencial	OMIM	Letal
DNMT3B	MUTADO	2	Esencial	OMIM	Letal
RBM9	LINKER	1	No	No	N/A
ESR1	MUTADO	1	No	OMIM	Infertilidad
HNF4A	MUTADO	1	Esencial	OMIM	Letal
NR2F2	MUTADO	1	Esencial	No	Letal
IDE	MUTADO	2	No	No	Acumulaciones B amiloides
KIF1A	LINKER	1	Esencial	No	Letal
RIMS2	MUTADO	2	No	No	Letal y metabolismo de glucosa alterado
ZBTB33	MUTADO	2	No	No	Normal
MDC1	LINKER	1	No	No	Células T y B anormales
GIT1	MUTADO	1	Esencial	No	Letal
USP7	LINKER	1	Esencial	No	Letal

**Tabla 21. Nuevos genes candidato en Mieloma Múltiple.** <sup>a</sup>La condición del gen hace referencia a si el gen candidato es un linker o es un gen mutado en el interactoma del Zinc. <sup>b</sup>La esencialidad se refiere a que son genes importantes para las funciones básicas celulares [1]. <sup>c</sup>OMIM es un grupo de genes relacionados con enfermedades mendelianas. <sup>d</sup>El fenotipo de ratón (<http://www.informatics.jax.org/>) permite definir el tipo de alteración ocasionada en el ratón cuando un gen específico es mutado. El fenotipo normal en el gen ZBTB33 hace referencia a machos y hembras viables sin cambios detectables en la composición de las células sanguíneas. N/A: no existe análisis de fenotipo en ratón para ese gen.

Debido al alto grado de homología que comparten los genes humanos con los de ratón, fue interesante analizar de qué manera afecta la mutación de estos genes al fenotipo del ratón (tabla 21). A partir de los resultados presentados en la tabla 21 pudimos determinar que la mayoría de los genes candidato modifican el fenotipo de ratón normal cuando están mutados y también encontramos que la mayoría de los genes que alteran el fenotipo de ratón son esenciales para la viabilidad celular y/o están relacionados con algunas enfermedades mendelianas. Esto podría sugerir que los nuevos genes candidato propuestos para MM son importantes para una adecuada función celular normal y, por lo tanto, puedan ser probables blancos de ser afectados en el proceso del cáncer.

Por otro lado, seleccionamos los valores globales de afectación de los genes candidato y observamos que sus valores se encuentran entre 3.47 y 5.07 con un 95% de confianza y presentan una mediana de 4.40. Interesantemente, los valores globales de afectación en los genes candidato tienden a ser más altos que los de las proteínas afectadas que no son candidato (mediana= 4.02). En el gráfico 7 se pueden observar estas comparaciones. Por otro lado, con la finalidad de analizar si los genes que consideramos candidato en MM comparten características con otros genes ya conocidos como candidato en otros tipos de cáncer, calculamos el valor global de afectación a un grupo de genes que fueron seleccionados como candidato en otros cánceres por su afectación en su E-F [19] y comparamos sus valores con los de los genes candidato que nosotros obtuvimos para Mieloma Múltiple y que también están afectados en su E-F (gráfico 8).



**Gráfico 8. Comparación de los valores globales de afectación.** Se compararon los valores globales que obtuvimos para las proteínas no afectadas en su estructura-función, para las proteínas afectadas en su estructura-función, para las \*proteínas que seleccionamos como candidato en MM por su afectación en estructura-función y para las \*\*proteínas seleccionadas previamente como candidato en cáncer de mama y cáncer colorectal [19]. El valor de la mediana se encuentra al lado de cada caja.

Como era de esperarse, la mediana del valor global de afectación de las proteínas no afectadas (mediana= 1.91) tiende a ser baja con respecto a los demás grupos de proteínas afectadas. Por otro lado, se puede ver que las proteínas afectadas en su E-F en MM tienen una mediana del valor global de afectación de 4.02 que es menor a la mediana que presentan las proteínas que obtuvimos como candidato en MM (mediana= 4.40). Esta diferencia en los valores fue la que esperábamos dado que la mayoría de las mutaciones candidato conocidas en el cáncer afectan ampliamente la estructura y/o función de las proteínas involucradas [19]. Por otro lado, se puede ver que tanto las proteínas que obtuvimos como candidato en MM (mediana= 4.40) como las proteínas candidato en otros tipos de cáncer (mediana=4.67), comparten la característica de tener valores globales de afectación altos. Esto sugiere que nuestra estrategia de selección de genes candidato correlaciona con las características observadas para otros genes candidato. Además, debido a la tendencia de las proteínas candidato a tener valores globales de afectación altos, es posible pensar que el valor global de afectación permite priorizar a las variantes funcionales más importantes de acuerdo al valor más alto obtenido. Con el objetivo de conocer cuáles de los nuevos genes candidato cumplen con más criterios para ser candidato

(y, por lo tanto, ser más importantes), en la tabla 22 fueron ordenados según su tipo de mutación y, en los casos de tratarse de mutaciones de cambio de aminoácido, por su valor de afectación global. Para aquellos genes que son candidato por el análisis de E-F y que además son HUB en el interactoma del Zn se les dio mayor peso como candidato dado que resultaron ser importantes por ambas estrategias de análisis (tabla 22).

<sup>a</sup> Valor global de afectación	Gen candidato	Condición de candidato
Desplazamiento del marco de lectura	ADAM17	E-F y HUB
inversión/”long range”	NCOR1	E-F y HUB
Codón de paro	THRB	E-F y HUB
3.39	CHD3	E-F y HUB
Sitio de corte y empalme	DNMT3B	E-F
Delección	RBM9	E-F
Inversión	ZBTB33	E-F
5.47	IDE	E-F
4.82	NR3C2	E-F
4.40	CUEDC2	E-F
3.03	KIF1A	E-F
2.49	RIMS2	E-F
1.73	HNF4A	HUB
1.53	ESR1	HUB
1.40	MDC1	HUB
1.03	USP7	HUB
0.88	NR2F2	HUB
0.54	GIT1	HUB

**Tabla 22. Los nuevos genes candidato ordenados en base a los criterios para ser gen candidato.** Los genes fueron ordenados de acuerdo al tipo de mutación, al valor global de afectación que presentan y de acuerdo a su condición de candidato (HUB, E-F o HUB y E-F). Los genes con los más altos valores globales de afectación o con mutaciones deletéreas o con ambas condiciones para ser candidato, fueron ubicados en la parte de arriba de la tabla. De esta forma los genes están ordenados desde aquellos que cumplen más condiciones para ser candidato hasta los que cumplen menos condiciones.

<sup>a</sup>Aquellos genes que no tienen un valor, es porque no tienen mutaciones de cambio de aminoácido.

A partir de este análisis podría considerarse que los genes candidato más importantes en MM (por cumplir un mayor número de características para serlo) son ADAM17, NCOR1, THRB y CHD3, y así sucesivamente en el orden descrito en la tabla 22.

### **12.1.8. Caracterización de los nuevos genes candidato y su relación con Mieloma Múltiple**

Dentro de las funciones de los genes candidato se pueden distinguir aquellas que comprenden mecanismos que permiten el desarrollo del cáncer en general y a otras que involucran procesos biológicos directamente asociados con mecanismos que participan dentro de la fisiopatología del Mieloma Múltiple:

#### **✓ CHD3, NCOR1 y ZBTB33**

Cada uno de estos tres genes participa en la remodelación de la cromatina a través de la desacetilación de histonas y en la regulación de la transcripción. A través de nuestro análisis pudimos observar que los tres genes se encuentran afectados en su estructura-función y que además dos de ellos (CHD3 y NCOR1) tienen la característica de ser HUBs en el interactoma del Zinc, por lo tanto, la alteración de estos genes podría estar afectando todas las interacciones proteína-proteína que involucra la afectación funcional de las proteínas con las que interactúan. Interesantemente, la secuenciación de los genomas de 38 pacientes con Mieloma Múltiple descubrió genes enriquecidos en mutaciones cuya función involucra a enzimas modificadoras de histonas [2]. Se ha observado que la desregulación de las enzimas modificadoras de histonas en Mieloma Múltiple conlleva a una sobre-expresión del gen HOXA9 [2], un factor de transcripción cuya expresión suele estar controlada por la metilación de histonas. Usualmente, HOXA9 se encuentra apagado durante el desarrollo linfoide y sobre-expresado en algunos mielomas, donde se ha visto su importancia para la supervivencia de las células de mieloma [2]. Esto llevó a HOXA9 a ser considerado un oncogén candidato en esta enfermedad. Por lo tanto, la afectación de los 3 genes candidato mencionados podría estar relacionada con la sobre-expresión de este oncogén a través de los mecanismos de desregulación de la cromatina y de la transcripción y, consecuentemente, en la supervivencia de las células de mieloma. ZBTB33 contribuye en la represión de los genes diana de la vía de señalización Wnt. A través de nuestro análisis, este gen se encontró afectado por la mutación, por lo tanto, su afectación podría estimular la proliferación y migración celular.

#### **✓ DNMT3B**

Este gen participa en la activación de la expresión génica de BAG1 el cual inhibe la apoptosis y también DNMT3B aumenta la actividad de BCL2, gen involucrado en suprimir la apoptosis en una variedad de sistemas celulares incluyendo células hematopoyéticas. Además este gen está involucrado

en la metilación *de novo* de todo el genoma. El 96% (Tabla 25) de los pacientes analizados en el presente estudio tienen expresado el gen DNMT3B, sin embargo en nuestro análisis de estructura-función esta proteína se encuentra afectada en dos pacientes, sugiriendo su posible rol en la desregulación de las funciones apoptóticas de BAG1 y BCL2.

✓ **MDC1**

MDC1 es un gen que detiene el ciclo celular en respuesta al daño del ADN y también está involucrado en su reparación. Además, interviene en la estabilización del TP53 y en la apoptosis. Este gen candidato es un HUB dentro del interactoma del Zinc, por lo tanto, los pacientes que lo tengan afectado, probablemente se encuentren con alteraciones en la vía en la que MDC1 está involucrado favoreciendo el mecanismo de supervivencia de las células de mieloma.

✓ **USP7**

USP7 es una hidrolasa que promueve la ubiquitinación de TP53 y su degradación proteosomal. Este gen es un HUB del interactoma, lo que sugiere que la expresión observada en el 100% de los pacientes podría estar favoreciendo la supervivencia de las células malignas a través de la degradación del TP53.

✓ **GIT1**

GIT1 está implicado en la organización del citoesqueleto e incrementa la migración celular, procesos necesarios para el mantenimiento funcional de una célula. Este gen es un HUB en el interactoma del Zinc por lo que su expresión en el 100% de los pacientes podría jugar un papel importante en mantener la estructura celular dada la alta tasa de duplicación de las células de mieloma.

✓ **THRB**

Este gen es un receptor hormonal nuclear que puede actuar como represor o activador de la transcripción. En diferentes tipos de cánceres se ha encontrado que este gen suele estar sub-expresado o disfuncional por lo que se ha sugerido que puede actuar como supresor de tumores [61]. La pérdida de función de THRB se ha asociado con crecimiento celular descontrolado y pérdida de diferenciación celular [61]. Interesantemente, a través de nuestro análisis de estructura-función encontramos que THRB está afectado y que además es HUB, coincidiendo con lo reportado en la literatura [61]. Por lo tanto, la disfunción de este gen en los pacientes con MM, probablemente contribuya en la supervivencia de las células malignas.



### ✓ ADAM17 e IDE

Dentro de la fisiopatología del Mieloma Múltiple se observa la formación de anticuerpos procoagulantes y secreción aumentada de citocinas proinflamatorias (interleucina 6 y TNF) [62]. ADAM17 juega un papel importante dentro de estos procesos ya que inhibe al receptor soluble de TNF-alfa aumentando la función de sus receptores de membrana. Consecuentemente, aumenta la actividad de TNF-alfa lo que promueve la activación y adhesión plaquetarias. En el Mieloma Múltiple los niveles de proteína M elevadas (>16 g/l) están relacionados con una mayor incidencia de complicaciones tromboticas [63]. Por otro lado, ADAM17 también es responsable de la liberación proteolítica de JAM3 (mediador de la angiogénesis) y de P75 involucrado en la señalización del receptor de antígeno, tanto para la expansión clonal como en la eliminación de las células linfoides. A través de nuestro análisis de genes candidato, ADAM17 resultó ser HUB y estar afectado. Esto podría sugerir que la disfunción de este gen podría prevenir el desarrollo de eventos tromboticos y que la expansión de células de mieloma esté más limitada en los pacientes con MM, a diferencia de aquellos que lo tengan sobre-expresado.

Por otro lado, durante el curso de la enfermedad, los pacientes con Mieloma Múltiple pueden desarrollar amiloidosis. La amiloidosis es una enfermedad en la que las proteínas se acumulan en órganos tales como corazón o riñones, dando lugar a daño orgánico y complicaciones asociadas con algunos de los tratamientos para el Mieloma Múltiple. Esta enfermedad es causada por la producción de un exceso de fragmentos de anticuerpos anormales (cadenas ligeras) liberados por las células plasmáticas. Estas cadenas ligeras se pegan unas a otras y se acumulan en distintos órganos del cuerpo. La sobreproducción de anticuerpos anormales pone a estos pacientes en riesgo de padecer amiloidosis. ADAM17 e IDE están involucrados en procesos que regulan la liberación de proteínas amiloides. Ambos genes se encuentran afectados en su E-F lo que sugiere que los pacientes con MM que presenten a estos genes disfuncionales podrían desarrollar amiloidosis durante el transcurso de la enfermedad.

### ✓ NR3C2

En el MM se han asociado altos niveles de potasio e hipertensión arterial con el avance de esta enfermedad. Si bien estas alteraciones pueden ser consecuencia del daño renal que sufren los pacientes con Mieloma Múltiple, el gen NR3C2 también podría estar implicado ya que, como receptor para mineralocorticoides, regula el volumen del líquido extracelular así como la concentración de iones. Debido a que este gen se encuentra afectado en su E-F con un alto valor global de afectación, su

disfunción podría estar ampliamente relacionada con el desarrollo e intensificación de la hipertensión arterial en estos pacientes.

#### ✓ **RBM9**

RBM9 actúa como un factor que inactiva a los receptores de estrógeno-alfa [65]. Se ha visto que las células de Mieloma Múltiple expresan receptores de estrógenos y la inhibición de estos receptores contribuye a la reducción de la proliferación celular y a la apoptosis [66]. Interesantemente, en nuestro estudio observamos que RBM9 está afectado en su estructura-función, sugiriendo cierta ventaja proliferativa a las células de mieloma.

#### ✓ **CUEDC2**

Este gen regula a la baja los niveles de proteínas ESR1 (receptor nuclear de estrógenos), involucradas en la proliferación y diferenciación celular a través de la vía del proteosoma-ubiquinasa. Como se mencionó anteriormente la expresión de ESR1 es esencial para la supervivencia de las células de MM [66]. Interesantemente, a través de nuestro análisis de estructura función, encontramos que CUEDC2 está afectado, sugiriendo que en los pacientes con afectación en el CUEDC2 podrían no estar regulando negativamente a ESR1 contribuyendo a la proliferación de las células de mieloma múltiple.

#### ✓ **KIF1A**

KIF1A es un transportador axonal de vesículas sinápticas. Se relaciona con desórdenes neurodegenerativos. Entre un 2 a 3% de los pacientes con mieloma desarrolla como síndrome paraneoplásico la neuropatía periférica sensitivo-motora. La causa auto inmune de este síndrome no está clara, sin embargo, se ha visto que KIF1A tiende a estar sub-expresado en pacientes con Mieloma Múltiple [67]. A partir de nuestro análisis de estructura-función, KIF1A resultó estar afectado por lo que su pérdida de función podría estar relacionada con el desarrollo de los eventos neurodegenerativos observados en pacientes con mieloma.

#### ✓ **HNF4A y NR2F2**

HNF4A Y NR2F2 son factores de transcripción que están involucrados en la regulación de la transcripción de las apolipoproteínas, un tipo de proteínas asociadas con la amiloidosis. A través de nuestro trabajo, pudimos observar que ambos genes son HUBs del interactoma y que HNF4A está

expresado en el 100% de los pacientes con MM y NR2F2 en el 86% de los pacientes, lo que probablemente podría involucrarlos en el desarrollo de amiloidosis en los pacientes con MM.

