Universidad Nacional Autónoma de México

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

Consideraciones en las Extensiones de los Análisis de Enriquecimiento de Datos ómicos

T E S I N A

QUE PARA OPTAR POR EL GRADO DE:

Especialista en Estadística Aplicada

PRESENTA:

Ricardo Omar Ramírez Flores

DIRECTOR:
Julio Saez Rodriguez

Ciudad de México, México, Octubre 2018

Write in recollection and amazement for yourself.
-Jack Kerouac

Reconocimientos

Gracias a Aurelién Dugourd por las discusiones y propuestas al trabajo, así como al grupo de Biomedicina Computacional de la Universidad de Heidelberg, dirigido por mi tutor el Prof. Julio Saez-Rodriguez. La curación de experimentos utilizada en este trabajo para la estimación de actividades de kinasas proviene del esfuerzo de Claudia Hernández, del EMBL-EBI. Luz García es quien facilitó los datos para la estimación de actividades de factores de transcripción. Los experimentos realizados por el grupo del Prof. Thorsten Cramer del Hospital de la RWTH Aachen fueron los utilizados como caso de estudio en este trabajo. Gracias a Miguel Ángel Ibarra por las sugerencias y correcciones al texto original, así como de su apoyo moral.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea el resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Ricardo Omar Ramírez Flores. Ciudad de México, México, Octubre 2018

Resumen

Los análisis de enriquecimiento se han vuelto un método de interpretación obligatorio en los estudios *ómicos* funcionales, principalmente por el ruido inherente de los análisis comparativos de componentes biológicos individuales y por la necesidad de ubicarlos en un contexto más significativo. A pesar de que existen varios esfuerzos para integrar y mejorar los métodos de enriquecimiento más ampliamente usados, aún existen desafíos analíticos y computacionales. En esta tesina se presenta una extensión al marco de trabajo generalizado de SAFE (Análisis de Significancia de Categorías Funcionales y de Expresión) para considerar pesos de contribución y confianza en el cálculo de scores de enriquecimiento y su significancia estadística. Se evaluó y comparó la efectividad de SAFE extendido en el contexto de la estimación de actividades de factores de transcripción y kinasas, y se desarrolló una herramienta computacional en *R*, *eGSA*, que incorpora las opciones de análisis ponderado expuestos aquí.

Índice general

Capítulo 1

Introducción

Tradicionalmente, las ciencias biológicas han sido relacionadas a investigaciones en las cuales solamente es posible obtener información relevante de un conjunto reducido de componentes, por ejemplo, la identificación de un gen o un pequeño grupo de proteínas relevantes en una enfermedad. Sin embargo, en las últimas dos décadas, una transición importante ha sucedido en esta área gracias al rápido desarrollo y evolución de tecnologías de alto rendimiento que permiten obtener mediciones moleculares a escalas genómicas o de sistemas. La ahora llamada era *ómica* de la biología, es un ambiente rico de datos en el que se han podido elucidar mecanismos celulares sin precedentes [Joyce, et al., 2006].

Los datos *ómicos* no engloban únicamente el catálogo de componentes que forman una célula, sino también las interacciones entre ellos y sus propiedades emergentes. A diferentes grados de resolución, es posible rastrear con las tecnologías actuales la organización celular en condiciones y tiempos específicos. No obstante, los retos que enfrenta el área involucran la correcta extracción de conocimiento del mar de información provista y la integración efectiva de diferentes clases de datos para mejorar el entendimiento de los sistemas. Es por esto que la Biología se ha vuelto un punto de encuentro interdisciplinario, donde se han explotado diferentes técnicas estadísticas y de ciencias de la computación [Joyce, et al., 2006].

La organización celular involucra un complicado conjunto de interacciones entre diferentes capas de complejidad biológica, en el que se codifica, traduce y ejecuta toda acción que rige la vida en el planeta. En el contexto de los datos ómicos, es posible identificar las capas de complejidad a partir de la clase de mediciones que se pueden realizar (Fig. 1.1):

Componentes: En esta clasificación se encuentran todos los datos que forman el catálogo de elementos funcionales de la célula tales como genes, transcritos de ARN, proteínas y metabolitos. Dejando a un lado los datos de secuencia genómica, que son descriptivos, el resto de las mediciones de componentes generan datos

continuos o discretos [Joyce, et al., 2006].

Regulaciones: Esta clase de datos contiene mediciones de las compuertas lógicas que existen en los procesos celulares. La expresión y función de un transcrito o una proteína depende de una regulación estricta ya sea a nivel de ADN, tales como las modificaciones epigenéticas [Roadmap Epigenomics Consortium, et al., 2015], o a nivel de componente, tales como las modificaciones post-traduccionales de proteínas [Hernandez-Armenta, et al., 2017] o los niveles de expresión de reguladores [Garcia-Alonso, et al., 2018]. Algunas de estas mediciones pueden ser binarias-lógicas o continuas, dependiendo del contexto.

Interacciones: Es posible medir también las interacciones directas (físicas) o indirectas que suceden entre diferentes componentes celulares. Estas interacciones generan redes pesadas a las cuáles se les pueden asignar niveles de confianza. Las redes de señalización (generadas por interacciones entre proteínas regidas por procesos regulatorios) [Turei, et al., 2016], las redes de dominios de DNA y las redes regulatorias transcripcionales [Ibarra-Arellano, et al., 2016], son ejemplos de este tipo de datos.

Funciones: En algunas circunstancias es posible medir directamente los niveles de algún proceso celular, sin embargo, la mayoría de las mediciones funcionales se refieren a la integración de las otras 3 clases de datos mencionadas anteriormente. Por ejemplo, es posible medir indirectamente las actividades de las rutas de señalización celular a partir de redes de interacción proteína-proteína, estados de regulación post-traduccional y niveles de transcritos [Schubert, et al. 2018].

La genómica funcional busca entender cuál es la relación entre el genotipo (la información genética) y el fenotipo (las características expresadas) a una escala global y sistémica. Este enfoque permite explicar cómo es que diferentes componentes celulares trabajan en conjunto para mantener la célula funcionando. Entre los esfuerzos de esta área se encuentran los métodos de enriquecimiento que incorporan conocimiento biológico existente a resultados provenientes de análisis estadísticos de componentes individuales.