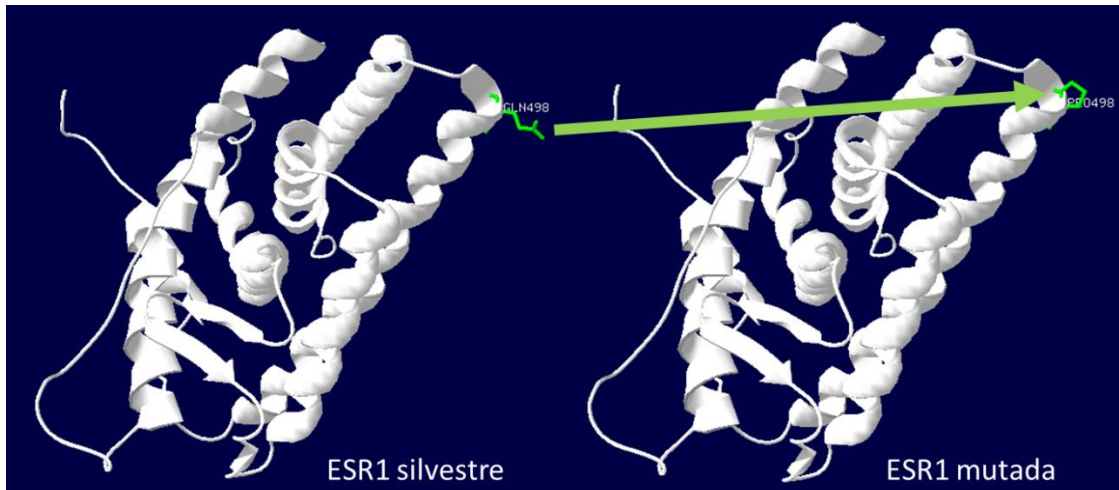
### **12.1.9. Genes candidato clínicamente accionables**

En esta sección, se profundizará en los genes candidato que son clínicamente accionables, es decir, son proteínas que pueden ser reguladas por diferentes drogas ya aprobadas por la FDA (“Food and Drug Administration”). Los genes ESR1, IDE y DNMT3B son clínicamente accionables, lo que abre la posibilidad de explorar nuevas alternativas para el tratamiento de Mieloma Múltiple y personalizarla según el gen blanco que esté afectado en cada paciente. A continuación se hará una descripción de cada uno de ellos.

#### **✓ ESR1 (Receptor de estrógenos)**

ESR1 es un receptor hormonal nuclear (específicamente, receptor de estrógenos), involucrado en la regulación génica en eucariotas y en la proliferación y diferenciación celular. Se ha visto que en diferentes tipos de cáncer los receptores de estrógeno suelen estar sobre-expresados [68], incluyendo a las células de mieloma [66]. Desde hace poco tiempo se ha considerado que la terapia endócrina podría ser efectiva para tratar Mieloma Múltiple a través del uso de moduladores selectivos del receptor de estrógeno (SERMs). Estos moduladores se comportan como antagonistas de los receptores de estrógenos en diferentes células cancerosas por lo que tienen propiedades antiproliferativas y proapoptóticas [69]. En dos estudios [70] [71], el tamoxifeno, modulador selectivo de receptores de estrógenos, generalmente usado en cáncer de mama, suprimió la síntesis de ADN, disminuyó la proliferación de células tumorales e indujo apoptosis en líneas celulares de MM. Otro SERM, el raloxifeno, utilizado para el tratamiento de la osteoporosis postmenopáusica, al tener una actividad similar a los estrógenos en huesos, ha sido probado en células de mieloma [72] encontrándose que disminuyó la viabilidad de dichas células, induciendo apoptosis. Además, el raloxifeno incrementó la respuesta citotóxica a la vincristina y al trióxido de arsénico (quimioterapéuticos en mieloma). También, el raloxifeno inhibió la actividad de la vía NFκB en las células de mieloma la cual está involucrada en la regulación de la proliferación y supervivencia celular. Por otro lado, en Mieloma Múltiple existe una actividad osteoclástica aumentada lo que provoca varias lesiones líticas. En este caso, raloxifeno también presenta la ventaja de tener la capacidad de inhibir la función osteoclástica, limitar la resorción ósea y favorecer las reparaciones de las lesiones líticas [72]. Interesantemente, a través de nuestro análisis, ESR1 resultó ser candidato debido a que es un HUB en el interactoma y además se encuentra expresado en todos los pacientes con MM (tabla 25), sugiriendo su posible rol

funcional en la proliferación celular de las células de mieloma. También observamos que la proteína ESR1 además de ser HUB, presenta una mutación de cambio de aminoácido (Q498P) que está estabilizando su estructura presentando un  $\Delta\Delta G = -1.07$  (figura 6) (ver ANEXO 1) lo que sugiere que la proteína no pierde su función sino al contrario podría estar mejorando la misma [53]. Por lo tanto, el raloxifeno y el tamoxifeno podrían ser medicamentos candidato para ser probados en células de mieloma y estudiar su posible efecto inhibitorio sobre ESR1.



**Figura 6. Estructura de la proteína ESR1.** Al mutar la glutamina (GLN) por prolina (PRO) en la posición 498 de ESR1, se estabilizó la proteína con un  $\Delta\Delta G = -1.07$ . Probablemente dicha estabilización se deba a que está cambiando un aminoácido polar (GLN) por uno no polar (PRO) lo que puede provocar que la estructura de ESR1 sufra un cambio conformacional y con ello la estabilización.

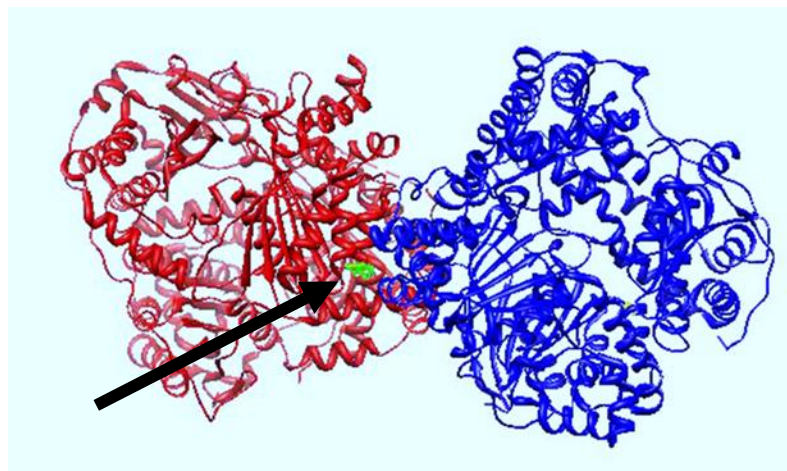
#### ✓ IDE (Enzima degradadora de insulina)

La enzima degradadora de insulina (IDE) es una metalopeptidasa de Zinc altamente conservada y distribuida en varios tejidos humanos, siendo particularmente abundante en cerebro, hígado y músculos [73] [74]. La actividad de IDE ha sido asociada con el catabolismo de la insulina y además participa en la regulación de las proteínas  $\beta$ -amiloides. Sin embargo, en la última década, se han realizado estudios que han vinculado a IDE con la etiología de algunas enfermedades, tales como la enfermedad de Alzheimer y Diabetes Mellitus tipo 2 [75] [76] [77], donde se ha asociado a esta proteína con diferentes procesos celulares tales como la señalización de los receptores de esteroides, y la modulación del proteosoma dependiente de insulina [78] [79]. Además, IDE se ha relacionado con múltiples funciones relacionadas al crecimiento celular [80]. Como se mencionó anteriormente, durante el curso de la enfermedad, los pacientes con Mieloma Múltiple pueden desarrollar amiloidosis, aunque la causa de la

misma es aún desconocida. El Mieloma Múltiple es un cáncer de células plasmáticas. Estas células son una parte importante del sistema inmunológico, responsables de la producción de anticuerpos, que son una de las defensas de primera línea contra la infección de que dispone el cuerpo. En el Mieloma Múltiple, las células plasmáticas producen un exceso de un tipo de anticuerpo anormal. La sobreproducción de anticuerpos anormales pone a estos pacientes en riesgo de padecer amiloidosis. Se calcula que entre el 10 por ciento y 15 por ciento de los pacientes con Mieloma Múltiple experimentarán síntomas causados por el desarrollo de amiloidosis durante el curso de su enfermedad. Sin embargo, hasta un 38 por ciento de los pacientes con mieloma pueden presentar amiloidosis pero no notar ninguno de sus síntomas. Por otro lado, se ha visto que IDE se encuentra sobre-expresado en células expuestas al estrés, sobre todo en tumores del sistema nervioso central y que su silenciamiento inhibe la proliferación del neuroblastoma [64]. Además, su silenciamiento se ha visto acompañado de una disminución de proteínas poliubiquitinadas por lo que recientemente se ha propuesto que IDE puede ser una proteína de choque térmico (proteínas que aumentan ante diferentes tipos de estrés celulares esenciales para evitar el daño celular asociado con la acumulación de proteínas desnaturalizadas) con implicaciones en la regulación del crecimiento celular y la progresión del cáncer, a través de la eliminación de las proteínas desnaturalizadas durante el estrés celular actuando sobre el sistema ubiquitina-proteosoma [64]. La vía ubiquitina-proteosoma juega un papel central en la destrucción de proteínas celulares, incluyendo proteínas reguladoras del ciclo celular. Debido a que esta vía es crítica para la proliferación y supervivencia de todas las células, particularmente de las células cancerosas, la inhibición del proteosoma es una probable y atractiva terapia anticáncer [81]. De manera interesante, a través de la secuenciación del genoma del Mieloma Múltiple [2], fueron encontradas mutaciones en 5 genes relacionados con la estabilización o eliminación de proteínas desnaturalizadas. Este descubrimiento fue de gran significancia clínica para Mieloma Múltiple ya que fue contemplado el uso de bortezomib (Velcade), medicamento ya aprobado el cual es un inhibidor del proteosoma, antiproliferativo y proapoptótico [82], el cual ha demostrado una actividad importante en Mieloma Múltiple en comparación con otros tipos de tumores [81]. El gen IDE hasta el momento no ha sido identificado como candidato en Mieloma Múltiple, pero de manera interesante, a través de nuestro estudio observamos que IDE es un gen candidato por ser un HUB del interactoma y por el análisis de estructura-función. Este gen lo encontramos expresado en todos los pacientes (tabla 25) lo que sugiere su posible participación en la proliferación y supervivencia de las células de mieloma a través de su actividad reguladora de la vía ubiquitina-proteosoma. Con base en esto, probablemente el bortezomib

podría ser un medicamento candidato para ser probado en células de mieloma donde el IDE esté expresado.

Por otro lado, a través de nuestro análisis de estructura-función encontramos que el IDE se encuentra afectado en su estructura por la mutación de cambio de aminoácido H134P la cual se encuentra justo en la interface de dimerización de la proteína y está enterrada en la estructura, como puede observarse en la figura 7 y que, además, presenta un valor global de afectación alto (5.47), siendo la estabilidad de la proteína una de las características mayormente afectadas ( $\Delta\Delta G = 6.78$ ). En este caso, probablemente los medicamentos mencionados no sirvan ya que el IDE se encuentra afectado y por lo tanto no es activo.



**Figura 7. Estructura de la proteína IDE.** Se puede observar que la mutación H134P se encuentra en la interface de dimerización (señalada con la flecha) sugiriendo que además de ser una mutación inestable para la proteína ( $\Delta\Delta G = 6.78$ ) podría además afectar el proceso de dimerización.

Además, mediante nuestro análisis encontramos que IDE es un blanco potencial de la bacitracina, antibiótico indicado para infecciones causadas por bacterias Gram positivas. Interesantemente, este medicamento también tiene la propiedad de ser un inhibidor de proteasas y otras enzimas [81-a], por lo que es un potencial candidato para ser probado en líneas celulares de mieloma. El efecto inhibitorio que tiene la bacitracina sobre el IDE de manera específica, sugiere que además de contribuir con la disminución de la proliferación de las células malignas, también podría disminuir el riesgo de desarrollar amiloidosis o sus complicaciones.

### ✓ DNMT3B

Este gen codifica para una metiltransferasa que participa en la metilación *de novo* de todo el genoma siendo esencial para el establecimiento de los patrones de metilación del ADN durante el desarrollo. Diversos estudios demuestran que genes supresores de tumor, son transcripcionalmente silenciados por hipermetilación, proceso propuesto como activador de la carcinogénesis [83]. En MM se han observado desregulaciones en las metiltransferasas [84]. A través de nuestro análisis de blancos farmacológicos encontramos que el DNMT3B puede ser regulado por la curcumina. La curcumina es un colorante natural procedente de la cúrcuma, especia obtenida del rizoma de la planta del mismo nombre cultivada principalmente en la India y utilizada desde la antigüedad para diversas aplicaciones. La cúrcuma es conocida en la medicina tradicional por sus funciones antitumoral, antioxidante, antiartritis, y antiinflamatorias. Los efectos antitumorales parecen derivar de la habilidad de la cúrcuma de inducir apoptosis en células cancerígenas, sin producir efectos citotóxicos en las células sanas [84-a]. La cúrcuma puede interferir con la actividad del factor de transcripción NF- $\kappa$ B, que ha sido ligado a través de múltiples estudios científicos, a un número de enfermedades inflamatorias, incluyendo cáncer y más recientemente fibrosis pulmonar [84-a]. En nuestro análisis encontramos que el gen DNMT3B está expresado en el 96% de los pacientes con MM, por lo tanto, sería interesante aplicar curcumina en líneas celulares de mieloma donde DNMT3B se encuentre expresado ya que se ha demostrado que la utilización de curcumina en pacientes con cáncer de mama logró bajar los niveles de DNMT3B [85].

Por otro lado, con nuestro análisis de estructura-función encontramos que DNMT3B está afectado. En este caso, la utilización de la curcumina no tendría efecto debido a que este gen no se encuentra funcional.

## 12.2. CAPITULO 2. Análisis de la expresión absoluta de genes

La finalidad de este capítulo es diseñar un método que nos permita determinar si los genes que se encuentran en el interactoma se encuentran o no expresados de manera absoluta. Nos interesa saber la expresión de los genes del interactoma dado que si el gen afectado por la mutación se está expresando, es más probable que la proteína mutada se encuentre en el citoplasma celular y entonces es mayormente posible que dicha proteína esté modificando la función celular de las interacciones proteína-proteína en donde ella participa. En caso contrario, si el gen afectado por la mutación no se expresa, lo más probable es que la proteína mutada no se encuentre en el citoplasma y por lo tanto el interactoma en el cual participa estaría siendo afectado debido a la no expresión de dicho gen. De esta manera, podremos identificar los módulos funcionales del interactoma que pudieran estar siendo alterados por las mutaciones y/o cambios en la expresión de los genes que los componen.

La integración del análisis de expresión absoluta con los análisis de interactomas nos permitirá tener una visión más integrada sobre las alteraciones biológicas que podrían estar contribuyendo en el desarrollo del Mieloma Múltiple.

### 12.2.1. Genes del cromosoma Y expresados en forma diferencial entre sexos

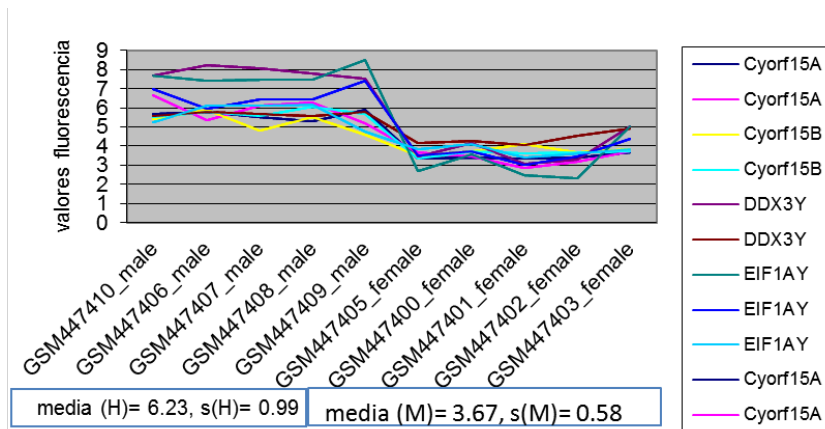
Como fue descrito previamente en materiales y métodos, primero identificamos cuáles son los genes que conforman el cromosoma Y a través del sitio bioinformático Genome Browser (tabla 23).

AK125149	CYorf15B	DQ595648	NLGN4Y	SFRS17A	TTY14	CDY2A
AL079291	CYorf16	DQ596138	P2RY8	SLC25A6	TTY2	CDY2B
AMELY	DAZ1	DQ596563	PCDH11Y	SRY	TTY3	DAZ1
ASMT	DAZ2	DQ597535	PLCXD1	TBL1X	TTY5	DAZ2
BC019893	DAZ3	DQ598589	PP1164	TBL1Y	TTY6	DAZ3
BC041884/64	DAZ4	EIF1AY	PPP2R3B	TGIF2LY	TTY7	DAZ4
BCORL2	DBY	FAM41AY1	PRKX	TMSB4Y	TTY9	DBY
BPY2	DDX3Y	GOLGA2LY1	PRY	TSPY1	USP9Y	DDX3Y
CD24	DHRX	GTPBP6	RBMY1A1	TSPY3	UTY	HSFY1
CD99	DKFZp666C074	HSFY1	RBMY1B	TTY15	VCY2	HSFY2
CDY1	DQ531594	HSFY2	RBMY1D	TTY16	VCY	KDM5D
CDY2A	DQ571366	ii-xr	RBMY1E	TTY17	VCY1B	PRY
CDY2B	DQ574760	IL3RA	RBMY1F	TTY21	VCY2	RBMY1A1
CR600952	DQ576041	JARID1C	RBMY1J	TTY22	XGPY2	RBMY1B
CRLF2	DQ577902	KDM5D	RBMY2EP	TTY23	XKRY	RBMY1D
CS706497	DQ579249	LOC401629	RBMY2FP	TTY6B	ZBED1	RBMY1E
CSF2RA	DQ585964	LOC728395	RBMY3AP	TTY6B	ZFY	RBMY1F
CSPG4LYP1	DQ592359	NCRNA00107	RPS4Y1	TTY10	BPY2	RBMY1J
CYorf15A	DQ593421	NCRNA00185	RPS4Y2	TTY13	CDY1	VCY
						VCY1B

**Tabla 23. Genes del cromosoma Y**

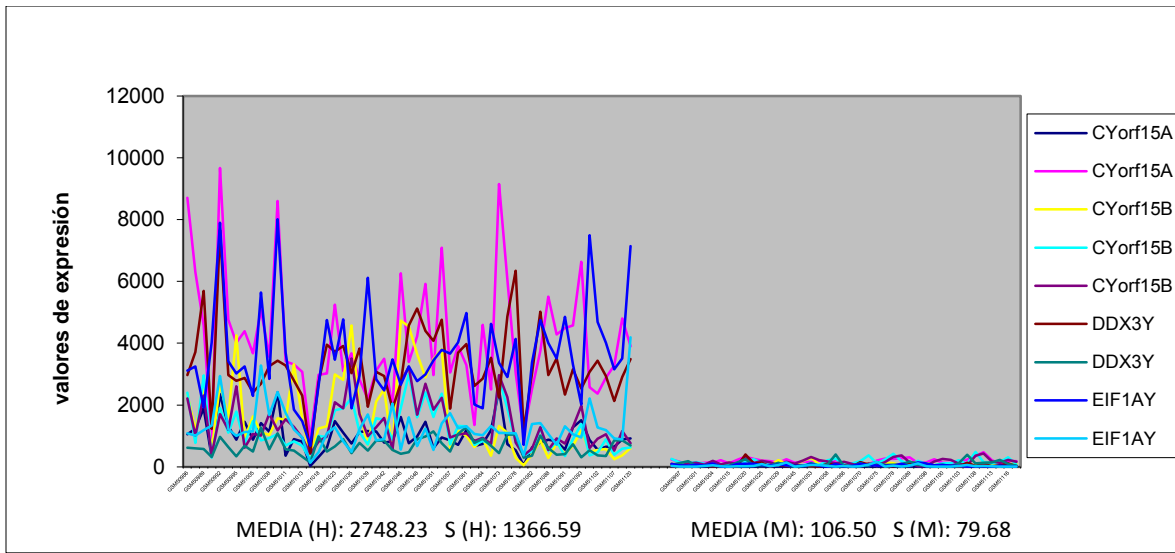


Para el desarrollo de nuestro método de expresión absoluta fue imprescindible determinar qué genes del cromosoma Y dan valores de fluorescencia significativamente bajos en la población femenina con respecto a la masculina, para tener la seguridad de que la fluorescencia que generen estos genes en los microarreglos de muestras de mujeres pueda ser considerada como ruido (debido a que las mujeres no tienen cromosoma Y). De esta forma, fue posible determinar un punto de corte de expresión absoluta para los diferentes microarreglos analizados (detallado posteriormente). Por lo tanto, primero identificamos qué genes presentaron valores de fluorescencia significativamente bajos en las mujeres con respecto a los hombres. Los valores de fluorescencia para esos genes fueron analizados inicialmente en un set de muestras de mucosa oral de 5 hombres y 5 mujeres (GSE17913), y en un set de 204 hombres y 210 mujeres con Mieloma Múltiple. Los genes del cromosoma Y que demostraron tener valores de fluorescencia significativamente altos en los hombres con respecto a las mujeres y que permitieron definir el punto de corte de expresión absoluta para definir a un gen como expresado son CYorf15A, CYorf15B, DDX3Y y EIF1AY (gráficos 9 y 10). En el set de muestras de mucosa oral (gráfico 9) pudimos observar para esos genes que la media de fluorescencia en hombres (media=6.23, n=5) fue mayor a la de mujeres (media=3.67, n=5) y corroboramos que la diferencia fue significativa al obtener una  $p=0.000019$ .



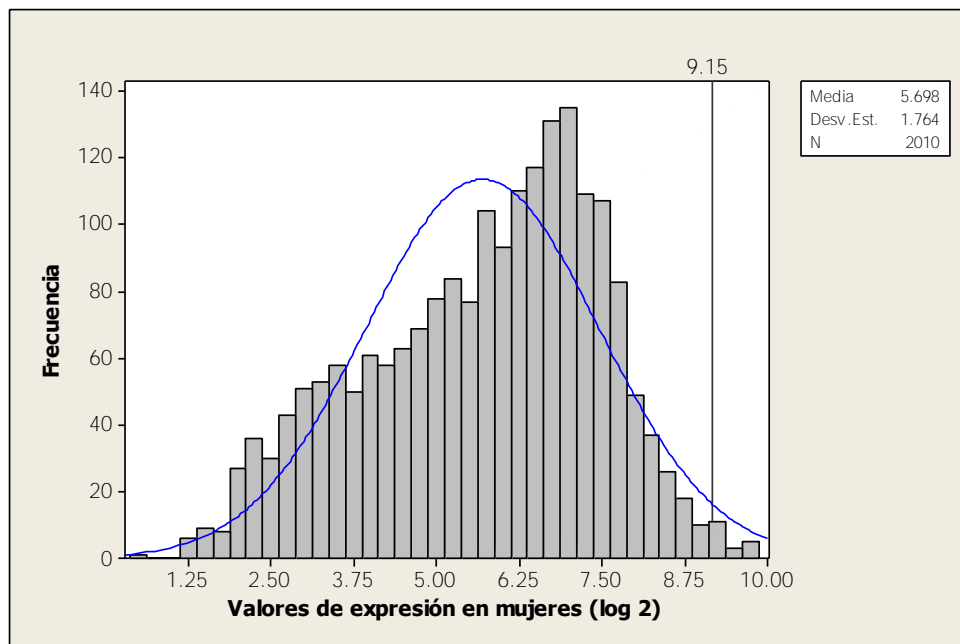
**Gráfico 9. Comparación de los valores de expresión de los genes del cromosoma Y entre sexos.** Comparación realizada en una muestra de mucosa oral de 5 hombres y 5 mujeres. **Media (H):** media de las fluorescencias para el set de hombres, **s(H):** desviación estándar de las fluorescencias para el set de hombres. **Media (M):** media de las fluorescencias para el set de mujeres, **s(M):** desviación estándar de las fluorescencias para el set de mujeres.

Posteriormente analizamos un set de individuos con Mieloma (GSE2658) (gráfico 10), donde la media de fluorescencia en hombres (media= 2748.23, n= 204) fue mayor que la media de fluorescencia en mujeres (media= 106.5, n=210). Corroboramos que la diferencia fue significativa con una  $p=0.002$ . Por lo tanto, podemos asumir que estos genes tienen una expresión significativamente diferente entre hombres y mujeres, característica que es imprescindible para el desarrollo del método para la determinación de expresión absoluta de genes.



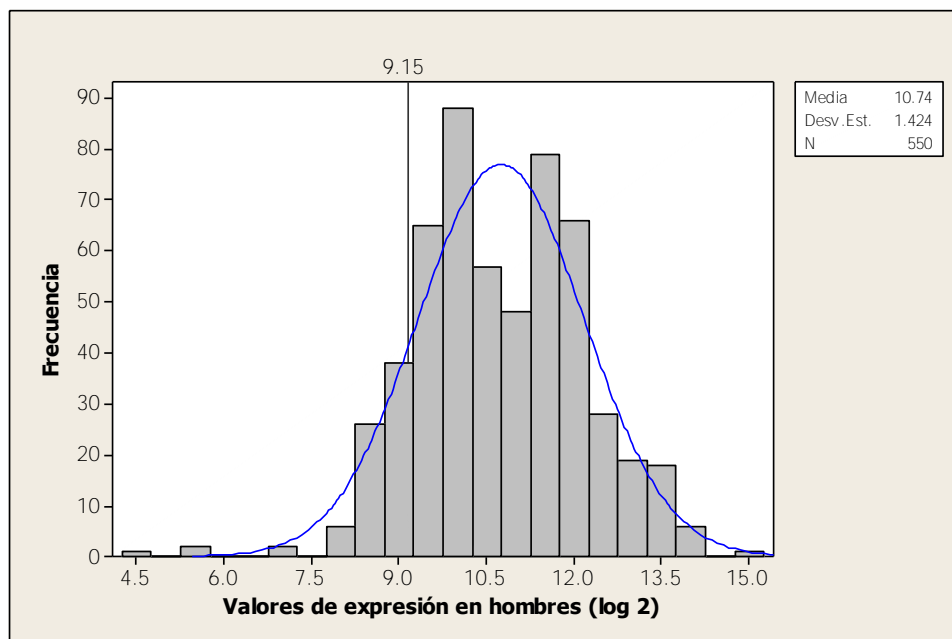
**Gráfico 10. Expresión de genes del cromosoma Y en hombres y mujeres con Mieloma Múltiple.** Comparación realizada en un set de 204 hombres y 210 mujeres con Mieloma Múltiple (GSE2658).

Una vez que ratificamos que las expresiones de los genes seleccionados del cromosoma Y fueron significativamente diferentes entre hombres y mujeres calculamos el punto de corte de expresión para el set de individuos con Mieloma Múltiple (GSE2658). La media de los valores de fluorescencia de estos genes en las mujeres con Mieloma Múltiple fue de 7.39 con una desviación estándar de  $\pm 1.76$  (gráfico 11). Por lo tanto, el 95% de las mujeres presentaron un valor máximo de fluorescencia de 9.15 para los 4 genes del cromosoma Y. A partir de esto, consideramos que los genes que tuvieran valores de fluorescencia mayores de 9.15 pueden considerarse como expresados, y los que están por debajo de ese valor no están siendo expresados.



**Gráfico 11. Histograma de la fluorescencia de los genes del cromosoma Y en microarreglos de mujeres con Mieloma Múltiple (GSE2658). Punto de corte= 9.15 (señalado en el gráfico).**

El 95% de la población femenina tiene fluorescencias menores a 9.15 para los 4 genes del cromosoma Y.

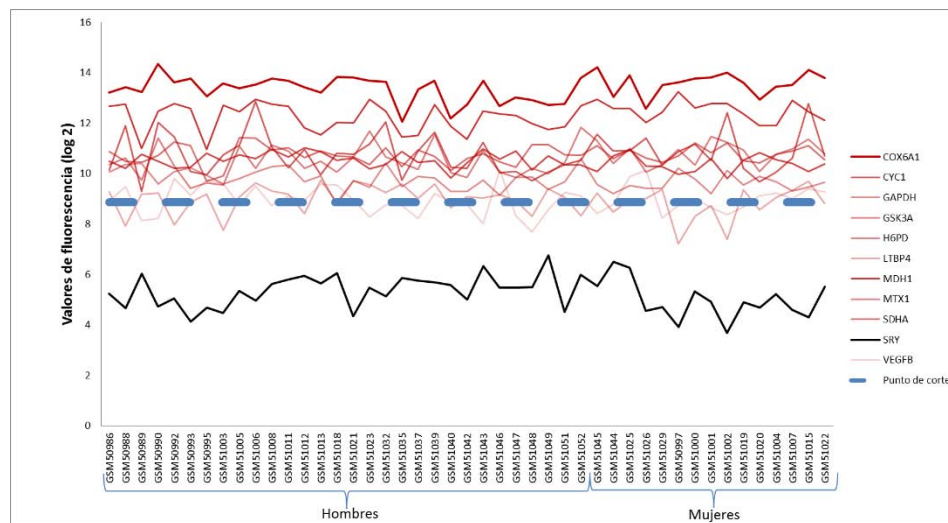


**Gráfico 12. Histograma de la fluorescencia de los genes del cromosoma Y en microarreglos de hombres con Mieloma Múltiple (GSE2658). Punto de corte= 9.15 (señalado en el gráfico).**

Dentro de la población masculina (media=10.74, DE=1.42) pudimos ver que la mayoría expresa los 4 genes seleccionados del cromosoma Y a partir del punto de corte establecido (gráfico 12).

### 12.2.2. Expresión absoluta de genes constitutivos “Housekeeping”

Posteriormente, analizamos la fluorescencia de los genes constitutivos o Housekeeping en base a nuestro punto de corte para la población con mieloma (9.15) (gráfico 13). La media de expresión en hombres fue de  $10.48 \pm 1.86$ , y la media de expresión en mujeres fue de  $8.34 \pm 3.23$ . Analizamos también, en base al punto de corte, la expresión del gen SRY donde la media de expresión en hombres fue de  $5.33 \pm 0.62$  y la media de expresión en mujeres fue de  $5.04 \pm 0.79$  (figura 5).

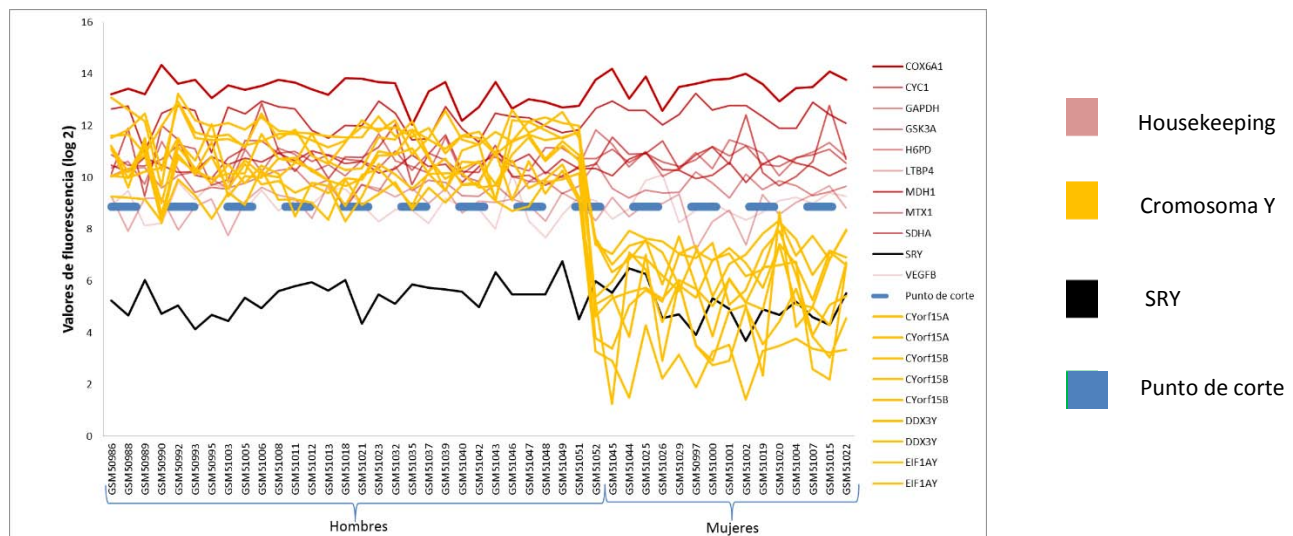


**Gráfico 13. Expresión de los genes constitutivos “Housekeeping” y del SRY según el punto de corte calculado.** Comparación realizada en hombres y mujeres con Mieloma Múltiple (GSE2658). El punto de corte es de 9.15 (señalado con la línea punteada azul).

En el gráfico 13 se puede observar que prácticamente todos los genes constitutivos o Housekeeping están expresados sin importar el sexo, aunque algunos tienen valores de fluorescencia que se encuentran en el límite del punto de corte de expresión, esto se puede explicar dado que los valores de fluorescencia varían drásticamente entre los genes constitutivos habiendo unos más expresados que otros [86]. Por ejemplo, el GAPDH, gen involucrado en el metabolismo de carbohidratos, es utilizado en varios estudios por ser uno de los genes constitutivos con mayor expresión en todas las células [86]. El gen VEGFB, factor de crecimiento endotelial vascular, es uno de los genes constitutivos que se encuentran al límite del punto de corte de expresión, es decir, con una baja expresión. Sin embargo,

dentro de su función, juega un papel menos importante en el sistema vascular que VEGFA (miembro de la misma familia), formando nuevos vasos sanguíneos, principalmente, durante algunos eventos patológicos. Esto podría explicar su bajo nivel de expresión. Se puede observar también, que los valores de fluorescencia de los genes constitutivos no varían entre sexos lo cual era de esperarse dado que la expresión de estos genes no es determinada por el sexo. Los valores de fluorescencia del gen SRY se encuentran debajo del punto de corte de expresión y tampoco varían entre hombres y mujeres. Este resultado es esperado ya que a pesar de que este gen pertenece al cromosoma Y, y es determinante del sexo, no debe estar expresado ni en hombres ni en mujeres en etapa adulta ya que solo se activa durante el desarrollo embrionario [60].

La comparación de la expresión de los genes constitutivos, del SRY y los genes del cromosoma Y en hombres y mujeres según nuestro punto de corte se puede observar en el gráfico 14.



**Gráfico 14. Expresión de genes del cromosoma Y, Housekeeping y SRY según el punto de corte.** Comparación realizada en hombres y mujeres con Mieloma Múltiple (GSE2658). El punto de corte es de 9.5 (señalado con la línea punteada azul).

En el gráfico anterior se pudo corroborar que el punto de corte de expresión puede diferenciar la expresión de los genes del cromosoma Y entre hombres y mujeres, la expresión de los genes constitutivos la cual debe ser independiente del sexo y la no expresión del gen SRY independientemente del sexo.

### 12.2.3. Expresión absoluta de genes de levadura

Por último, aplicamos la estrategia de punto de corte de expresión a las fluorescencias reportadas en el estudio donde se hibridó ARN de levadura en microarreglos para ARN humano (GSE22974). Para definir el valor del punto de corte de expresión, buscamos los valores de fluorescencia de las sondas del microarreglo correspondientes a los 4 genes del cromosoma Y seleccionados previamente para el método de expresión absoluta. A partir de los valores obtenidos, pudimos definir que el punto de corte de expresión fue de 9.17. Considerando que las sondas correspondientes a los 4 genes seleccionados del cromosoma Y pueden determinar el ruido de fluorescencia en las sondas de microarreglos humanos hibridados con ARN de levadura, se pudo determinar con el valor de fluorescencia obtenido para dichas sondas que de las 54644 sondas presentes en el microarreglo humano (HG133 Plus2), sólo 1178 hibridan con ARN de levadura (2.15%), lo que indicaría que estarían siendo expresadas (gráfico 15). Por lo tanto, el punto de corte aplicado indica que el 97.85% de las sondas del microarreglo humano hibridado con levadura no estarían expresadas lo cual es algo esperado dado que las sondas de microarreglos de humano no deben hibridar con ARN de levadura.