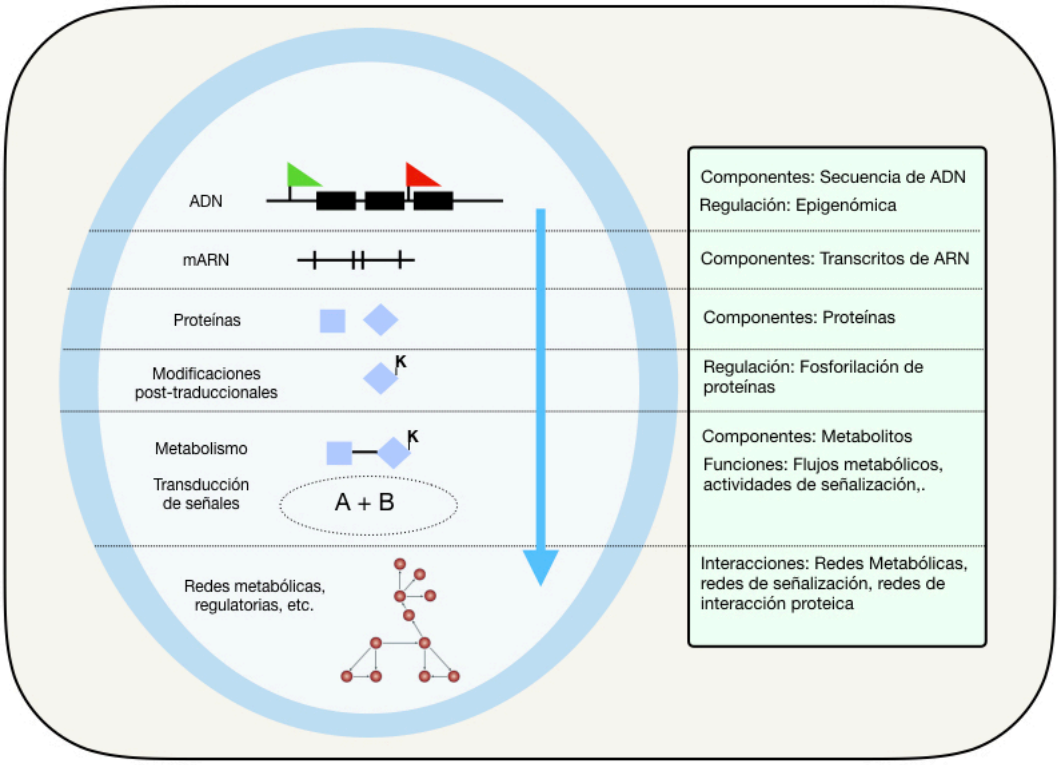

Figura 1.1: La organización celular puede observarse a grandes rasgos a partir de las diferentes mediciones que realizan las tecnologías de alto rendimiento. En la visión más sencilla, la información codificada en el ADN es transformada en proteínas a partir de un intermediario de ARN (transcrito). Diferentes grupos de proteínas trabajan en conjunto en el metabolismo o en rutas de señalización, por lo que puede pensarse que las proteínas son parte de la maquinaria funcional de la célula. Diferentes procesos regulatorios suceden a diferentes niveles de complejidad, de tal forma que los niveles de expresión de los componentes celulares pueden ser controlados (modificaciones al ADN o a las proteínas, por ejemplo).

Capítulo 2

Métodos de Enriquecimiento

2.1. Análisis de Significancia de Categorías Funcionales y de Expresión

En general, la mayoría de los estudios *ómicos* son comparativos, es decir, se contrastan respuestas a diferentes estímulos y se identifican los componentes celulares que cambian de manera estadísticamente significativa. Aunque es importante identificar componentes individuales asociados con las respuestas, la mayoría de los fenómenos biológicos y las enfermedades humanas ocurren a través relaciones funcionales, esto es, por medio de interacciones entre varios componentes [Barry, et al., 2005]. Los métodos de enriquecimiento surgen como herramientas que tienen como objetivo medir los niveles de actividad de estas funciones.

La característica principal de los métodos de enriquecimiento es el uso de conocimiento biológico existente para agrupar e interpretar resultados individuales. Lo anterior se logra al incorporar al análisis interpretativo conjuntos funcionales que relacionan los componentes individuales con funciones celulares [Varemo, et al., 2013]. Por ejemplo, en el cálculo de actividades de reguladores genéticos, los conjuntos funcionales son definidos como grupos de genes que son regulados conjuntamente [Garcia-Alonso, et al., 2017]. En principio, los conjuntos funcionales pueden ser construidos de un número ilimitado de formas, no obstante, usualmente se utilizan bases de datos de categorías funcionales o mediciones de interacción entre componentes celulares.

Dependiendo del tipo de prueba estadística que utilicen, los métodos de enriquecimiento pueden ser clasificados de diferentes maneras, sin embargo, los métodos de interés en este trabajo son aquellos que siguen la estructura de los análisis de significancia de categorías funcionales y de expresión o SAFE, por sus siglas en inglés (*Significance Analysis of Function and Expression*). El marco de referencia SAFE es un método generalizado de enriquecimiento *ab initio* de dos etapas que incorpora la in-

formación completa de análisis comparativos, ya sea usando el conjunto de p-values o el conjunto de sus estadísticos de prueba asociados (Fig. 2.1). Primero, estadísticos locales que miden la asociación entre los niveles de un componente celular y una respuesta son calculados. Después, estadísticos globales o *scores* de enriquecimiento son construidos como funciones de los estadísticos locales, de tal forma que pueda determinarse si los componentes celulares dentro un conjunto funcional tienen cambios más extremos que el resto de los componentes celulares medidos. Como último paso, mediante métodos de permutación, la significancia de cada estadístico global es calculada y eventualmente corregida por pruebas de hipótesis múltiples [Alhamdoosh, et al., 2017, Barry, et al., 2005, Varemo, et al., 2013].

En el ejemplo del cálculo de la actividad de reguladores genéticos en humanos durante una enfermedad, un análisis de enriquecimiento bajo el marco de SAFE seguiría la siguiente estructura (Fig 2.1) [Garcia-Alonso, et al., 2017]:

1. Medir los niveles de transcritos de todos los genes en algún tipo de célula de interés bajo las condiciones de estudio: Enfermo y sano.
2. Realizar un análisis comparativo entre ambas condiciones a nivel transcrito.
3. Utilizando una base de datos o el resultado de una inferencia computacional de interacciones, determinar los componentes que son controlados por los reguladores de interés.
4. Utilizar los p-values locales o sus estadísticos asociados para calcular estadísticos globales o en este caso, la actividad de cada regulador.
5. Calcular la significancia de cada estadístico global mediante análisis de permutación, ya sea de etiquetas de condición o de los conjuntos funcionales.
6. Corregir por pruebas de hipótesis múltiples e interpretar.

2.2. Limitaciones y desafíos

A pesar de ser un marco de referencia para todos los métodos de enriquecimiento más usados, SAFE aún carece de dos aspectos fundamentales en su definición:

- El primer aspecto está relacionado con la integración de datos provenientes de diferentes formas de calcular estadísticos globales. Dada la diversidad de formas que existen de calcular estadísticos globales y de sus diferentes desempeños dependiendo de la complejidad, escala y grado de ruido de los datos locales, es inadecuado confiar en los resultados de un solo método. Como solución a este problema se ha propuesto en diferentes aplicaciones el uso de resultados consenso

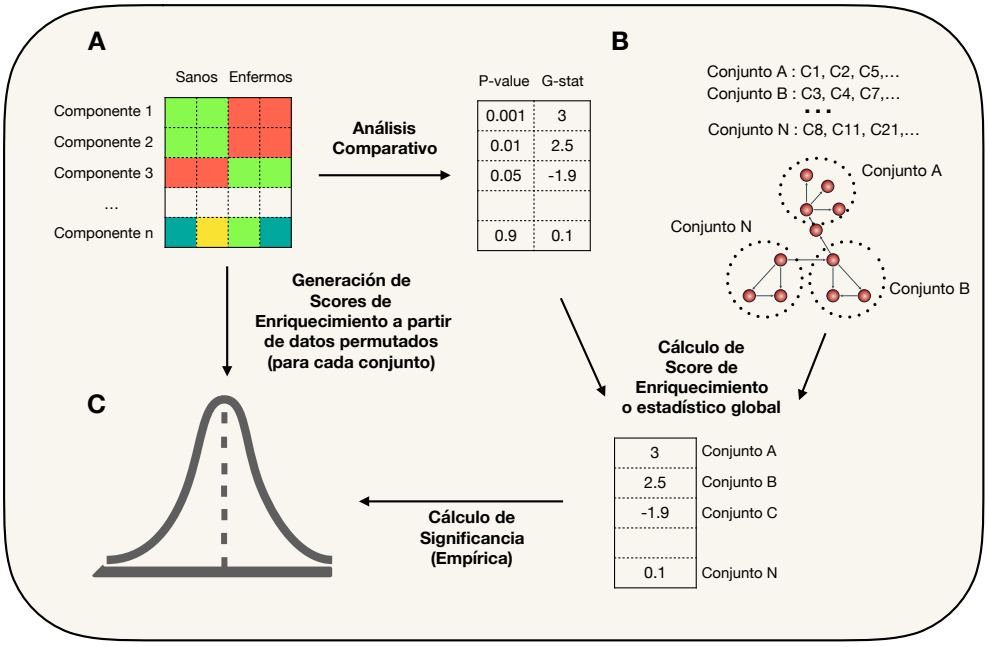

Figura 2.1: SAFE es un marco de trabajo que puede entenderse por 3 partes. El primer paso involucra el cálculo de estadísticos locales y su significancia a partir de un análisis comparativo (A). Luego, con los resultados generados, una colección de conjuntos funcionales y un método de enriquecimiento, se calculan estadísticos globales (*scores* de Enriquecimiento) (B). Para cada estadístico global se calcula su significancia empírica a partir de la permutación de etiquetas de muestras o de conjuntos funcionales, esto es, se repite el cálculo del *score* de Enriquecimiento por cada permutación y se calcula la proporción de estadísticos aleatorios que fueron igual o más extremos que el estadístico original (C).

que son calculados a partir de medias o medianas de "rankings" de estadísticos o p-values globales [Varemo, et al., 2013].

- El segundo aspecto se refiere al uso de la información contenida dentro de los conjuntos funcionales. Además de la lista de miembros dentro de un conjunto funcional, es necesario considerar los niveles de confianza de asociación que existen entre los conjuntos y los componentes, así como el grado de contribución de información (o de ruido) que cada componente contiene. También es posible definir direcciones de interacción, de tal forma que estadísticos globales direccionales puedan ser calculados. Aunque algunas herramientas bioinformáticas han propuesto de manera independiente diferentes formas de agregar información extra al cálculo de estadísticos globales, aún no existe una implementación generalizada de estas [Alvarez, et al., 2016].

Capítulo 3

Análisis Extendido de Conjuntos de Genes

En este trabajo se presenta *eGSA* (Análisis Extendido de Conjuntos de Genes), una herramienta bioinformática de análisis de enriquecimiento funcional que extiende el marco de trabajo de SAFE para poder corregir el cálculo de estadísticos globales y su significancia a partir del uso de conjuntos funcionales ponderados asociados a una covariable que evalúa confianza. *eGSA* calcula estadísticos globales, mejor conocidos como *scores* de enriquecimiento (ES), con diferentes métodos reportados en la literatura y permite la integración de sus resultados en un análisis de consenso (Fig. 3.1). *eGSA* representa un esfuerzo para integrar y extender el poder de análisis de 3 herramientas ampliamente usadas de manera independiente: PIANO [Varemo, et al. 2013], VIPER [Alvarez, et al. 2016] y IHW [Ignatiadis, et al. 2016]. El código fue desarrollado en *R* y está disponible en: *https://github.com/roramirezf/GSEAenhancement*. A lo largo de este capítulo se describirán los métodos estadísticos y las justificaciones de desarrollo de la librería.

3.1. Uso de pesos para la corrección de estadísticos globales y su significancia

Como se mencionó con anterioridad, los conjuntos funcionales capturan las relaciones que existen entre procesos celulares y componentes celulares individuales medidos por tecnologías de alto rendimiento. Las relaciones proceso-componente por construcción tienen tres cualidades que pueden considerarse en el cálculo de estadísticos globales y su significancia (Fig. 3.2):

- **Contribución**: Llamamos contribución a la cantidad de información relacionada a un proceso contenida en un componente. Dependiendo de la clase de proceso

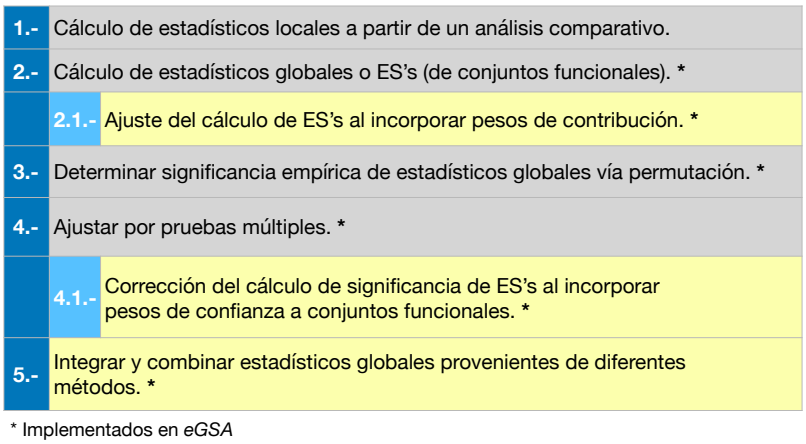

Figura 3.1: Descripción del marco de trabajo SAFE extendido: *eGSA* implementa una extensión de análisis funcionales en cualquier tipo de dato *ómico* al incorporar el uso de pesos de contribución y confianza en el cálculo de *scores* de enriquecimiento. En amarillo se muestran las extensiones y con una estrella las funciones implementadas en *eGSA*

funcional medido, es la interpretación específica que se le da a la contribución. En este trabajo se utilizaron datos de transcripción genética para inferir actividades de reguladores de transcripción [Garcia-Alonso, et al., 2017] y datos de fosforilación proteica para inferir actividades de kinasas (reguladores post-traduccionales de proteínas) [Hernandez-Armenta, et al., 2017]. En ambos casos el peso de contribución se refiere a: 1) El grado de control que ejerce el regulador a cada componente blanco o 2) el grado de "correlación" que existe entre la actividad de un regulador y el nivel de expresión de cada componente.

Los pesos de contribución modifican el cálculo de estadísticos globales, ya que permiten balancear el efecto que tienen los componentes. Esto es importante en estudios de datos masivos como los genómicos. Supongamos que un conjunto funcional está compuesto por 2 clases de componentes en partes iguales, aquellos que cambiaron de manera significativa en el análisis local y aquellos en los que no fue posible encontrar un cambio en el nivel de expresión. Si además consideramos que todos los que cambiaron tienen un mayor peso de contribución, entonces un método ponderado de estimación de ES puede identificar que el regulador está activo dado a que los genes blanco más susceptibles a su efecto cambiaron. Un método no ponderado, ante esta situación sería incapaz de capturar la actividad del regulador (Fig. 3.3). eGSA incorpora los pesos de contribución al marco de SAFE en 3 funciones de cálculo de Scores de enriquecimiento: Estadístico de la media [Varemo, et al., 2013], GSEA [Subramanian, et al., 2005] y aREA [Alvarez, et al., 2016] (descritas en posteriores secciones).