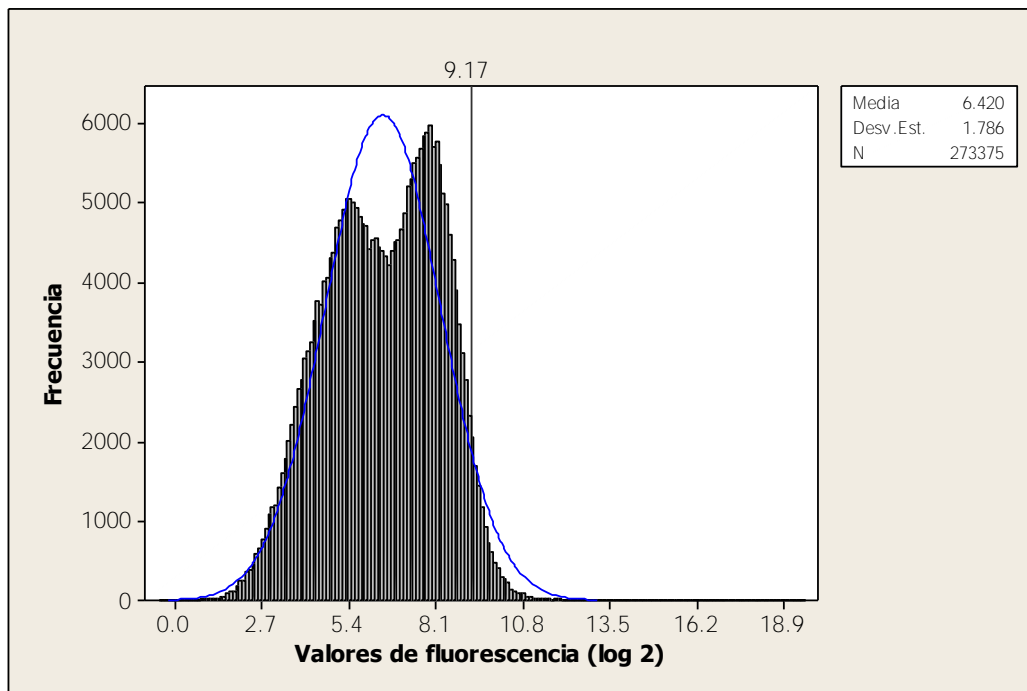
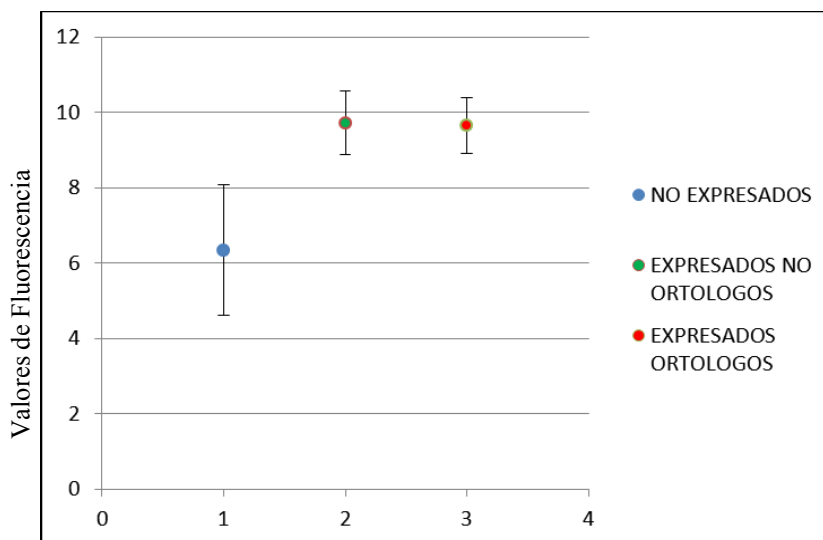


Gráfico 15. Determinación de la expresión de las sondas de microarreglo humano (HG133plus 2) en 5 muestras de levadura según el punto de corte. Punto de corte= 9.17 (señalado en la gráfica).

Número de sondas del microarreglo	54644	100%
Número de sondas no expresadas	53466	98%
Número de sondas expresadas	1178	2.16%

**Tabla 24. Número de sondas expresadas y no expresadas en el microarreglo humano hibridado con levadura.**

Como no era de esperarse que hayan sondas expresadas, atribuimos la expresión del 2.16% de las sondas a la existencia de genes ortólogos entre levadura y humano. Para verificar si esto es posible recurrimos a la base de datos de Affymetrix para la plataforma HGU133 plus2 y buscamos dentro del set de sondas expresadas cuáles representan a genes ortólogos entre humano y levadura. De las 1178 sondas expresadas observamos que 93 corresponden a genes ortólogos entre humano y levadura. Para verificar si la ortología contribuyó en la expresión de sondas humanas a partir de muestras de levadura, comparamos la media de la fluorescencia de los genes expresados ortólogos ( $9.65 \pm 0.73$ ) con la media de los genes expresados no ortólogos ( $9.71 \pm 0.84$ ) (gráfico 16).



**Gráfico 16. Comparación de las fluorescencias entre el set de sondas no expresadas, sondas expresadas ortólogas y sondas expresadas no ortólogas.**

No observamos diferencia significativa entre ambos sets, por lo tanto, no podemos atribuir la expresión de las sondas a la homología de estos genes entre humano y levadura. Es probable que estas sondas tengan un ruido de fondo muy alto y por ello presenten fluorescencias tan elevadas y en consecuencia pasen nuestro punto de corte.

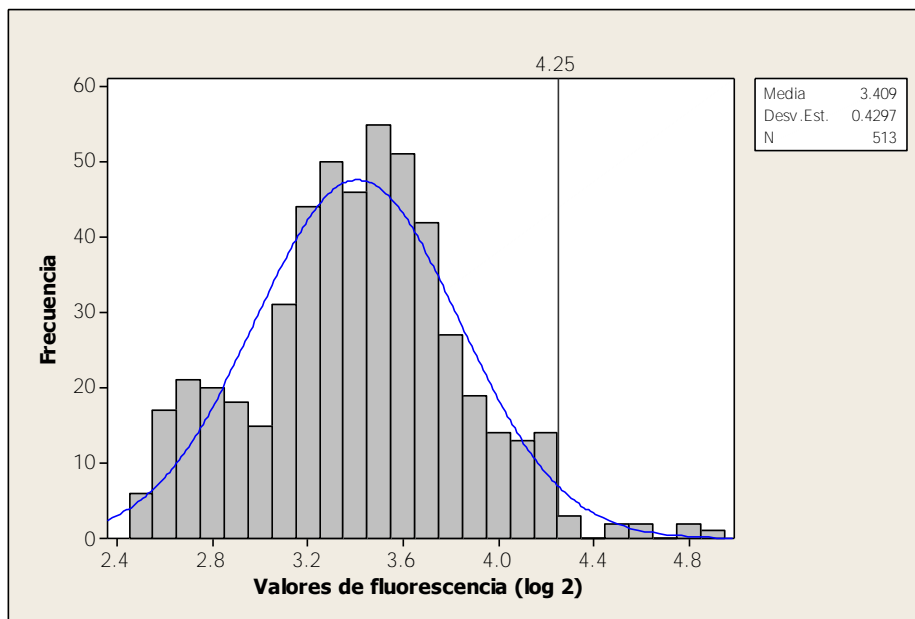
Con la finalidad de analizar si la expresión de las sondas se debe a que presentan un ruido de fondo muy alto, analizamos la fluorescencia de las sondas de los mismos microarreglos hibridados con ARN de levadura a través del programa “The Gene Expression Barcode 2.0” [87] [88]. Este programa estima medidas absolutas de expresión de genes a través de las fluorescencias ya reportadas de sondas de microarreglos de humano y de ratón. Por lo tanto, decidimos evaluar cuáles y cuántas sondas no se están expresando según Barcode y comparamos sus resultados con los nuestros. Pudimos observar que el 94% (n=51196) de las sondas humanas no hibridan con ARN de levadura según Barcode a diferencia de lo reportado por nuestro método de corte en donde el 98% (n=53466) de las sondas no se expresan. Por otro lado, de las 1178 sondas expresadas según nuestro método, 617 también están expresadas según Barcode. Esto confirma que es probable que las 617 sondas presenten un ruido de fondo alto ya que sus expresiones se vieron reflejadas de igual manera por ambos métodos. A partir de ello, pudimos ver que nuestro método reporta resultados muy similares a los del Gene Expression Barcode, pero con una mayor especificidad dado que nuestro método reporta menor cantidad de sondas humanas expresadas con ARN de levadura.

#### **12.2.4. Expresión absoluta de los genes candidato**

Una vez corroborada la estrategia de expresión absoluta, ésta fue aplicada a los 255 pacientes con Mieloma Múltiple del MMRC cuyo genoma fue secuenciado y que cuentan con microarreglos de expresión (GSE26760).

El cálculo del punto de corte sobre los microarreglos de las mujeres dio un valor máximo de fluorescencia de 4.25 con un 95% de confianza (gráfico 17). Por lo tanto, consideramos como gen expresado todo aquel que tenga más de 4.25 de valor de fluorescencia en este set de pacientes.





**Gráfico 17. Cálculo del punto de corte a partir de los microarreglos de mujeres con Mieloma Múltiple del MMRC (GSE26760).** El punto de corte es de 4.25 (señalado en el gráfico). Cálculo del punto de corte= media + 2.96 x Desv. Est.

A través del punto de corte pudimos asociar a los genes candidato con su estado de expresión en la población en general de pacientes con Mieloma Múltiple del MMRC. En la tabla 25 se presentan los resultados obtenidos.

Gen	Condición de candidato	Nº de pacientes que lo expresa	% de pacientes que lo expresa	Promedio de expresión poblacional (fluorescencia)
ADAM17	E-F y HUB	255	100%	6.33
NCOR1	E-F y HUB	255	100%	6.77
THRB	E-F y HUB	255	100%	6.86
CHD3	E-F y HUB	255	100%	5.86
DNMT3B	E-F	245	96%	4.97
RBM9	E-F	255	100%	7.76
ZBTB33	E-F	255	100%	7.85
IDE	E-F	255	100%	7.91
NR3C2	E-F	56	22%	3.94
CUEDC2	E-F	255	100%	8.47
KIF1A	E-F	255	100%	5.22
RIMS2	E-F	255	100%	6.03
HNF4A	HUB	255	100%	5.83
ESR1	HUB	255	100%	6.11
USP7	HUB	255	100%	9.04
NR2F2	HUB	220	86%	4.66
GIT1	HUB	255	100%	6.93
MDC1	HUB	255	100%	6.05

**Tabla 25. Expresión de los genes candidato.** En total 255 pacientes con Mieloma Múltiple del MMRC tienen microarreglos de expresión (están incluidos los 38 pacientes secuenciados) a través de los cuales pudimos obtener la información de expresión de los genes candidato. Los datos de fluorescencia se presentan en log2. El punto de corte de expresión absoluta fue de 4.25.

Observamos que casi todos los genes candidato tienden a estar expresados en el 100% de la población de pacientes con Mieloma Múltiple. Los genes candidato que se encontraron afectados en su E-F y que además se encontraron expresados probablemente estén contribuyendo con el fenotipo de la enfermedad, dado que la proteína alterada podría estar presente en el citoplasma celular. Solo el gen NR3C2 fue el gen que presentó una tendencia a no expresarse en los pacientes con mieloma aunque estuvo afectado en su E-F. Probablemente, este gen solo contribuya al fenotipo de la enfermedad en los pocos pacientes que sí lo expresan. El gen USP7, tiene un valor de expresión alto. Probablemente esté sobre-expresado por lo que al ser HUB, podría estar afectando la conectividad entre los genes vecinos del interactoma y por lo tanto, la función en la que se encuentran involucrados.

### 13. DISCUSIÓN

A partir de los análisis integrados de genómica, transcriptómica, interactómica y de estructura-función sobre las proteínas mutadas en Mieloma Múltiple que tienen la característica de presentar motivos de unión a Zinc hemos podido descubrir genes candidato nuevos y otros ya reportados tanto en mieloma como en otros cánceres. A través del análisis de las proteínas que unen Zinc encontramos que éstas generan redes de interacción proteína-proteína tanto biológica como estadísticamente significativas en esta enfermedad a diferencia del análisis de las proteínas mutadas que unen otros elementos esenciales. Interesantemente, al buscar la relación que existe entre el Zinc y el cáncer pudimos encontrar en la literatura que el Zinc es el segundo metal más abundante en el cuerpo humano y que constituye un componente esencial en la estructura de diversas proteínas que están involucradas en procesos de crecimiento celular, síntesis de ADN, transcripción, división, desarrollo y diferenciación celular [89]. Estos procesos biológicos contienen funciones moleculares que normalmente suelen ser alteradas o afectadas por el cáncer [90]. Debido a que las perturbaciones en el funcionamiento de una red normal pueden desencadenar una cascada de fallas las cuales podrían conducir al desarrollo de la enfermedad [11], es decir, pueden conducir a una alteración del funcionamiento normal de algunas redes celulares y, como consecuencia, al posible desarrollo de enfermedades complejas como el cáncer [7] [25] [17] [11] [18], el estudio de las mutaciones y expresiones de proteínas que unen Zinc involucradas en redes de interacción podrían ayudar a determinar los procesos biológicos y patogénicos que alteran el funcionamiento celular y que transforman a una célula normal en una célula cancerosa, para finalmente, encontrar posibles genes candidato en esta enfermedad.

En los últimos decenios, se han realizado varios estudios para investigar la participación de micronutrientes en los mecanismos antioxidantes y anticarcinogénesis que han sido implicados en el desarrollo del cáncer. En particular, el Zinc, como elemento traza, ha suscitado un gran interés por parte de la mayoría de los investigadores por su asociación en los procesos bioquímicos y de defensa antioxidante [91] [92] [93], jugando un rol preventivo contra la producción de radicales libres y el crecimiento maligno protegiendo a la célula del estrés oxidativo [94]. Por otro lado, en estudios más recientes, se ha demostrado que el Zinc puede tener un papel importante en la inmunidad mediada por células [95]. Por lo tanto, cabe la pregunta sobre si los niveles de Zinc alterados son uno de los causantes del cáncer o son consecuencia de dicho proceso patológico. Algunos estudios han demostrado que su deficiencia a través de modelos animales está involucrada en la transformación maligna de las células [96]. La deficiencia de Zinc ha sido asociada con daño y cambios oxidativos en el ADN que podrían incrementar el riesgo de cáncer. En algunos tumores, incluyendo los de pulmón,

riñón e hígado, se han reportado cantidades alteradas de metales [97] [98] [99]. En particular, el metabolismo del Zinc se ha visto afectado en el cáncer de próstata, mama y de páncreas [100] [101] [102] En el estudio realizado por Dar et al. [103], la deficiencia o exceso de los elementos traza estuvieron implicados en el desarrollo de algunos cánceres a través del cual fue reportado un elevado nivel de cobre y bajos niveles de Zinc en el plasma de pacientes con cáncer de esófago. En otros estudios, bajos niveles plasmáticos de Zinc han sido encontrados en pacientes con algún tipo de neoplasia en comparación con individuos sanos [104] [105] [106]. Por ejemplo, Kuo et al. [107] reportó niveles séricos de Zinc significativamente bajos en pacientes con cáncer de mama comparados con un grupo control y sugirió que los niveles plasmáticos de Zinc podrían ser usados como un posible marcador pronóstico y terapéutico en esta enfermedad, coincidiendo con la opinión de otros investigadores en cáncer [108] [109] [110] [111]. La asociación entre los niveles plasmáticos de Zinc y el riesgo de desarrollo de cáncer también ha sido evaluado por Gupta et al. [112] y más recientemente por Adzersen et al. [113]. Estos estudios demostraron la existencia de asociaciones inversas y significativas entre los niveles de Zinc en plasma y una dieta rica en Zinc con respecto al riesgo de desarrollo de cáncer de mama que podrían ser importantes para el desarrollo de estrategias de prevención de la enfermedad. Con respecto a esto, el estudio de Fresno et al. [114] reportó que la administración oral de altas dosis de Zinc tuvo efectos antimetastásicos en un modelo de tumor experimental. Sin embargo, los mecanismos por los cuales el Zinc contribuye a la carcinogénesis todavía son controversiales. Cuando los tumores están presentes, pueden ocurrir alteraciones en la distribución del Zinc, comprometiendo su función. Una reducción de los niveles de Zinc en plasma puede ser resultado de un incremento en la expresión de genes codificadores de proteínas transportadoras de Zinc que promueven el transporte de este mineral de la sangre al tejido tumoral. Kagara et al. [115], Taylor [116] y Louis and Cousins [117] demostraron un incremento de los niveles de Zinc dentro del tejido tumoral de pacientes con cáncer de mama que fue asociado con un incremento de la expresión de las proteínas transportadoras Zip 10, Zip 7 y Zip 6. Estos investigadores atribuyeron la reducción de los niveles de Zinc plasmáticos a la sobre-expresión de los genes codificadores de proteínas transportadoras de Zinc que pueden transferir este mineral del plasma hacia el interior del tumor. Por otro lado, el estudio de las metalotioneínas (MTs) y su asociación con el Zinc, han tomado relevancia en el estudio de diferentes cánceres ya que dentro de sus funciones se encuentran el mantenimiento de la homeostasis de los elementos traza y la eliminación de radicales libres [118]. A través de la unión y el desprendimiento de Zinc, las MTs pueden regular los niveles de Zinc de una célula a otra, por lo que son componentes claves del sistema de señales de Zinc en las células. En base

a esto, en el estudio de Kawata et al. [96], los niveles de metalotioneínas y Zinc se encontraron bajos en el carcinoma hepatocelular. De acuerdo con estos datos, la disminución de las concentraciones plasmáticas de Zinc fueron observadas después de episodios de estrés, trauma y en varias enfermedades malignas [119] [112]. Hasta el momento, no hay datos que reporten niveles de Zinc plasmáticos en individuos con Mieloma Múltiple, sin embargo, los estudios mencionados podrían sugerir que la detección de bajos niveles de Zinc plasmáticos podría considerarse como un factor de riesgo para cáncer y al mismo tiempo funcionar como un posible marcador para prevenir la carcinogénesis a través de suplementos de este mineral. Por otro lado, la disminución de los niveles de Zinc podrían estar causando una deficiencia en la función de las proteínas que dependen de él afectando las importantes funciones celulares en las que éstas se encuentran involucradas, como la activación y unión de ciertos factores de transcripción. Al mismo tiempo, si las metalotioneínas no tienen los niveles de Zinc suficientes, no podrán contribuir en la antioxidación celular necesaria para prevenir la conversión de una célula normal a una cancerosa. Esto, no solo aclara la importancia del estudio de los niveles de Zinc en cáncer, sino también da sustento a que la búsqueda de genes candidato a partir del análisis de las proteínas que tienen motivos de unión a Zinc podría ser una estrategia certera dentro de los estudios asociados a la investigación del cáncer, tal como lo hicimos en nuestro trabajo.

Interesantemente, el grupo de Won et al. [40] desarrolló una aproximación integradora muy similar a la nuestra para la búsqueda de genes candidato en cáncer, basada en la integración de los análisis de función de las proteínas, de expresión, y de redes de interacción proteína-proteína. Para el análisis de función de las proteínas, ellos seleccionaron miles de genes asociados a cáncer de mama con mutaciones de cambio de aminoácido, y utilizaron solo 3 predictores basados en conservación de secuencia para ver si la proteína estaba afectada. Para integrar la información de los tres predictores, utilizaron un método bayesiano para estimar qué tan probable es que la proteína esté afectada por la mutación. Sin embargo, nosotros integramos la información obtenida de una mayor cantidad de predictores basados en conservación de secuencia (cinco) y, además, analizamos la estabilidad de la estructura de las proteínas mutadas. Por lo tanto, nuestra estrategia para la predicción de la funcionalidad de las proteínas mutadas integró una mayor cantidad de herramientas y permitió que el análisis fuera más detallado, otorgando una mayor confianza en nuestros resultados de predicción. En lo que refiere a los análisis de expresión y a la red de interacción, ellos buscaron los perfiles de expresión del tejido mamario para generar solo con los genes expresados una red de interacción proteína-proteína a la cual le fue integrada los resultados de predicción de las mutaciones de las proteínas expresadas en mama, para finalmente seleccionar como candidato a los genes que estaban

afectados según el método bayesiano y expresados en el interactoma. A diferencia de su estudio, nosotros primero seleccionamos las proteínas mutadas en Mieloma Múltiple que consideramos importantes para el desarrollo del cáncer y que aportan la novedad por estar relacionadas al Zn. A partir de ellas, construimos el interactoma donde no solo analizamos de una manera más integrada la estructura-función de las proteínas mutadas, sino también analizamos el interactoma determinando cuáles son los genes que son el centro de muchas conexiones entre proteínas y que están mutados y/o afectados. Por otro lado, nuestra selección de genes candidato no estuvo basada en su expresión, sino que estuvo basada en el análisis de estructura-función y de interactoma. El análisis de expresión lo utilizamos para conocer cuales genes candidato estarían expresándose en los pacientes secuenciados pero además también estudiamos la expresión de los genes candidato en una muestra extra de 217 pacientes no secuenciados. Además, la información de expresión que analizamos, proviene de los mismos pacientes a los que les fue secuenciado su genoma, por lo tanto, nuestra estrategia otorga la posibilidad de obtener los genes candidato por paciente contribuyendo a una medicina P4.

## 14. CONCLUSIONES

En este trabajo nosotros demostramos que tanto el Zinc como elemento esencial y las proteínas que lo unen, son participantes activos e importantes para el desarrollo del Mieloma Múltiple. También pudimos determinar que el interactoma que forman estas proteínas en esta enfermedad involucra vías bioquímicas y de señalización estadísticamente significativas y que están relacionadas con la regulación del sistema inmune, con la regulación de la transcripción y del ciclo celular, con la apoptosis, la proliferación celular, con las vías NF-kB y TGF-beta entre otras, las cuales encontramos alteradas a través de los análisis de estructura-función de las proteínas involucradas. Interessantemente, muchas de estas vías ya han sido previamente reportadas por la secuenciación de los genomas de pacientes con Mieloma Múltiple, lo que le da mayor confianza a las nuevas vías que hemos encontrado. Con la integración del análisis de estructura-función y de expresión absoluta hemos podido sugerir cuáles proteínas del interactoma se encuentran afectadas para poder determinar las interacciones proteína-proteína que pudieran estar alteradas. Interessantemente, 26 pacientes tienen mutaciones deletéreas en el interactoma del Zinc, por lo tanto, a partir de este análisis integrativo, podría ser posible explicar qué vías están afectadas en el 70% de los pacientes cuyos genomas fueron secuenciados. A partir de la integración de estos análisis descubrimos 28 genes candidato de los cuales 18 todavía no han sido reportados como tales en Mieloma Múltiple. Interessantemente, dentro de nuestros diez genes candidato restante, 5 genes ya han sido reportados como tales en otros tipos de cáncer y otros 5 en Mieloma Múltiple. Esto nos brinda la confianza necesaria con respecto a la estrategia que estamos empleando para detectar nuevos genes candidato. A través de nuestros análisis, pudimos notar que los genes ADAM17, NCOR1, THRB y CHD3 fueron los que cumplieron la mayor cantidad de criterios para ser candidato ya que mostraron estar afectados en su estructura-función y ser HUBs del interactoma al mismo tiempo. Además, resultaron ser esenciales para la función celular en el humano, estar involucrados en enfermedades monogénicas y alterar el fenotipo de ratón. Esto sugiere la importancia de ser estudiados con mayor profundidad en los pacientes con MM.

Al analizar los valores globales de afectación que presentan nuestros genes candidato pudimos observar que tienden a tener valores de afectación altos, y que ésta característica también la comparten con genes candidato que fueron obtenidos para otros tipos de cáncer a través de los análisis funcionales de las proteínas con mutaciones de cambio de aminoácido [19]. Esto podría sugerir que el valor global de afectación permite priorizar a las variantes funcionales de acuerdo a sus efectos deletéreos sobre la estructura-función de una proteína para diferenciar a las proteínas candidato en MM de las que no lo son, y que nuestra estrategia para la selección de genes candidato es confiable.

Al integrar el análisis de expresión absoluta a los genes del interactoma, fue posible observar que casi todos los pacientes con mieloma a los cuales les fue analizado su transcriptoma, están expresando casi todos los genes candidato afectados en su estructura-función y, también, a los genes que son HUBs. Esto nos permite concluir que probablemente la proteína afectada está presente en la célula, afectando al módulo del interactoma al que pertenece y, por lo tanto, promoviendo el desarrollo de la enfermedad. Al integrar la información de genes candidato y del interactoma, es posible ver cómo cada paciente tiene diferentes genes candidato afectados pero la vía afectada o módulo del interactoma es la misma. Por lo tanto, este tipo de estrategia de análisis podría contribuir con el desarrollo de una medicina P4 (personalizada, predictiva, preventiva y participatoria) donde sería posible aplicar un tratamiento enfocado directamente al gen o genes candidato afectados en cada paciente. Interesantemente, los nuevos genes candidato ESR1, IDE y DNMT3B son clínicamente accionables, es decir, pueden ser regulados por medicamentos ya aprobados en la actualidad, sugiriendo que tanto el raloxifeno como el tamoxifeno podrían ser medicamentos candidato para ser probados en células de mieloma y estudiar su posible efecto inhibitorio sobre ESR1. Además, encontramos que IDE es un blanco potencial de la bacitracina, utilizado hasta el momento como antibiótico, por lo que también es un potencial candidato para ser probado en células de mieloma. Por último, encontramos que la curcumina disminuye los niveles de DNMT3B por lo que sería interesante probarla en líneas celulares de mieloma donde DNMT3B se encuentre sobre-expresado. Esto abre la posibilidad de explorar nuevas alternativas terapéuticas para Mieloma Múltiple.



## 15. PERSPECTIVAS

Este estudio nos ha dejado varias puertas abiertas por explorar. Por un lado, el hecho de que existan genes candidato que son clínicamente accionables abre la oportunidad de explorar si las líneas celulares de mieloma responden a los diferentes medicamentos que se han mencionado en esta tesis. Por otro lado, sería interesante poder aplicar las estrategias de integración de ómicas y de estructura-función que empleamos para descubrir genes candidato en Mieloma Múltiple, en otros tipos de cánceres y analizar si también es posible llegar al descubrimiento de nuevos genes candidato a partir del análisis de las proteínas que unen Zinc. Con este tipo de estudio, no solo pudimos demostrar que es posible determinar algunos de los factores que están contribuyendo con el desarrollo del Mieloma Múltiple a partir del análisis de las proteínas que unen Zinc, sino también, fue posible observar que a través de la utilización y aplicación de los diferentes análisis bioinformáticos es posible descubrir los procesos biológicos que están siendo afectados desde una perspectiva de sistemas. Esto nos permite pensar que a futuro podríamos reducir la cantidad de genes a secuenciar en los pacientes con Mieloma Múltiple y tal vez en otros tipos de cáncer, y limitarnos a solo secuenciar los genes que están asociados al Zinc, estrategia que nos permitiría bajar los costos de secuenciación del genoma completo, y facilitar el análisis y entendimiento de la gran cantidad de datos que suelen ser obtenidos por las tecnologías ómicas. Finalmente, sería importante comprobar de manera experimental que los genes candidato mencionados estén realmente siendo afectados en un set de pacientes con Mieloma Múltiple, y al mismo tiempo, analizar si los valores plasmáticos del Zinc se encuentran disminuidos en esta enfermedad. Estos análisis también nos permitirían evaluar si el Zinc puede ser utilizado como un marcador tanto preventivo como predictivo de la enfermedad. De esta manera, la contribución de esta estrategia de análisis hacia una medicina personalizada podría ser evaluada y considerada en los pacientes con Mieloma Múltiple.

## 16. BIBLIOGRAFÍA

- 1-a. *Guía de Práctica Clínica. Diagnóstico y Tratamiento de Mieloma Múltiple. México: Secretaría de Salud, 2010.*
- 2-a. *Labardini et al, Oncoguía. Mieloma Múltiple. Instituto Nacional de Cancerología. México DF. Cancerología 6 (2011): 121 – 124.*
1. Georgi, B., B.F. Voight, and M. Bucan, *From mouse to human: evolutionary genomics analysis of human orthologs of essential genes.* PLoS Genet. 9(5): p. e1003484.
2. Chapman, M.A., et al., *Initial genome sequencing and analysis of multiple myeloma.* Nature. 471(7339): p. 467-72.
3. Braggio, E. and R. Fonseca, *Unraveling the multiple myeloma genome in the next-generation sequencing era: challenges to translating knowledge into the clinic.* Expert Rev Hematol. 4(6): p. 579-81.
4. Paul, S. and E. Regulier, *[Molecular basis of oncogenesis].* Ann Biol Clin (Paris), 2001. 59(4): p. 393-402.
5. Sejpal, S.V., A. Bhate, and W. Small, *Palliative radiation therapy in the management of brain metastases, spinal cord compression, and bone metastases.* Semin Intervent Radiol, 2007. 24(4): p. 363-74.
6. Barnes, E.A., J.L. Palmer, and E. Bruera, *Prevalence of symptom control and palliative care abstracts presented at the Annual Meeting of the American Society for Therapeutic Radiology and Oncology.* Int J Radiat Oncol Biol Phys, 2002. 54(1): p. 211-4.
7. Tian, Q., N.D. Price, and L. Hood, *Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine.* J Intern Med. 271(2): p. 111-21.
8. Ostrowski, J. and L.S. Wyrwicz, *Integrating genomics, proteomics and bioinformatics in translational studies of molecular medicine.* Expert Rev Mol Diagn, 2009. 9(6): p. 623-30.
9. Meyerson, M., S. Gabriel, and G. Getz, *Advances in understanding cancer genomes through second-generation sequencing.* Nat Rev Genet. 11(10): p. 685-96.
10. Berger, M.F., et al., *Integrative analysis of the melanoma transcriptome.* Genome Res. 20(4): p. 413-27.
11. del Sol, A., et al., *Diseases as network perturbations.* Curr Opin Biotechnol. 21(4): p. 566-71.
12. Edelman, L.B., J.A. Eddy, and N.D. Price, *In silico models of cancer.* Wiley Interdiscip Rev Syst Biol Med. 2(4): p. 438-59.
13. Elnitski, L.L., et al., *The ENCODEdb portal: simplified access to ENCODE Consortium data.* Genome Res, 2007. 17(6): p. 954-9.
14. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes.* Nature, 2007. 446(7132): p. 153-8.
15. Kaminker, J.S., et al., *CanPredict: a computational tool for predicting cancer-associated missense mutations.* Nucleic Acids Res, 2007. 35(Web Server issue): p. W595-8.
16. Leary, R.J., et al., *Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers.* Proc Natl Acad Sci U S A, 2008. 105(42): p. 16224-9.
17. Vandin, F., E. Upfal, and B.J. Raphael, *De novo discovery of mutated driver pathways in cancer.* Genome Res. 22(2): p. 375-85.
18. Cerami, E., et al., *Automated network analysis identifies core pathways in glioblastoma.* PLoS One. 5(2): p. e8918.
19. Shi, Z. and J. Moulton, *Structural and functional impact of cancer-related missense somatic mutations.* J Mol Biol. 413(2): p. 495-512.
20. Aloy, P. and R.B. Russell, *Structural systems biology: modelling protein interactions.* Nat Rev Mol Cell Biol, 2006. 7(3): p. 188-97.
21. Furlong, L.I., *Human diseases through the lens of network biology.* Trends Genet. 29(3): p. 150-9.
22. Snyderman, R. and J. Langheier, *Prospective health care: the second transformation of medicine.* Genome Biol, 2006. 7(2): p. 104.
23. Auffray, C., Z. Chen, and L. Hood, *Systems medicine: the future of medical genomics and healthcare.* Genome Med, 2009. 1(1): p. 2.
24. Hood, L. and M. Flores, *A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory.* N Biotechnol. 29(6): p. 613-24.
25. Nicholson, J.K., *Global systems biology, personalized medicine and molecular epidemiology.* Mol Syst Biol, 2006. 2: p. 52.
- 25-a. Yannone, S.M., *Metals in Biology: Defining Metalloproteomes,* Curr Opin Biotechnol. Feb 2012; 23(1):89-95.