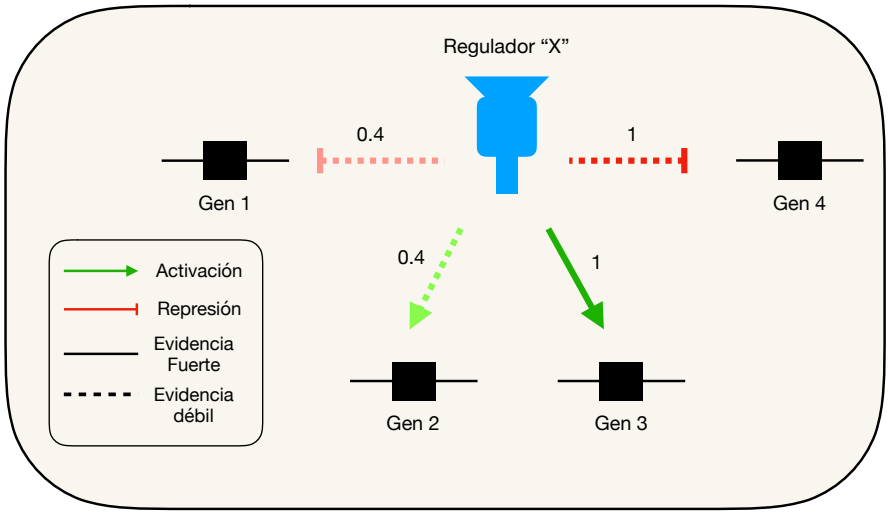

Figura 3.2: Ejemplo de información auxiliar en conjuntos funcionales: En esta figura se muestra un regulón, un conjunto funcional que describe un grupo de genes que son regulados conjuntamente. Como puede observarse, hay características de dirección, contribución y confianza para cada interacción regulador-gen. Un método de enriquecimiento ideal debiera utilizar toda esta información en el cálculo de estadísticos globales.

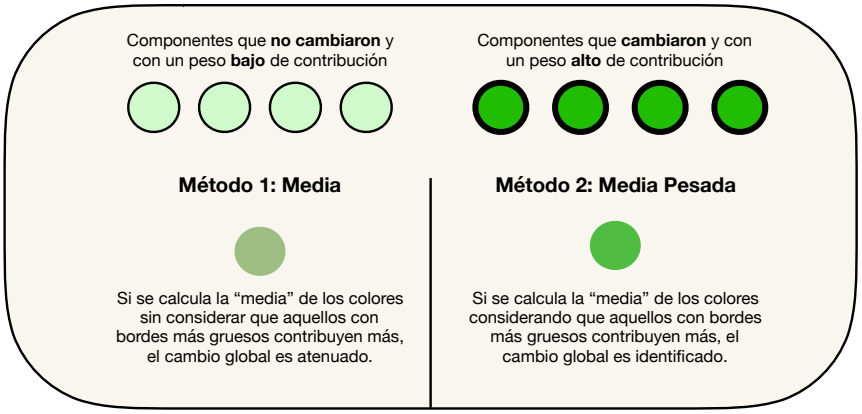

Figura 3.3: Efectos del uso de métodos de enriquecimiento pesados por contribución: Como se mencionó en el texto principal, los pesos de contribución pueden ser determinantes en el cálculo de estadísticos globales, especialmente porque permiten corregir subestimaciones o sobre-estimaciones.

- **Confianza**:Los conjuntos funcionales pueden ser construidos principalmente por curación manual de evidencia experimental, inferencias computacionales o ambas, por lo tanto es posible asignar un nivel de incertidumbre a cada relación proceso-componente. Los pesos de confianza están pensados para corregir el nivel de significancia de los estadísticos globales, especialmente en el contexto de la corrección por hipótesis múltiples. En este sentido, la corrección de *p-values* será más estricta para conjuntos funcionales de menor confianza, ya que es posible suponer que entre menor sea la confianza, mayor será su parecido a un conjunto azaroso. Los pesos de confianza son variables independientes a la hipótesis nula de la prueba de enriquecimiento, donde:

  H_0: Conjunto funcional no está enriquecido en el estudio comparativo (Proceso celular no está sucediendo)
  H_a: Conjunto funcional enriquecido (Proceso celular está sucediendo)

  en otras palabras, un conjunto funcional compuesto por componentes de alta confianza se espera que tenga un *p-value* más bajo que un un conjunto funcional compuesto por componentes de baja confianza, si y solo si, ambos conjuntos están enriquecidos. *eGSA* incorpora esta corrección al utilizar la estrategia de "Ponderación Independiente de Hipótesis", o IHW por sus siglas en inglés (Independent Hypothesis Weighting) [Ignatiadis, et al., 2016], descrita en posteriores secciones.

- **Dirección**: Es posible definir una dirección de interacción en los conjuntos funcionales, dada a la inherente cualidad direccional de los procesos celulares. Utilizando el ejemplo de los reguladores genéticos, algunos factores de transcripción tienen funciones duales de activación y represión. Algunos métodos de enriquecimiento pueden utilizar esta información en el cálculo de estadísticos globales [Alvarez, et al. 2016], sin embargo, en general se crean conjuntos funcionales por dirección [Varemo, et al. 2013]. *eGSA* no utiliza esta información en su implementación inicial.

3.2. Cálculo de estadísticos globales o *scores* de enriquecimiento

Para el cálculo de ES's, se implementaron 3 de los métodos más usados en aplicaciones de genómica funcional: El estadístico de la media [Varemo, et al. 2013], GSEA [Subramanian, et al. 2005] y aREA [Alvarez, et al., 2016] . Estos métodos difieren en la forma en la que integran la información de los estadísticos locales, sin embargo, comparten las estrategia de estimación de la significancia de estadísticos globales vía permutación, un aspecto necesario en la integración de resultados por análisis de consenso. Por otra parte, los 3 métodos pueden ser modificados para incluir un cálculo pesado de enriquecimiento, lo que permite el uso de pesos de contribución.

3.2.1. Cálculo de significancia empírica vía permutación

Las pruebas de hipótesis que están relacionadas a los análisis de enriquecimiento pueden ser de dos tipos:

- **Competitivas**: Una prueba es competitiva cuando se evalúa si un conjunto funcional está más enriquecido que una serie de conjuntos aleatorios [Varemo, et al. 2013].

  H_0: El valor de enriquecimiento de un conjunto funcional A es igual al de conjuntos aleatorios
  H_a: El valor de enriquecimiento de un conjunto funcional A es diferente al de conjuntos aleatorios

  En este caso, para cada conjunto funcional se genera una serie conjuntos aleatorios de su mismo tamaño (se sugieren al menos 1000 permutaciones [Varemo, et al. 2013]) y para cada conjunto aleatorio se calcula un ES. Entonces, el p-value empírico del estadístico global de un conjunto funcional es igual a la proporción de ES's aleatorios cuyo valor absoluto sea mayor o igual al valor absoluto del Score

de Enriquecimiento del conjunto funcional. Esta clase de prueba es la codificada en eGSA.

- **Contenidas en sí mismas**: Esta clase de pruebas evalúa la asociación de un conjunto funcional con el fenotipo, es decir, evalúa si un proceso celular está sucediendo en cierta condición, ignorando el resto de los componentes celulares.

  H_0: El valor de enriquecimiento de un conjunto funcional A no está relacionado con el cambio de condiciones.
  H_a: El valor de enriquecimiento de un conjunto funcional A está relacionado con el cambio de condiciones.

  El cálculo empírico de *p-values* para esta clase de pruebas, requiere la permutación de etiquetas de muestras previo al cálculo de estadísticos locales. Al igual que las pruebas competitivas, se requiere una gran cantidad de aleatorizaciones. Para cada permutación de etiquetas se calcula el ES de los conjuntos funcionales originales. El *p-value* del ES de un conjunto funcional es el número de estadísticos globales provenientes del mismo conjunto pero calculados con estadísticos locales de datos permutados, cuyo valor absoluto sea mayor o igual al calculado que el ES a probar. Dado a que esta clase de pruebas necesita de un número grande de muestras, no suele ser tan usado en aplicaciones generales, aunque sea más relevante para estudios biológicos por mantener las relaciones entre genes que por naturaleza no son independientes.

3.2.2. Métodos de enriquecimiento implementados en *eGSA*

A continuación se describirán los métodos utilizados en el cálculo de ES (o estadísticos globales):

- **Estadístico de la media:** Este método es el más sencillo de todos, ya que sólo requiere del cálculo de la media de los estadísticos locales de cada conjunto funcional. En su versión ponderada, es posible darle prioridad a ciertos componentes, dado a que el denominador en el cálculo es la suma de los pesos.

  Sea G un conjunto funcional, con n componentes los cuáles tienen un correspondiente estadístico t_i, y un peso de contribución w_i, $i \in [0, ..., n]$, se define el estadístico de la media como lo siguiente:

  $$ES_G = \frac{\sum_{i=1}^{n} w_i * t_i}{\sum_{i=1}^{n} w_i}$$

  En el caso de la versión no ponderada, $w_i = 1$, para toda i.

- **GSEA**: El análisis de enriquecimiento de conjuntos de genes o GSEA por sus siglas en inglés (*Gene Set Enrichment Analysis*) es el método de enriquecimiento más popular dentro del marco de trabajo SAFE. Utiliza una lista ordenada de estadísticos locales de los componentes, de tal forma que los más significativos se encuentran en los extremos de la lista (recordemos que los componentes pueden aumentar o disminuir sus niveles). Para cada conjunto funcional se calcula una suma total (*running sum*) sobre la lista ordenada empezando con el primer estadístico local y terminando con el último. La suma aumentará cada vez que se encuentre un estadístico local perteneciente a un componente del conjunto funcional y disminuirá en caso contrario. Los pesos de contribución y un parámetro de ajuste determinan la magnitud del aumento de la suma. El valor de enriquecimiento es la máxima desviación de 0 de la suma [Subramanian, et al., 2005] (Fig. 3.4).

  Matemáticamente, definimos a G como un conjunto funcional, con n componentes los cuáles tienen un correspondiente estadístico t_i, y un peso de contribución w_i, $i \in [0, ..., n]$. Definimos como $\bar{G}$, al conjunto de componentes fuera del conjunto G con $w_i = 0$. También definamos al conjunto total de componentes medidos y ordenados por t_i como N. Se define el estadístico global de G como:

  $$ES_G = max(abs(F_j^G - F_j^{\bar{G}})), j \in [1, ..., N]$$

  F siendo funciones de distribución acumuladas empíricas ponderadas.

  $$F_j^G = \frac{\sum_{s=1}^{j} |w_s|^\alpha \cdot \mathbb{1}_{\{componente_s \in G\}}}{\sum_{s=1}^{N} |w_s|^\alpha \cdot \mathbb{1}_{\{componente_s \in G\}}}$$

  $$F_j^{\bar{G}} = \frac{\sum_{s=1}^{j} \mathbb{1}_{\{componente_s \in \bar{G}\}}}{N - n}$$

  con α siendo un peso de escalamiento, usualmente igualado a 1. En el caso de la versión no pesada, $w_s = 1$, para toda s.

- **aREA:** El análisis de enriquecimiento analítico basado en posiciones, aREA, por sus siglas en inglés (*analytic Rank-based Enrichment Analysis*), es uno de los métodos más efectivos y modernos que se han desarrollado en el área de la genómica funcional. aREA mide un cambio global de las posiciones de los miembros de un conjunto funcional cuando son proyectados en una lista de componentes, ordenada por sus valores de estadísticos locales. El Score de enriquecimiento de un conjunto funcional es la media de de las posiciones cuantil-transformadas de sus componentes en la lista ordenada. Debido a que este texto no pretende ahondar

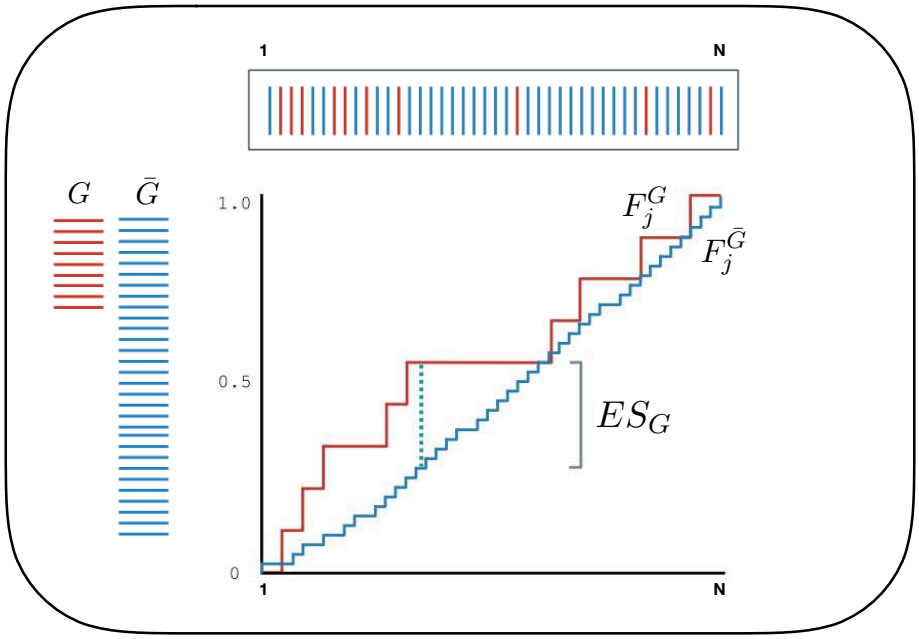

Figura 3.4: GSEA: Este método de enriquecimiento puede ser entendido como el cálculo de la máxima diferencia que existe entre 2 funciones de distribución acumuladas, una definida por los elementos de un conjunto funcional G y la otra por el resto de los componentes celulares medidos $\bar{G}$

en las justificaciones del método, la explicación de aREA se limitará a mencionar que integra la información de 2 cálculos de GSEA con diferentes ordenamientos de una misma lista de componentes y la información de contribución. [Alvarez, et al., 2016]

3.2.3. Ponderación Independiente de Hipótesis como método de corrección de significancia

En estudios ómicos comparativos, debido a la gran cantidad de pruebas de hipótesis que se realizan, es mandatorio corregir por pruebas múltiples. Un procedimiento de pruebas múltiples rechazará R hipótesis de un conjunto de m hipótesis distintas $H_1, \ldots, H_m$. V de esas hipótesis son en realidad nulas, por lo que se cometerán V errores tipo 1. Errores generalizados del tipo 1 usualmente son definidas como esperanzas de funciones de V y R. Uno de los primeros acercamientos de evaluación múltiple se enfoca en controlar la tasa de error de Familia (*Family Wise Error, FWER*), definida como $Pr[V \geq 1]$ a un nivel α establecido. En muchas aplicaciones, en especial aquellas de cualidad exploratoria, FWER resulta muy conservador, por lo que se opta por utilizar la tasa de falsos descubrimientos (*False Discovery Rate, FDR*). La *FDR* está definida como el valor esperado de la proporción de falsos descubrimientos:

$$FDR = E[\frac{V}{R}]$$

El procedimiento Benjamini-Hochberg permite fácilmente controlar la FDR a un nivel especificado α. El método funciona ordenando los *p-values* $P_{(1)} \leq \ldots \leq P_{(m)}$ para luego rechazar todas las hipótesis con un p-value $\leq P_{k*}$, donde $k* = max\{k|P_{(k)} \leq k\alpha/m, k \geq 1\}$ [Benjamini, et al., 1995, Ignatiadis, et al., 2016].