- 25-b. Lu Y. *Metalloprotein and metallo-DNA/RNAzyme design: current approaches, success measures, and future challenges*. Inorg. Chem., 2006, 45 (25), pp 9930–9940.
26. Carter, H., et al., *Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations*. Cancer Res, 2009. 69(16): p. 6660-7.
27. Torkamani, A. and N.J. Schork, *Identification of rare cancer driver mutations by network reconstruction*. Genome Res, 2009. 19(9): p. 1570-8.
28. Reimand, J. and G.D. Bader, *Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers*. Mol Syst Biol. 9: p. 637.
29. Hodis, E., et al., *A landscape of driver mutations in melanoma*. Cell. 150(2): p. 251-63.
30. Chen, W., et al., *Targets of genome copy number reduction in primary breast cancers identified by integrative genomics*. Genes Chromosomes Cancer, 2007. 46(3): p. 288-301.
31. Zhou, D., et al., *Exome capture sequencing of adenoma reveals genetic alterations in multiple cellular pathways at the early stage of colorectal tumorigenesis*. PLoS One, 2013. 8(1): p. e53310.
32. Lohr, J.G., et al., *Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing*. Proc Natl Acad Sci U S A. 109(10): p. 3879-84.
33. Bahlis, N.J., *Darwinian evolution and tiding clones in multiple myeloma*. Blood. 120(5): p. 927-8.
34. Keats, J.J., et al., *Clonal competition with alternating dominance in multiple myeloma*. Blood. 120(5): p. 1067-76.
35. Harris, M.A., et al., *The Gene Ontology (GO) database and informatics resource*. Nucleic Acids Res, 2004. 32(Database issue): p. D258-61.
36. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function*. Nucleic Acids Res, 2003. 31(13): p. 3812-4.
37. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2*. Curr Protoc Hum Genet. Chapter 7: p. Unit7 20.
38. Ramensky, V., P. Bork, and S. Sunyaev, *Human non-synonymous SNPs: server and survey*. Nucleic Acids Res, 2002. 30(17): p. 3894-900.
39. Ding, L., et al., *Somatic mutations affect key pathways in lung adenocarcinoma*. Nature, 2008. 455(7216): p. 1069-75.
40. Won, H.H., J.W. Kim, and D. Lee, *A Bayesian ensemble approach with a disease gene network predicts damaging effects of missense variants of human cancers*. Hum Genet. 132(1): p. 15-27.
41. Wood, L.D., et al., *The genomic landscapes of human breast and colorectal cancers*. Science, 2007. 318(5853): p. 1108-13.
42. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nat Methods. 7(4): p. 248-9.
43. Nishi, H., et al., *Cancer missense mutations alter binding properties of proteins and their interaction networks*. PLoS One. 8(6): p. e66273.
44. Jonsson, P.F. and P.A. Bates, *Global topological features of cancer proteins in the human interactome*. Bioinformatics, 2006. 22(18): p. 2291-7.
45. Barabasi, A.L., N. Gulbahce, and J. Loscalzo, *Network medicine: a network-based approach to human disease*. Nat Rev Genet. 12(1): p. 56-68.
46. Xu, J. and Y. Li, *Discovering disease-genes by topological features in human protein-protein interaction network*. Bioinformatics, 2006. 22(22): p. 2800-5.
47. Jeong, H., et al., *Lethality and centrality in protein networks*. Nature, 2001. 411(6833): p. 41-2.
48. Yu, H., et al., *High-quality binary protein interaction map of the yeast interactome network*. Science, 2008. 322(5898): p. 104-10.
49. Magrane, M. and U. Consortium, *UniProt Knowledgebase: a hub of integrated protein data*. Database (Oxford). 2011: p. bar009.
50. Futreal, P.A., et al., *A census of human cancer genes*. Nat Rev Cancer, 2004. 4(3): p. 177-83.
- 50\_a. Wang, J., Duncan, D., Shi, Z., Zhang, B. (2013). WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res, 41 (Web Server issue), W77-83.
51. Schymkowitz, J., et al., *The FoldX web server: an online force field*. Nucleic Acids Res, 2005. 33(Web Server issue): p. W382-8.
52. Lara-Gonzalez, S., et al., *Structural and thermodynamic folding characterization of triosephosphate isomerases from Trichomonas vaginalis reveals the role of destabilizing mutations following gene duplication*. Proteins. 82(1): p. 22-33.
53. Reva, B., Y. Antipin, and C. Sander, *Predicting the functional impact of protein mutations: application to cancer genomics*. Nucleic Acids Res. 39(17): p. e118.

54. Sim, N.L., et al., *SIFT web server: predicting effects of amino acid substitutions on proteins*. Nucleic Acids Res. 40(Web Server issue): p. W452-7.
55. Ng, P.C. and S. Henikoff, *Predicting the effects of amino acid substitutions on protein function*. Annu Rev Genomics Hum Genet, 2006. 7: p. 61-80.
56. Bromberg, Y. and B. Rost, *SNAP: predict effect of non-synonymous polymorphisms on function*. Nucleic Acids Res, 2007. 35(11): p. 3823-35.
57. Ashkenazy, H., et al., *ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids*. Nucleic Acids Res. 38(Web Server issue): p. W529-33.
58. Chen, C., et al., *Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods*. PLoS One. 6(2): p. e17238.
59. Zhan, F., et al., *The molecular classification of multiple myeloma*. Blood, 2006. 108(6): p. 2020-8.
60. Kashimada, K. and P. Koopman, *Sry: the master switch in mammalian sex determination*. Development. 137(23): p. 3921-30.
61. Kim, W.G. and S.Y. Cheng, *Thyroid hormone receptors and cancer*. Biochim Biophys Acta. 1830(7): p. 3928-36.
62. Fox, E.A. and S.R. Kahn, *The relationship between inflammation and venous thrombosis. A systematic review of clinical studies*. Thromb Haemost, 2005. 94(2): p. 362-5.
63. Auwerda, J.J., et al., *Prothrombotic coagulation abnormalities in patients with newly diagnosed multiple myeloma*. Haematologica, 2007. 92(2): p. 279-80.
64. Tundo, G.R., et al., *Insulin-degrading enzyme (IDE): a novel heat shock-like protein*. J Biol Chem. 288(4): p. 2281-9.
65. Norris, J.D., et al., *A negative coregulator for the human ER*. Mol Endocrinol, 2002. 16(3): p. 459-68.
66. Fassas, A.B., et al., *Tamoxifen-based treatment induces clinically meaningful responses in multiple myeloma patients with relapsing disease after autotransplantation*. Leuk Lymphoma, 2001. 42(6): p. 1323-8.
67. Bruns, I., et al., *Multiple myeloma-related deregulation of bone marrow-derived CD34(+) hematopoietic stem and progenitor cells*. Blood. 120(13): p. 2620-30.
68. Deroo, B.J. and K.S. Korach, *Estrogen receptors and human disease*. J Clin Invest, 2006. 116(3): p. 561-70.
69. Doisneau-Sixou, S.F., et al., *Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells*. Endocr Relat Cancer, 2003. 10(2): p. 179-86.
70. Otsuki, T., et al., *Estrogen receptors in human myeloma cells*. Cancer Res, 2000. 60(5): p. 1434-41.
71. Hayon, T., et al., *Non-steroidal antiestrogens induce apoptosis in HL60 and MOLT3 leukemic cells; involvement of reactive oxygen radicals and protein kinase C*. Anticancer Res, 1999. 19(3A): p. 2089-93.
72. Olivier, S., et al., *Raloxifene-induced myeloma cell apoptosis: a study of nuclear factor-kappaB inhibition and gene expression signature*. Mol Pharmacol, 2006. 69(5): p. 1615-23.
73. Kirschner, R.J. and A.L. Goldberg, *A high molecular weight metalloendoprotease from the cytosol of mammalian cells*. J Biol Chem, 1983. 258(2): p. 967-76.
74. Duckworth, W.C., R.G. Bennett, and F.G. Hamel, *Insulin degradation: progress and potential*. Endocr Rev, 1998. 19(5): p. 608-24.
75. Hersh, L.B., *The insulysin (insulin degrading enzyme) enigma*. Cell Mol Life Sci, 2006. 63(21): p. 2432-4.
76. Abdul-Hay, S.O., et al., *Deletion of insulin-degrading enzyme elicits antipodal, age-dependent effects on glucose and insulin tolerance*. PLoS One. 6(6): p. e20818.
77. Tundo, G., et al., *Somatostatin modulates insulin-degrading-enzyme metabolism: implications for the regulation of microglia activity in AD*. PLoS One. 7(4): p. e34376.
78. Kupfer, S.R., E.M. Wilson, and F.S. French, *Androgen and glucocorticoid receptors interact with insulin degrading enzyme*. J Biol Chem, 1994. 269(32): p. 20622-8.
79. Bennett, R.G., et al., *Insulin inhibition of the proteasome is dependent on degradation of insulin by insulin-degrading enzyme*. J Endocrinol, 2003. 177(3): p. 399-405.
80. Kuo, W.L., A.G. Montag, and M.R. Rosner, *Insulin-degrading enzyme is differentially expressed and developmentally regulated in various rat tissues*. Endocrinology, 1993. 132(2): p. 604-11.
81. Adams, J., *The development of proteasome inhibitors as anticancer drugs*. Cancer Cell, 2004. 5(5): p. 417-21.
- 81-a. Brewer, G.A. Analytical Profiles of Drug Substances, volumen 9, Academic Press, New York, N.Y. USA, 1980, págs. 1 - 69.)
82. Richardson, P.G., et al., *Bortezomib: proteasome inhibition as an effective anticancer therapy*. Annu Rev Med, 2006. 57: p. 33-47.

83. Linhart, H.G., et al., *Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing*. Genes Dev, 2007. 21(23): p. 3110-22.
84. Heuck, C.J., et al., *Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis*. J Immunol. 190(6): p. 2966-75.
- 84-a. Aggarwal, B.B., et al., *Suppression of the nuclear factor-kappaB activation pathway by spice-derived phytochemicals: reasoning for seasoning*. Ann N Y Acad Sci, 2004 Dec. 1030: p.434-41.
85. Mirza, S., et al., *Expression of DNA methyltransferases in breast cancer patients and to analyze the effect of natural compounds on DNA methyltransferases and associated proteins*. J Breast Cancer. 16(1): p. 23-31.
86. de Jonge, H.J., et al., *Evidence based selection of housekeeping genes*. PLoS One, 2007. 2(9): p. e898.
87. Zilliox, M.J. and R.A. Irizarry, *A gene expression bar code for microarray data*. Nat Methods, 2007. 4(11): p. 911-3.
88. McCall, M.N., et al., *The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes*. Nucleic Acids Res. 39(Database issue): p. D1011-5.
89. John, E., et al., *Zinc in innate and adaptive tumor immunity*. J Transl Med. 8: p. 118.
90. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell. 144(5): p. 646-74.
91. McCall, K.A., C. Huang, and C.A. Fierke, *Function and mechanism of Zinc metalloenzymes*. J Nutr, 2000. 130(5S Suppl): p. 1437S-46S.
92. Wood, R.J., *Assessment of marginal Zinc status in humans*. J Nutr, 2000. 130(5S Suppl): p. 1350S-4S.
93. Bargellini, A., et al., *Trace elements, anxiety and immune parameters in patients affected by cancer*. J Trace Elem Med Biol, 2003. 17 Suppl 1: p. 3-9.
94. Zuo, X.L., et al., *Levels of selenium, Zinc, copper, and antioxidant enzyme activity in patients with leukemia*. Biol Trace Elem Res, 2006. 114(1-3): p. 41-53.
95. Rink, L. and H. Haase, *Zinc homeostasis and immunity*. Trends Immunol, 2007. 28(1): p. 1-4.
96. Kawata, T., et al., *An improved diagnostic method for chronic hepatic disorder: analyses of metallothionein isoforms and trace metals in the liver of patients with hepatocellular carcinoma as determined by capillary zone electrophoresis and inductively coupled plasma-mass spectrometry*. Biol Pharm Bull, 2006. 29(3): p. 403-9.
97. Martin Mateo, M.C., J. Rabadan, and J. Boustamante, *Comparative analysis of certain metals and tumor markers in bronchopulmonary cancer and colorectal cancers. Metals and tumor markers in the neoplastic process*. Clin Physiol Biochem, 1990. 8(5): p. 261-6.
98. Kwiatek, W.M., et al., *Preliminary study on the distribution of selected elements in cancerous and non-cancerous kidney tissues*. J Trace Elem Med Biol, 2002. 16(3): p. 155-60.
99. Gurusamy, K., *Trace element concentration in primary liver cancers--a systematic review*. Biol Trace Elem Res, 2007. 118(3): p. 191-206.
100. Costello, L.C. and R.B. Franklin, *The clinical relevance of the metabolism of prostate cancer; Zinc and tumor suppression: connecting the dots*. Mol Cancer, 2006. 5: p. 17.
101. Margalioth, E.J., J.G. Schenker, and M. Chevion, *Copper and Zinc levels in normal and malignant tissues*. Cancer, 1983. 52(5): p. 868-72.
102. Logsdon, C.D., et al., *Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer*. Cancer Res, 2003. 63(10): p. 2649-57.
103. Dar, N.A., et al., *Association between copper excess, Zinc deficiency, and TP53 mutations in esophageal squamous cell carcinoma from Kashmir Valley, India--a high risk area*. Nutr Cancer, 2008. 60(5): p. 585-91.
104. Lipman, T.O., et al., *Esophageal Zinc content in human squamous esophageal cancer*. J Am Coll Nutr, 1987. 6(1): p. 41-6.
105. Mellow, M.H., et al., *Plasma Zinc and vitamin A in human squamous carcinoma of the esophagus*. Cancer, 1983. 51(9): p. 1615-20.
106. Poo, J.L., et al., *Diagnostic value of the copper/Zinc ratio in digestive cancer: a case control study*. Arch Med Res, 1997. 28(2): p. 259-63.
107. Kuo, H.W., et al., *Serum and tissue trace elements in patients with breast cancer in Taiwan*. Biol Trace Elem Res, 2002. 89(1): p. 1-11.
108. Oyama, T., et al., *A case-case study comparing the usefulness of serum trace elements (Cu, Zn and Se) and tumor markers (CEA, SCC and SLX) in non-small cell lung cancer patients*. Anticancer Res, 2003. 23(1B): p. 605-12.
109. Oyama, T., et al., *Efficiency of serum copper/Zinc ratio for differential diagnosis of patients with and without lung cancer*. Biol Trace Elem Res, 1994. 42(2): p. 115-27.

110. Doerr, T.D., et al., *Zinc deficiency in head and neck cancer patients*. J Am Coll Nutr, 1997. 16(5): p. 418-22.
111. Kopanski, Z., et al., *The clinical value of the determinations in the serum of Zinc concentration in women with breast cancer*. Folia Histochem Cytobiol, 2001. 39 Suppl 2: p. 84-6.
112. Gupta, S.K., et al., *Serum and tissue trace elements in colorectal cancer*. J Surg Oncol, 1993. 52(3): p. 172-5.
113. Adzersen, K.H., et al., *Raw and cooked vegetables, fruits, selected micronutrients, and breast cancer risk: a case-control study in Germany*. Nutr Cancer, 2003. 46(2): p. 131-7.
114. Fresno, M., et al., *Localization of metallothionein in breast carcinomas. An immunohistochemical study*. Virchows Arch A Pathol Anat Histopathol, 1993. 423(3): p. 215-9.
115. Kagara, N., et al., *Zinc and its transporter ZIP10 are involved in invasive behavior of breast cancer cells*. Cancer Sci, 2007. 98(5): p. 692-7.
116. Taylor, K.M., et al., *ZIP7-mediated intracellular Zinc transport contributes to aberrant growth factor signaling in antihormone-resistant breast cancer Cells*. Endocrinology, 2008. 149(10): p. 4912-20.
117. Lichten, L.A. and R.J. Cousins, *Mammalian Zinc transporters: nutritional and physiologic regulation*. Annu Rev Nutr, 2009. 29: p. 153-76.
118. Viarengo, A., et al., *Antioxidant role of metallothioneins: a comparative overview*. Cell Mol Biol (Noisy-le-grand), 2000. 46(2): p. 407-17.
119. Zowczak, M., et al., *Oxidase activity of ceruloplasmin and concentrations of copper and Zinc in serum of cancer patients*. J Trace Elem Med Biol, 2001. 15(2-3): p. 193-6.

## 17. ANEXOS

### ANEXO 1. Resultados del análisis de predicción de la estructura-función de las proteínas mutadas que unen Zn

En la siguiente tabla, se muestran los programas utilizados y los resultados de predicción obtenidos de cada uno de ellos para cada proteína de unión a Zinc mutada en los pacientes con Mieloma Múltiple.