La Ponderación Independiente de Hipótesis o IHW por sus siglas en inglés (Independent Hypothesis Weighting), es un método de corrección de *p-values* que complementa la corrección de pruebas múltiples. El objetivo de IHW es asignar un peso WH_i de ponderación a cada hipótesis de tal forma que algunas sean prioridad. IHW asigna pesos a *p-values*, usando una covariable informativa continua o categórica que refleja la información de las propiedades estadísticas de las hipótesis y al mismo tiempo se mantiene independiente a los *p-values* bajo la hipótesis nula. IHW en el mejor de los casos aumenta la potencia de las pruebas y sigue controlando los errores tipo 1 si se combina con otro método de corrección de hipótesis múltiples. En el peor de los casos en el que los pesos asignados no sean completamente correctos, IHW disminuirá sutilmente la potencia de las pruebas a comparación de una corrección sin pesos.

La implementación básica de IHW es la siguiente. Primero, se dividen las pruebas entre grupos basados en una covariable, a cada grupo se le asocia un peso de tal manera que todas las hipótesis dentro de un grupo se les asigne el mismo peso. Para cada posible selección de pesos, se aplica un procedimiento *Benjamini-Hochberg* [Benjamini,

et al., 1995] ponderado a un nivel alpha y se calculan el número de descubrimientos. Finalmente se seleccionan los pesos que generan el mayor número de descubrimientos. IHW utiliza métodos de aprendizaje estadístico con los que regulariza y optimiza el cálculo de parámetros. El algoritmo final va más allá de la discusión de este trabajo y se recomienda consultar Ignatiadis, et al., si es de interés conocer las pruebas matemáticas del método. IHW necesita de un proceso de división de datos para poder entrenar los estimadores de pesos, por lo que al menos 2000 pruebas tienen que corregirse para que el rendimiento sea el óptimo.

Es mandatorio que antes de utilizar la correción por IHW, se evalúe si la covariable de confianza definida sigue las suposiciones del método. La independencia con la hipótesis nula puede verificarse si en los histogramas de *p-values* separados por los niveles de la covariable de confianza se observan colas uniformes. Por otra parte, las gráficas de funciones de distribución acumuladas de los*p-values* separadas por los niveles de la covariable de confianza, dan indicios del poder de la covariable para reflejar información estadística de las hipótesis. Curvas claramente separadas y sin cruzamientos son señales suficientes para continuar con la corrección.

Para eGSA, el uso de IHW tiene 2 propósitos:

1. En estudios con un número pequeño de conjuntos funcionales ($\leq$ 2000), permite darle prioridad a pruebas de conjuntos funcionales de alta confianza (como los determinados por evidencia experimental) sin disminuir bruscamente el número de hipótesis rechazadas.

2. En análisis masivos, con un número grande de conjuntos funcionales ($\geq$2000), además de lo anterior, se incrementa la potencia de las pruebas, por lo que hay un número mayor de hipótesis rechazadas.

Capítulo 4

Uso de Pesos de Contribución en Análisis Funcionales de una Colección de Referencia de Datos de Actividades de Kinasas

4.1. Datos de referencia de actividades de kinasas

Para probar la efectividad del uso de pesos de contribución en el mejoramiento del cálculo de estadísticos globales, se utilizaron los datos de referencia de fosforilación proteica provistos por Hernandez-Armenta, et al.. Estos datos consisten en 91 experimentos de perturbación que miden la actividad de kinasas específicas. Los experimentos consisten en el aumento o decrecimiento de la actividad regulatoria de kinasas específicas y la subsecuente medición de niveles de fosforilación de varias proteínas (sitio específico). Básicamente los experimentos "desconectan" una función y luego miden qué sucede en una capa de complejidad biológica. Los métodos de enriquecimiento, en este caso, permiten medir los niveles de actividad de reguladores de proteínas.

Los conjuntos funcionales fueron recuperados de *PhosphoSitePlus* [Hornbeck, at al., 2012] y están definidos como relaciones kinasa-sustratos, esto es, cada kinasa tiene asignada un grupo de "sitios" en los que puede ejercer su actividad reguladora. Para algunos sitios es posible calcular un puntaje de afinidad a partir de una matriz de pesos específica de la kinasa, y refleja en términos simples qué tan probable es que la kinasa regule ese sitio: a mayor afinidad, mayor probabilidad. Para cada kinasa, sólo se seleccionaron los sitios a los que un puntaje de afinidad pudo ser calculado. En re-

sumen, cada conjunto funcional es una kinasa diferente junto a sus sitios ponderados, y la matriz de datos contiene el "nivel" de regulación de los diferentes sitios en varias condiciones.

Cada columna de la matriz de datos representa una comparación entre un estado perturbado y un estado nativo, por lo que cada elemento en ella representa el logaritmo base 2 del cambio de los niveles de fosforilación de un sitio específico (*log fold change*). Dada la construcción de estos datos de referencia, es posible asignar un verdadero positivo a cada columna, es decir, si la columna i representa la pertubación de la kinasa "X", entonces el conjunto funcional "X" es el verdadero positivo de esa columna, dado a que se espera observar cambios extremos en los elementos de ese conjunto. A cada uno de esos pares perturbación-conjunto se le denomina par positivo. Se recuperaron todos los pares positivos en los que el conjunto funcional tuviera al menos 4 sitios reportados y ese fue el conjunto de datos final a analizar: 76 pares positivos, que abarcan 54 experimentos únicos y 14 conjuntos funcionales (kinasas). Cada conjunto funcional tiene un análogo "no ponderado", el cuál es idéntico a nivel componente, sin embargo, difiere en los puntajes de afinidad, los cuales se igualaron a 1 para todos los sitios.

4.2. Pesos de contribución mejoran ligeramente la identificación de procesos celulares

Se calculó el ES de cada conjunto funcional de cada par positivo utilizando su correspondiente experimento y los 3 métodos mencionados anteriormente en sus versiones ponderadas y no ponderadas. A la par, se calculó el ES de un grupo de 76 pares aleatorios, que son análogos en tamaño a los pares positivos. Este proceso se repitió 100 veces, tanto para los conjuntos ponderados, como para los no ponderados. Cada conjunto funcional aleatorio es un par análogo a un par positivo, el cual está compuesto por un número igual de componentes que su correspondiente par positivo, pero estos componentes son tomados al azar de la lista de sitios medidos en el respectivo experimento del par. Para generar las versiones ponderadas, se mantuvieron los pesos del par positivo original pero repartidos aleatoriamente en los nuevos componentes aleatorios.

Este proceso generó 100 listas de ES's por tipo de análisis (ponderado y no ponderado), que contienen resultados de los verdaderos positivos (76 estadísticos globales) y resultados de una iteración de la aleatorización (76 estadísticos aleatorios). Cada lista fue ordenada de mayor a menor y cada elemento de la lista fue etiquetado dependiendo de su par de origen. De cada lista se calculó una curva de precisión y exhaustividad en la posición (*Precision-Recall curve at rank* [Garcia-Alonso, et al.,2017]) como se explica a continuación.

Sea una lista ordenada con n elementos que pueden ser clasificados en dos grupos H

y $\bar{H}$, los cuáles representan los grupos de verdaderos positivos y verdaderos negativos, por cada posición i definimos precisión y exhaustividad como:

$$Precision_i = \frac{\sum_1^i \mathbb{1}_{elemento_i \in H}}{n}$$

$$Exhaustividad_i = \frac{\sum_1^i \mathbb{1}_{elemento_i \in H}}{sum_1^n \mathbb{1}_{elemento_i \in H}}$$

Esta curva nos permite evaluar la efectividad del método en identificar verdaderos positivos, ya que se espera que los ES de los pares positivos sean mayores a los de los pares aleatorios si es que los métodos funcionan correctamente. De cada una de las 200 curvas generadas se recuperó el área debajo de la curva (AUC) y la precisión cuando la exhaustividad era del 0.5 (PREC).

En la figura 4.1 se muestran las comparaciones de los valores de AUC que se realizaron entre los análisis ponderados y no ponderados, y entre los diferentes métodos de enriquecimiento utilizados. Al igual que en Hernandez-Armenta, et al., se observó que las versiones pesadas de los métodos tienen un ligero mejor desempeño que los no ponderados (Prueba t de Student de dos colas, valor de corte: 0.05, Tabla 4.1). A diferencia del estudio mencionado anteriormente, este trabajo incorpora las pruebas de desempeño para 2 nuevos métodos: aREA y el estadístico de la media. Los resultados para GSEA replican lo reportado por Hernandez-Armenta, et al.. aREA es un método que ha mostrado un nivel de rendimiento superior en estudios funcionales de actividades regulatorias [Alvarez, et al. 2016], por lo que es esperado que sus valores de AUC sean superiores a los de los estadísticos de media y GSEA. En contraste, los valores de PREC fueron similares entre los análisis ponderados y no ponderados, a excepción de los obtenidos con los estadísticos de media (Prueba t de Student de dos colas, valor de corte: 0.05, Tabla 4.2, Fig. 4.2).

Lo anterior muestra que el uso de los pesos de contribución puede aumentar el poder de identificación de actividades de conjuntos funcionales, no obstante, es necesario aclarar que en estudios individuales el mayor beneficio del uso de pesos de contribución es la corrección de subestimaciones o sobre-estimaciones de estadísticos. Esto es relevante si se utilizan los métodos de enriquecimiento no solo en su fase interpretativa, sino también como métodos de reducción de dimensiones que permiten resumir la información contenida en miles de componentes celulares en cientos de procesos celulares.

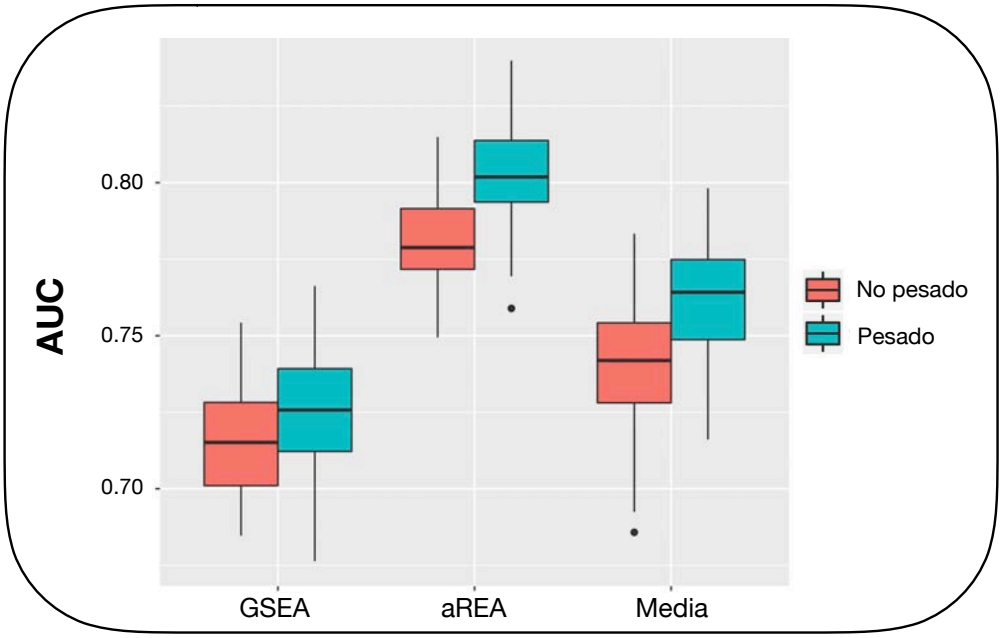

Figura 4.1: Comparación de las AUC's de el conjunto de listas ordenadas de pares positivos y pares aleatorios. Se observa una ligera tendencia de los métodos pesados a tener valores de AUC mayores a los de los métodos no ponderados. Esto significa que los métodos ponderados tienen mayor poder en la identificación de verdaderas actividades de conjuntos funcionales

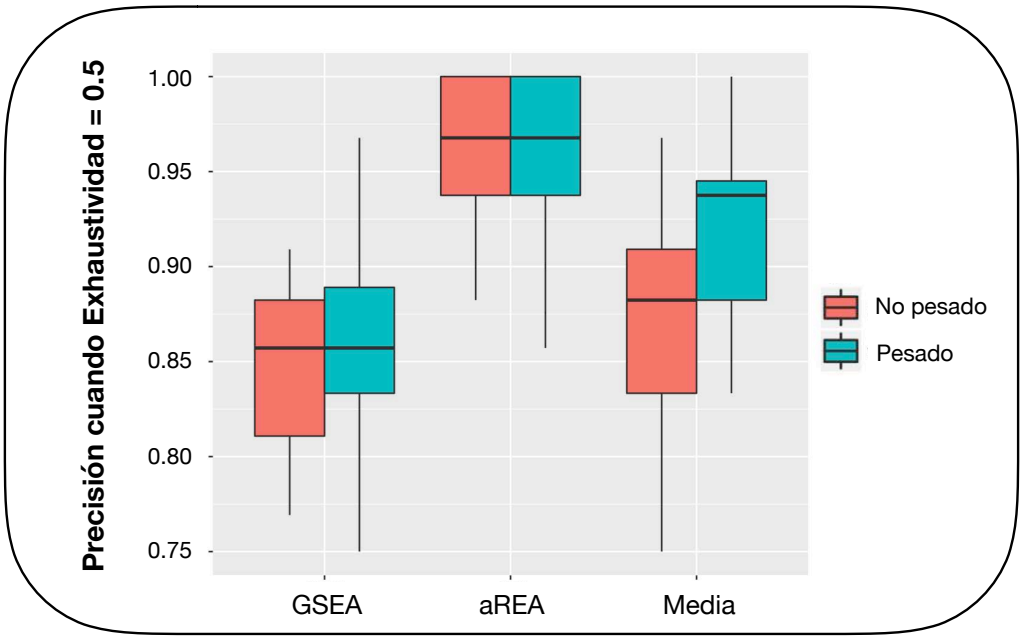

Figura 4.2: Comparación de las PREC's del conjunto de listas ordenadas de pares positivos y pares aleatorios. A diferencia de las AUC's, sólo es posible observar diferencia entre métodos ponderados y no ponderados, cuando los estadísticos globales son calculados con medias