Nombre de la proteína	Variante de aminoácido	FIS		SIFT		SNAP	POLIPHEN2		FOLDx	CONSURF		Nº de predictores que reportan que la proteína está afectada	¿La mutación afecta a la proteína?
		Predicción	Nº de secuencias	Predicción	Nº de secuencias	Predicción	Predicción	Nº de secuencias	ΔE	Conservación (%)	Nº de secuencias		
ADAM17	D161N	low	19	TOLERATED	68	Neutral	neutral	108	--	51.43	36	0	NO
ADAM18	E457K	medium	590	TOLERATED	93	Non-neutral	deleterious	214	--	71.79	78	3	SI
ADAM29	M356T	high	169	DAMAGING	92	Non-neutral	deleterious	72	--	97.67	129	5	SI
ADAM9	R53G	low	366	TOLERATED	71	Non-neutral	deleterious	62	--	51.26	308	2	NO
ADAMTS1	G594A	high	700	DAMAGING	84	Non-neutral	deleterious	264	--	85.33	157	5	SI
ADAMTS18	S487I	neutral	380	DAMAGING	78	Neutral	deleterious	212	--	20.69	119	2	NO
ADAMTS19	N475K	neutral	233	TOLERATED	79	Neutral	deleterious	64	--	--	--	1	NO
ADAMTS20	D1514G	low	600	TOLERATED	77	Non-neutral	neutral	146	--	41.67	47	1	NO
ADAMTS8	Y703*	--	--	--	--	--	--	--	--	97.22	145	2*	SI
ADAMTS9	N663T	medium	291	TOLERATED	74	Neutral	deleterious	136	--	66	51	2	NO
ADAMTS9	R1431H	medium	634	TOLERATED	76	Non-neutral	deleterious	160	--	25.49	51	3	SI
ADAMTS1	W1710C	medium	568	TOLERATED	76	Non-neutral	deleterious	34	--	76.92	17	3	SI
ADAMTS13	G497R	high	346	DAMAGING	89	Non-neutral	deleterious	47	--	100	22	5	SI
ADNP	D814N	neutral	8	TOLERATED	13	Neutral	neutral	24	--	66.67	9	0	NO
ADNP2	L565S	low	6	TOLERATED	79	Neutral	deleterious	29	--	91.67	12	2	NO
AFG3L2	I486V	medium	700	DAMAGING	80	Non-neutral	deleterious	122	--	97.33	444	5	SI
AGAP2	G506E	low	174	TOLERATED	11	Neutral	deleterious	27	2.34	20	30	2 (PDB)	SI
ARID2	M1625I	neutral	34	TOLERATED	33	Neutral	neutral	35	--	22.22	10	0	NO
ASH1L	R2691*	--	--	--	--	--	--	--	--	100	18	2*	SI
ASH1L	K1355N	low	7	DAMAGING	21	Non-neutral	deleterious	27	--	100	18	4	SI
ASH1L	M1356L	neutral	7	TOLERATED	21	Neutral	neutral	27	--	57.89	18	0	NO
ASH2L	F254S	low	26	DAMAGING*	2	Neutral	deleterious	25	5.86	68.57	35	3 (PDB)	SI
BRCA1	C44F	high	700	DAMAGING*	17	Non-neutral	deleterious	35	9.46	100	12	6	SI
BSN	E795*	--	--	--	--	--	--	--	--	61.54	17	2*	SI
CA3	G111R	medium	697	DAMAGING	81	Non-neutral	deleterious	36	--	--	--	4	SI
CALR3	A299P	medium	290	DAMAGING	93	Neutral	deleterious	35	--	--	--	3	SI
CBLB	W194*	--	--	--	--	--	--	--	Normal: -75.59, Mutated: -20.10	100	14	2*	SI
CHD3	G1755V	neutral	2	TOLERATED	56	Neutral	neutral	108	--	60.47	63	0	NO
CHD3	P2049S	neutral	6	TOLERATED	7	Neutral	neutral	18	--	43.75	63	0	NO
CHD3	V1462I	medium	32	DAMAGING	71	Neutral	deleterious	28	--	96.83	63	4	SI
CLIP1	I466M	low	30	DAMAGING	7	Non-neutral	deleterious	63	--	72.09	57	3	SI
CPSF3	S589L	low	57	TOLERATED	66	Neutral	neutral	224	--	55.83	191	0	NO
CTCF	G510V	low	700	DAMAGING	89	Non-neutral	deleterious	37	--	72	25	3	SI
DNAAJ1	Q159P	high	700	DAMAGING	74	Neutral	neutral	152	--	30.41	404	2	NO
DNMT3B	K421E	neutral	37	TOLERATED	46	Neutral	neutral	76	--	55.77	52	0	NO
EGR1	Q9H	low	13	DAMAGING*	8	Neutral	deleterious	22	--	100	29	3	SI
EHMT1	A988T	low	700	DAMAGING*	183	Neutral	neutral	157	3.27	83.33	30	3 (PDB)	SI
ESR1	Q498P	neutral	180	DAMAGING*	71	Neutral	deleterious	85	-1.07	66.2	143	2	NO



FNTB	R291C	high	347	DAMAGING	74	Non-neutral	deleterious	78	2.61	100	154	6	SI
FOXP1	P407L	medium	48	DAMAGING	70	Neutral	neutral	355	--	87.5	48	3	SI
FOXP2	S100T	neutral	33			Non-neutral	deleterious	49	--	19.8	101	2	NO
GIT1	S368G	neutral	19	TOLERATED	5	Neutral	neutral	187	--	36.51	64	0	NO
HNH4A	R331H	medium	238	DAMAGING	69	Neutral	deleterious	204	0.16	31.76	191	3	NO
IDE	H134P	high	514	DAMAGING	68	Non-neutral	deleterious	77	6.78	71.94	143	5	SI
INTS12	M445I	low	8	DAMAGING*	4	Neutral	neutral	30	--	100	10	2	NO
LMO2	E135D	neutral	552	TOLERATED	65	Neutral	deleterious	29	0.13	52.35	174	1	NO
MBD1	G471A	low	7	TOLERATED	33	Neutral	deleterious	77	--	96.3	28	2	NO
MDM4	Q262H	neutral	13	DAMAGING	11	Neutral	neutral	38	--	37.5	21	1	NO
METAP2	A2S	neutral	1	DAMAGING*	15	Neutral	deleterious	20	--	100	191	3	SI
MID2	R105H	medium	182	DAMAGING	64	Non-neutral	deleterious	117	--	87.5	27	5	SI
MIPEP	L171F	low	100	TOLERATED	78	Neutral	neutral	95	--	23.64	59	0	NO
MKRN3	D483V	low	107	DAMAGING	46	Neutral	neutral	25	--	68.42	67	1	NO
MLL	M2022I	low	60	DAMAGING*	17	Non-neutral	deleterious	37	--	100	9	4	SI
MLL2	F3556S	low	10	TOLERATED	41	Non-neutral	deleterious	34	--	--	--	2	NO
MLL2	H5475L	high	700	DAMAGING	55	Non-neutral	deleterious	32	--	--	--	4	SI
MLL3	Y987H	medium	22	DAMAGING	51	Non-neutral	deleterious	35	--	100	16	5	SI
MLPH	V112F	medium	69	TOLERATED	55	Neutral	deleterious	80	-0.63	74.42	44	2	NO
MMP16	Y459D	high	456	DAMAGING	72	Non-neutral	deleterious	241	--	77.85	278	4	SI
MT1A	C36F	medium	56	DAMAGING*	82	Non-neutral	neutral	17	10.57	92.59	54	5 (PDB)	SI
MYO9A	T1974P	low	183	TOLERATED	45	Non-neutral	deleterious	55	--	66.67	13	2	NO
MYSM1	Q387H	neutral	123	TOLERATED	39	Neutral	deleterious	36	-0.25	46.67	15	1	NO
MYT1	M1035T	medium	21	DAMAGING*	8	Non-neutral	deleterious	124	--	98.18	55	5	SI
NR2F2	S223L	neutral	405	TOLERATED	43	Neutral	deleterious	66	-0.1	19.58	222	1	NO
NR3C2	E785K	medium	110	DAMAGING*	50	Non-neutral	deleterious	477	-0.43	50	12	4	SI
NR3C2	F626C	high	699	DAMAGING	61	Non-neutral	deleterious	442	--	100	14	5	SI
OSR2	T55M	low	25	DAMAGING*	21	Neutral	deleterious	45	--	100	57	3	SI
PDE3A	N681I	neutral	166	TOLERATED	57	Neutral	deleterious	47	-1.1	31.25	33	1	NO
PDE8B	V344G	low	609	DAMAGING*	4	Non-neutral	neutral	95	--	11.76	53	2	NO
PDZD8	R138H	low	24	TOLERATED	60	Non-neutral	deleterious	37	--	76.92	14	2	NO
PHC1	V250L	neutral	25	TOLERATED	66	Neutral	neutral	24	--	90.9	10	1	NO
POLR3A	D66H	low	360	TOLERATED	66	Neutral	neutral	52	--	12.95	329	0	NO
PRICKLE2	R390W	low	17	DAMAGING	25	Non-neutral	deleterious	34	--	60.71	28	3	SI
PRKCZ	Y126D	medium	20	DAMAGING	45	Non-neutral	deleterious	132	--	59.32	171	4	SI
RAG1	G954D	low	11	DAMAGING*	36	Non-neutral	deleterious	44	--	88.4	250	4	SI
RAG2	R39*	--	--	--	--	--	--	--	--	--	--	2*	SI
RBM5	W547L	medium	84	DAMAGING	64	Non-neutral	deleterious	101	--	100	39	5	SI
RELN	D207N	low	13	TOLERANT	27	Neutral	deleterious	20	--	85.71	12	2	NO
RIMS2	P973L	low	18	TOLERANT	41	Neutral	neutral	123	--	60.47	43	0	NO
RIMS2	R354W	medium	18	TOLERANT	45	Neutral	deleterious	120	--	95.24	43	3	SI
RNF168	L71H	medium	16	DAMAGING	84	Neutral	deleterious	87	2.62	100	21	5 (PDB)	SI
RNF19A	I795V	low	34	TOLERANT	11	Neutral	neutral	21	--	72.73	22	0	NO
RNF213	E2081K	medium	16	DAMAGING*	1	Non-neutral	neutral	56	--	100	14	4	SI
RNF26	C404F	high	700	DAMAGING	79	Non-neutral	deleterious	35	--	100	15	5	SI
RORB	C299F	low	128	TOLERANT	73	Neutral	deleterious	116	4.98	31.67	62	2 (PDB)	SI
SF3A3	D144Y	medium	107	DAMAGING	71	Neutral	deleterious	29	--	15.28	75	3	SI
SMAP2	N200K	medium	25	TOLERANT	62	Neutral	neutral	22	--	75.93	57	1	NO
SOBP	R245G	low	14	TOLERANT	55	Non-neutral	neutral	35	--	86.36	21	2	NO
SP140	S474T	low	22	TOLERANT	29	Neutral	neutral	23	--	100	14	1	NO



SP8	S84F	low	6	TOLERANT	49	Non-neutral	deleterious	18	--	45.45	37	2	NO
SRRM2	S510F	neutral	4	DAMAGING	62	Non-neutral	deleterious	28	--	33.53	170	3	SI
ST18	E4K	low	18	DAMAGING	20	Neutral	neutral	32	--	71.43	55	1	NO
STAC	R169C	medium	17	DAMAGING	36	Non-neutral	deleterious	38	--	100	43	5	SI
TADA2A	E33K	low	106	TOLERANT	70	Neutral	neutral	35	--	47.62	21	0	NO
TET1	H1547P	low	36	TOLERANT	72	Neutral	deleterious	78	--	42.86	14	1	NO
TET1	Q1504H	neutral	36	TOLERANT	72	Neutral	deleterious	79	--	78.57	14	1	NO
THRB	R354*	--	--	--	--	--	--	--	Normal: -15.46, Mutated: 19.83	100	66	2* (PDB)	SI
TIMP2	G114E	high	80	DAMAGING	73	Non-neutral	deleterious	45	18.63	100	54	6	SI
TLL1	R219*	--	--	--	--	--	--	--	Normal: -55.11, Mutated: -0.80	84.42	81	2* (PDB)	SI
TP53	C275S	medium	79	DAMAGING	84	Non-neutral	deleterious	121	-0.25	100	65	5	SI
TP53	R196*	--	--	--	--	--	--	--	Normal: 11.09, Mutated: 45.52	96.88	65	2* (PDB)	SI
TP53	Y205D	medium	79	DAMAGING	85	Non-neutral	deleterious	120	6.16	98.44	65	6	SI
TRAF3	D551V	medium	99	DAMAGING*	7	Non-neutral	deleterious	123	1.14	82.14	59	6	SI
TRAFD1	N519K	low	17	TOLERANT	27	Non-neutral	deleterious	33	--	35.71	17	2	NO
TRIM10	T41P	low	700	TOLERANT	84	Non-neutral	deleterious	32	3.52	54.29	75	3 (PDB)	SI
TRIM17	V132L	medium	437	TOLERANT	85	Neutral	deleterious	34	4.25	56.41	80	3 (PDB)	SI
TRIM29	S301N	neutral	4	TOLERANT	76	Neutral	neutral	39	--	57.14	14	0	NO
TRIM47	G445E	neutral	120	TOLERANT	52	Neutral	deleterious	33	--	100	11	2	NO
TRIM6	V372I	low	451	TOLERANT	92	Neutral	neutral	30	--	--	--	0	NO
TSHZ2	F634L	neutral	23	TOLERANT	52	Neutral	neutral	37	--	31.11	47	0	NO
TSHZ2	L757Q	low	23	TOLERANT	69	Neutral	deleterious	37	--	19.15	47	1	NO
TSHZ3	R1004W	neutral	14	TOLERANT	34	Non-neutral	deleterious	27	--	28	50	2	NO
TSHZ3	S58G	neutral	17	DAMAGING	64	Neutral	neutral	26	--	85	50	2	NO
UBA2	R612K	neutral	3	TOLERANT	57	Neutral	deleterious	37	--	74.42	135	1	NO
UPF1	A771T	low	699	DAMAGING*	65	Neutral	neutral	301	0.65	40.35	115	2 (PDB)	SI
USP19	N535Y	medium	563	DAMAGING	41	Non-neutral	deleterious	50	--	100	17	5	SI
USP22	L114V	neutral	74	TOLERANT	62	Non-neutral	neutral	108	--	12.96	58	1	NO
VAT1L	G262V	medium	52	DAMAGING	70	Neutral	deleterious	36	--	63.46	52	3	SI
WHSC1	A332D	low	14	DAMAGING	61	Neutral	deleterious	53	--	95.45	24	3	SI
WHSC1	C1191F	high	572	DAMAGING	78	Non-neutral	deleterious	45	30.05	100	24	6	SI
WHSC1L1	K888N	neutral	100	TOLERANT	3	Non-neutral	deleterious	26	--	87.5	24	3	SI
WHSC1L1	Y887C	neutral	100	DAMAGING*	3	Non-neutral	deleterious	26	--	62.5	24	3	SI
ZBTB3	R503C	low	700	DAMAGING	77	Non-neutral	deleterious	30	--	62.5	24	3	SI
ZBTB33	I395M	neutral	5	TOLERANT	47	Neutral	deleterious	27	--	33.33	12	1	NO
ZBTB45	P384L	neutral	4	TOLERANT	86	Neutral	deleterious	23	--	56.25	17	1	NO
ZC3H14	E339K	medium	14	TOLERANT	47	Neutral	deleterious	49	--	100	17	3	SI
ZC3H15	Q264R	low	146	TOLERANT	70	Neutral	deleterious	46	--	17.7	115	1	NO
ZC3H18	S53T	neutral	14	TOLERANT	49	Neutral	neutral	32	--	22.22	13	0	NO
ZCCHC5	E217K	neutral	4	TOLERANT	67	Neutral	neutral	24	--	44.44	17	0	NO
ZFPM2	S647P	neutral	8	TOLERANT	33	Neutral	neutral	32	--	--	--	0	NO
ZFYVE1	M152V	neutral	39	TOLERANT	46	Neutral	neutral	43	--	13.89	37	0	NO
ZKSCAN3	Q361H	low	700	TOLERANT	92	Non-neutral	deleterious	35	--	62	311	1	NO
ZKSCAN5	E427K	low	700	DAMAGING	72	Non-neutral	deleterious	30	0.15	91.22	423	4	SI
ZMYM4	I1182V	neutral	32	TOLERANT	60	Neutral	neutral	77	--	21.21	34	0	NO
ZNF148	S591A	neutral	14	TOLERANT	16	Neutral	neutral	20	--	52.63	21	0	NO
ZNF193	Y310C	medium	700	DAMAGING	91	Non-neutral	deleterious	28	--	85.33	222	5	SI
ZNF319	H528R	low	700	TOLERANT	77	Non-neutral	deleterious	52	--	34.23	435	2	NO
ZNF329	E235*	--	--	--	--	--	--	--	--	40	363	2*	SI
ZNF416	Y522N	medium	700	DAMAGING	91	Non-neutral	deleterious	27	--	90.07	518	5	SI

ZNF417	C30Y	low	699	DAMAGING	30	Neutral	neutral	23	--	7.2	250	1	NO
ZNF423	E426Q	neutral	700	TOLERANT	64	Neutral	neutral	138	--	56.25	16	0	NO
ZNF462	P744L	neutral	7	TOLERANT	47	Neutral	neutral	25	--	--	--	0	NO
ZNF483	R543Q	low	700	DAMAGING	11	Non-neutral	deleterious	28	--	78.52	506	3	SI
ZNF484	G474R	medium	700	TOLERANT	89	Non-neutral	deleterious	39	1.55	97.32	494	5 (PDB)	SI
ZNF492	P420T	medium	700	DAMAGING	96	Neutral	deleterious	123	--	88.59	490	4	SI
ZNF492	T325N	neutral	700	TOLERANT	97	Neutral	neutral	114	--	76.03	490	0	NO
ZNF527	Y302N	medium	699	DAMAGING	86	Non-neutral	neutral	30	--	79.87	521	3	SI
ZNF530	Q398E	neutral	700	TOLERANT	94	Non-neutral	deleterious	128	--	67.35	503	2	NO
ZNF559	E326K	neutral	700	TOLERANT	93	Non-neutral	deleterious	105	--	72.48	538	2	NO
ZNF569	H516Y	high	700	DAMAGING	84	Non-neutral	deleterious	28	--	100	518	5	SI
ZNF630	Q394K	neutral	700	TOLERANT	94	Neutral	neutral	20	--	59.59	522	0	NO
ZNF646	Q1397R	neutral	8	TOLERANT	74	Non-neutral	neutral	28	--	63.64	11	0	NO
ZNF667	T213A	low	700	TOLERANT	78	Non-neutral	neutral	31	--	19.33	533	1	NO
ZNF700	V134F	low	546	DAMAGING	62	Neutral	deleterious	64	--	14.19	504	2	NO
ZNF749	E358Q	low	700	TOLERANT	92	Non-neutral	neutral	17	--	--	--	1	NO
ZNF761	P47L	--	--	--	--	--	neutral	80	--	--	--	0	NO
ZNF804A	E338K	low	23	TOLERANT	37	Neutral	neutral	33	--	57.14	20	0	NO
ZNF83	E293V	neutral	700	TOLERANT	100	Non-neutral	neutral	19	--	100	384	2	NO
ZNF830	W365C	low	9	DAMAGING	82	Non-neutral	deleterious	35	--	100	16	4	SI
ZNF878	S510C	medium	700	TOLERANT	97	Non-neutral	deleterious	48	--	52	515	3	SI
ZNRF3	A807P	low	11	DAMAGING*	22	Neutral	deleterious	34	--	83.33	14	3	SI
ZRANB3	R171S	low	699	DAMAGING	87	Neutral	neutral	74	--	31.58	20	1	NO
ZSCAN4	R384Q	low	700	TOLERANT	91	Neutral	deleterious	65	--	--	--	1	NO
ZZEF1	D2773H	low	9	TOLERANT	28	Non-neutral	deleterious	53	--	100	16	3	SI
ZZEF1	H1877P	neutral	9	DAMAGING	39	Neutral	deleterious	48	--	41.18	16	2	NO
ZZEF1	T2924N	neutral	9	DAMAGING	65	Neutral	neutral	45	--	71.43	16	1	NO

## **ANEXO 2. Determinación de una fórmula para integrar los resultados de los predictores de E-F en un valor global de afectación**

Se quiso obtener una fórmula capaz de integrar a los 6 predictores de E-F y que pudiera reflejar la certeza de que una proteína está afectada considerando que cuanto mayor es el número de predictores que predicen que está afectada, mayor debe ser el valor global de afectación obtenido por la fórmula. Las variantes que consideramos que definen con mayor validez la afectación de una proteína dentro de cada predictor son el número de secuencias homólogas con las que se realiza el análisis de predicción y el puntaje final obtenido por cada predictor. El número de secuencias lo consideramos importante para la fórmula ya que cuanto mayor es el número de secuencias con el que se hace el análisis es más probable que el resultado final del predictor sea certero. En base a esto iniciamos desarrollando una fórmula que nos permitiera integrar toda esta información (fórmula 1):

*Fórmula 1:*

*Valor global de afectación= (predicción SIFT) (nº de seq SIFT) + (predicción FIS)(nº de seq FIS) + ... otros predictores + predicción FoldX* ***r= 0.56***

Para ver si existe una correlación positiva entre el valor global de afectación obtenido por la fórmula y el número de predictores que predicen que está afectada la proteína, determinamos la correlación de Pearson. Para esta fórmula obtuvimos una  $r= 0.56$ .

Como fue una correlación baja, probamos aplicar la misma fórmula pero normalizando los valores obtenidos por FoldX a una escala de -1 a 1, de tal forma que la escala sea como la de los otros predictores. El valor de -1 hace referencia a los valores de Foldx que tuvieron un delta  $\Delta\Delta G \leq 0.5$  (estructura de la proteína estable), los valores iguales a 0 corresponden a las proteínas que no pudieron ser analizadas por FoldX, y los valores de 1 corresponden a los valores de Foldx con un  $\Delta\Delta G > 0.5$  (estructura de la proteína inestable) (fórmula 2).

*Fórmula 2:*

*Valor global de afectación= (predicción SIFT) (nº de seq SIFT) + (predicción FIS)(nº de seq FIS) + ... otros predictores + predicción FoldX (normalizado de -1 a 1).* ***r= 0.75***

Con esta fórmula obtuvimos una  $r=0.75$ . Obtuvimos de esta manera una correlación más alta, sin embargo, probamos quitar FoldX de la fórmula para ver si este predictor está metiendo ruido en los resultados (fórmula 3).

Fórmula 3:

Valor global de afectación= (predicción SIFT) (nº de seq SIFT) + (predicción FIS)(nº de seq FIS) + ... otros predictores

$$r=0.55$$

Obtuvimos una  $r=0.55$ , por lo tanto, de estas tres fórmulas concluimos que la fórmula 2 hasta ahora es la que mejor se ajusta a una correlación positiva, es decir, cuanto mayor es la cantidad de predictores que dicen que la proteína está afectada, el valor global de afectación tiende a ser más alto.

Por otro lado, para ver si podíamos aumentar la  $r$ , probamos sumar todos los valores obtenidos por todos los predictores y multiplicarlos con la suma de todas las secuencias correspondientes a los predictores (fórmula 4).

Fórmula 4:

Valor global de afectación= (suma de las predicciones de todos los programas)(suma de nº de secuencias de todos los programas)

$$r=0.63$$

Obtuvimos una correlación baja con una  $r=0.63$ . Decidimos entonces sumar el valor de FoldX a la fórmula 4 (fórmula 5).

Fórmula 5:

Valor global de afectación= (suma de las predicciones de todos los programas)(suma de nº de secuencias de todos los programas)+ FoldX

$$r=0.56$$

La fórmula 5 pero con FoldX normalizado de -1 a 1 nos dio una mejor correlación con una  $r=0.65$  (fórmula 6)

Fórmula 6:

Valor global de afectación= (suma de las predicciones de todos los programas)(suma de nº de secuencias de todos los programas)+ FoldX (-1 a 1)

$$r=0.65$$

Pudimos notar que de estas tres últimas fórmulas, nos dio una mejor correlación la que incluye la suma de FoldX normalizado (fórmula 6), sin embargo dicha correlación es más baja con respecto a la fórmula 2. Probamos entonces eliminar las multiplicaciones de las fórmulas anteriores pensando que probablemente multiplicar las secuencias baja el valor de  $r$ . Por lo tanto, sumamos los valores de predicción de los programas con el número de secuencias de cada uno de ellos (fórmula 7).

Fórmula 7:

Valor global de afectación= (predicción SIFT) + (nº de seq SIFT) + (predicción FIS)+ (nº de seq FIS) + ... otros predictores

**r= 0.81**

En este caso obtuvimos una r alta igual a 0.81. Probamos entonces sumarle FoldX (fórmula 8)

Fórmula 8:

Valor global de afectación= (predicción SIFT) + (nº de seq SIFT) + (predicción FIS)+ (nº de seq FIS) + ... otros predictores + FoldX

**r= 0.63**

Como bajó mucho la correlación le sumamos a la fórmula 7 a FoldX normalizado (fórmula 9)

Fórmula 9:

Valor global de afectación= (predicción SIFT) + (nº de seq SIFT) + (predicción FIS)+ (nº de seq FIS) + ... otros predictores + FoldX (-1 a 1)

**r= 0.83**

Con la fórmula 9 obtuvimos una muy buena correlación. Sin embargo, como el valor de predicción que se obtiene de cada programa ya se encuentra influenciado por el número de secuencias con las que hizo la predicción, decidimos quitar el número de secuencias de las fórmulas y probar si el número de secuencias también está metiendo algún ruido en las correlaciones que estuvimos obteniendo. Por lo tanto, quitamos el número de secuencias a la fórmula 7 (fórmula 10).

Fórmula 10:

Valor global de afectación= suma del valor de predicción de todos los programas

**r= 0.91**

Posteriormente, a la fórmula 10 le sumamos por un lado a Foldx sin normalizar y por el otro a FoldX normalizado (fórmulas 11 y 12).

Fórmula 11:

Valor global de afectación= suma del valor de predicción de todos los programas + FoldX

**r= 0.62**

Fórmula 12:

Valor global de afectación= suma del valor de predicción de todos los programas + FoldX (-1 a 1) **r= 0.91**

La suma de todos los predictores con FoldX normalizado y sin FoldX nos dio la correlación más alta con una  $r=0.91$ . Por lo tanto, decidimos utilizar la fórmula 12, que es la que integra a todos los predictores.

A continuación, se presentan los gráficos de correlación correspondientes a las fórmulas 9 y 12 y los análisis que se hicieron con cada una de ellas, ya que la fórmula 9 (gráfico 1-anexo) presentó la correlación más alta integrando el número de secuencias y la fórmula 12 (gráfico 2-anexo) presentó la correlación más alta sin integrar el número de secuencias.

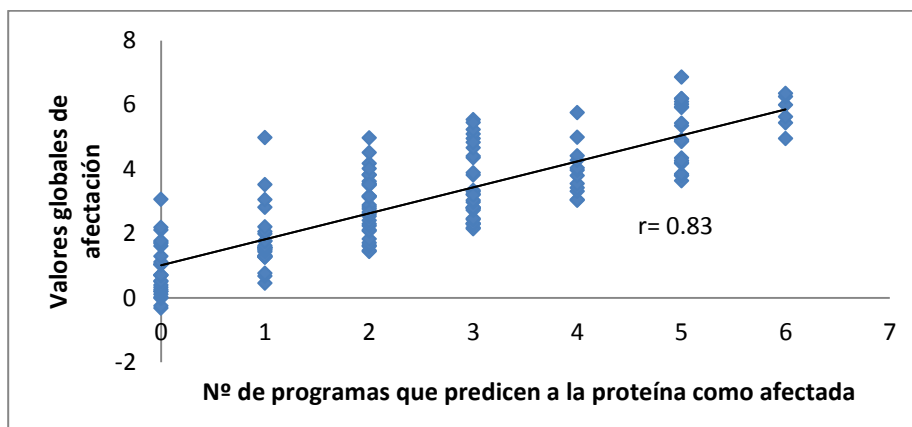


Gráfico 1-anexo. Correlación de los valores globales de afectación de la fórmula 9 con respecto al número de programas que las predicen como afectadas

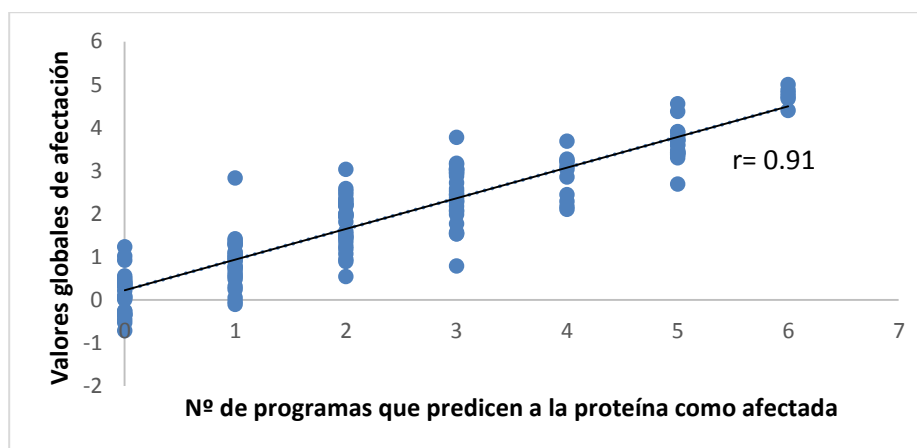
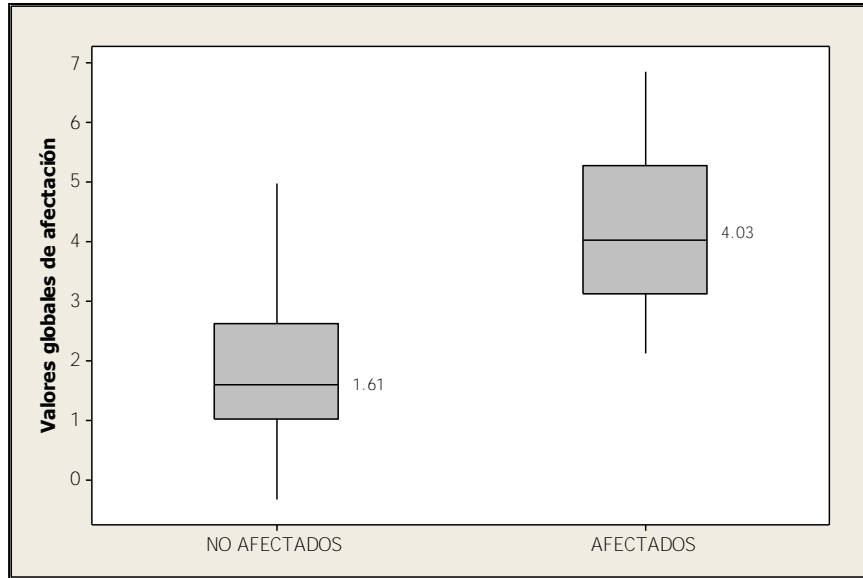
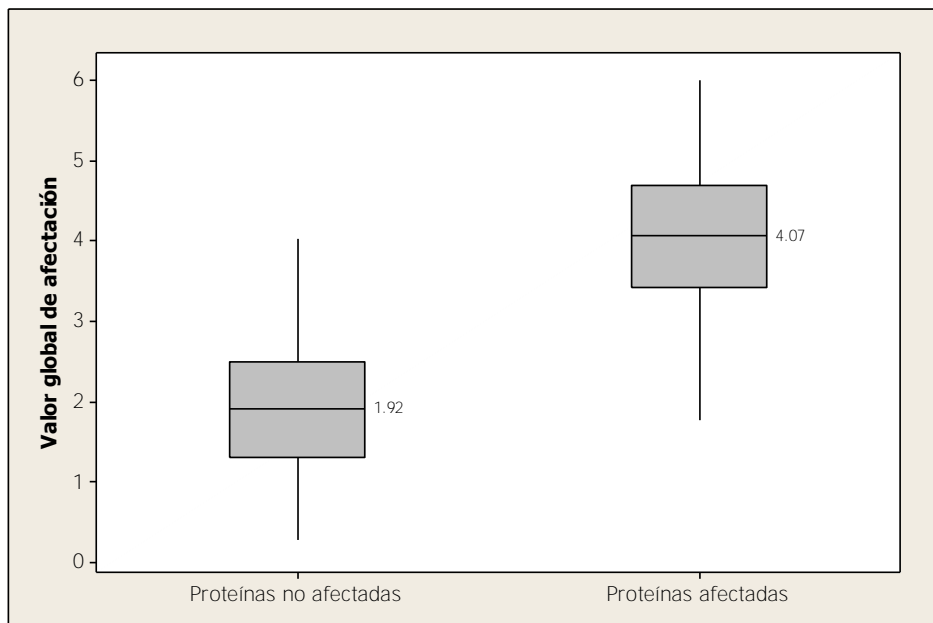


Gráfico 2-anexo. Correlación de los valores globales de afectación de la fórmula 12 con respecto al número de programas que las predicen como afectadas

Se analizó con cada fórmula cuáles valores globales de afectación presentan las proteínas que no están afectadas y las que sí lo están según el número de predictores (ver sección análisis E-F). Se representaron los datos en gráficos de cajas (gráficos 3-anexo y 4-anexo) y posteriormente se calculó el intervalo de confianza obtenido con cada fórmula.



**Gráfico 3-anexo. Valores globales de afectación de las proteínas afectadas y no afectadas (fórmula 9).** Fórmula 9: Valor global de afectación= (predicción SIFT) + (nº de seq SIFT) + (predicción FIS)+ (nº de seq FIS) + ... otros predictores + FoldX (-1 a 1)).



**Gráfico 4-anexo. Valores globales de afectación de las proteínas afectadas y no afectadas (fórmula 12).** Fórmula 12: Valor global de afectación= suma del valor de predicción de todos los programas + FoldX (-1 a 1).

Se puede ver que los valores globales de afectación obtenidos por la fórmula 12 tienden a estar más distanciados entre las proteínas afectadas y no afectadas que los de la fórmula 9. Además, se puede visualizar que existe una mayor variabilidad en los valores globales de afectación obtenidos por la fórmula 9 en comparación con la fórmula 12.

Al calcular los intervalos de confianza al 95%, la fórmula 9 presenta intervalos de 1.55 a 2.09 para las proteínas no afectadas y de 3.9 a 4.5 para las proteínas afectadas. El IC al 95% para la fórmula 12 es de 1.75 a 2.14 para las proteínas no afectadas y de 3.77 a 4.19 para las afectadas.

Debido a que con la fórmula 12 se obtiene una mejor correlación con respecto al número de predictores que predicen que una proteína está afectada, a que los valores globales de afectación tienden a estar más separados entre las proteínas afectadas y no afectadas y además existe menos variabilidad entre los valores globales de afectación dentro de cada grupo de proteínas, la fórmula 12 fue la que se decidió utilizar para integrar los diferentes predictores y para evaluar el grado de afectación de cada proteína mutada.



### ANEXO 3. Scripts para interactomas

Para correr interactomas se debe utilizar Linux

#### 1. Script para generar 1000 listas al azar:

Utilizar el script: listas\_random.pl (es el programa que ejecuta listas al azar y genera l sublistas de g elementos de la lista i)

Entrada: lista de genes de la cual se van a generar las listas al azar

Archivo: ALTERED-GENE-LIST-HYPERMUTATORS-EXCLUDED.csv

Salida: 1000 listas de genes al azar

Archivo: listas\_random.csv

Opciones del script: -g número de elementos

-l número de sublistas

-i nombre del archivo csv con la lista

-o nombre del archivo de salida

Poner en la terminal:

```
Perl listas_random.pl -i "Nombre del archivo que contiene la lista
de genes de donde se crearan las listas al azar.csv" -l 1000 -g 70 -
o listas_random.csv
```

#### 2. Script para correr los interactomas:

Utilizar el script: correr\_listas.pl (es el programa que corre las mil listas al azar de 70 genes cada una del archivo i).

Entrada: 1000 listas de genes al azar

Archivo: listas\_random.csv

Salida: 1000 interactomas analizados

Carpetas: "log", "modules", "Node-Type", "Networks" y "report"

Opciones del script: -i nombre del archivo csv con las listas al azar

Poner en la terminal:

```
perl correr_listas.pl -i listas-random.csv
```

### 3. Script para extraer tabla de datos de los interactomas:

Cuando finalicen las corridas de los interactomas se podrá extraer una tabla con los datos obtenidos de cada interactoma con el global model y el local model.

Para ello se utilizó el script: `extraer_tabla.pl` (extrae los datos obtenidos por interactoma en una tabla)

Salida: tabla con los datos de los 1000 interactomas

Archivo: `tabla.csv`

Poner en la terminal:

```
perl extraer_tabla.pl > tabla.csv
```

Los datos obtenidos por tabla de los 1000 interactomas son los siguientes:

Interactoma analizado	N° de genes analizados	Nodos	Edges	linkers	módulos	modularidad
--------------------------	---------------------------	-------	-------	---------	---------	-------------