Método	Estadístico t	p-value	Medición
Media	5.433538	3.126218e-07	AUC
aREA	7.252438	4.882442e-11	AUC
GSEA	2.648367	9.223513e-03	AUC
Media	5.144082	1.108096e-06	PREC
aREA	1.241848	2.167874e-01	PREC
GSEA	1.192620	2.354593e-01	PREC

Tabla 4.1: Comparación de AUC entre métodos ponderados y no ponderados:
Prueba t de Student a dos colas

Capítulo 5

La Ponderación Independiente de Hipótesis Puede Aumentar el Número de Descubrimientos sin Afectar el Control del Error tipo 1

5.1. Datos de referencia de actividades de factores de transcripción

Como se mencionó con anterioridad, IHW tiene un diferente alcance de corrección de *p-values* dependiendo de las dimensiones del estudio. Con este análisis se tiene el objetivo de mostrar cómo es que IHW aumenta el número de hipótesis rechazadas en estudios con miles de pruebas por corregir. Se utilizó un conjunto de datos de referencia de expresión genética, que forman parte de la continuación del trabajo, aún no publicado, de Garcia-Alonso, et al.. Estos datos representan una colección de 127 experimentos de perturbación de actividades de factores de transcripción (reguladores de la expresión genética). Cada columna contiene estadísticos t moderados provenientes de un análisis comparativo, entre una condición basal y la perturbación experimental, de la expresión de 22,066 genes. Valores altos de los estadísticos, no importando si son positivos o negativos, representan cambios de expresión significativos.

Los conjuntos funcionales utilizados en este estudio se llaman regulones y están definidos como grupos de genes que son regulados conjuntamente por un factor de transcripción, una proteína con función regulatoria. Se recuperaron los regulones de la

base de datos *RegNetwork* [Liu, et al. 2015]. Esta base de datos proporciona interacciones de alta, media y baja confianza. Se ignoraron las interacciones de media confianza y se generaron regulones exclusivamente compuestos por interacciones de alta y de baja confianza. En total 1685 regulones pudieron construirse, de los cuáles se filtraron 612 por tener menos de 4 componentes. Luego de esta serie de filtros, se mantuvieron 1073 regulones, 183 de alta confianza y 890 de baja confianza.

5.2. IHW corrige los resultados de los métodos de enriquecimiento al priorizar hipótesis de alta confianza

Para cada regulón se calculó su ES y su significancia en cada experimento utilizando solamente aREA por cuestiones de poder de cómputo. Lo anterior generó 134,911 *p-values*. En la figura 5.1 es posible observar la distribución de los *p-values* sin corregir de los ES calculados (Fig. 5.1-A). Esta distribución tiene una cola uniforme, cumpliendo uno de los supuestos necesarios para la corrección por pruebas múltiples [Ignatiadis, et al., 2016]. Con el objetivo de probar que la variable de confianza asociada a los *p-values* obtenidos, está relacionada al rechazo de la hipótesis nula, se segregaron los *p-values* dependiendo del tipo de interacciones dentro del regulón evaluado (Fig. 5.1-B), y se construyeron funciones de distribución acumulada para cada grupo (Fig. 5.1-C). Ambas gráficas cumplen con las condiciones especificadas por Ignatiadis, et al., por lo tanto se procedió a usar IHW para calcular los pesos de corrección (se especificó que la lista de *p-values* se separara en 5 grupos para entrenar el algoritmo).

IHW entrena su algoritmo con diferentes listas de *p-values* con la finalidad de evitar la sobre estimación de los pesos de corrección [Ignatiadis, et al., 2016]. Por cada grupo calcula un peso para cada nivel de la covariable usada, en este caso, un peso de corrección para los *p-values* de pruebas de conjuntos funcionales de baja confianza y otro para los conjuntos de confianza alta (Fig. 5.2). En general, se espera que para cada grupo de *p-values* la estimación de los pesos sea similar dentro del mismo nivel de la covariable.

Sea W_i un peso asignado a un p-value p_i, se define un p-value ponderado p_i^* como:

$$p_i^* = \frac{p_i}{W_i}$$

Es por lo anterior, que pesos mayores a 1 dan prioridad a las hipótesis, mientras que pesos menores a 1, penalizan a las hipótesis.

Los *p-values* corregidos por IHW en seguida fueron nuevamente corregidos, pero esta vez utilizando un procedimiento Benjamini-Hochberg agrupado [Hu, et al., 2010], que en términos prácticos es un procedimiento BH pero que utiliza *p-values* corregidos. En la tabla 5.1 es posible observar el número de hipótesis a probar en este análisis y el

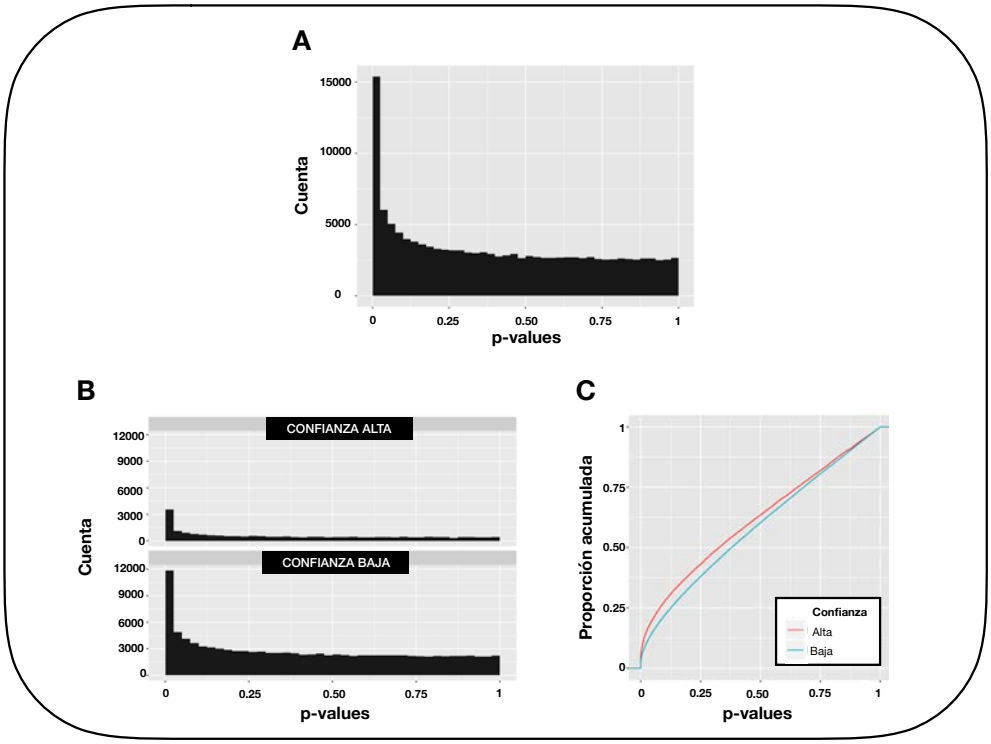

Figura 5.1: Prueba de suposiciones de la covariable de confianza: El histograma de los *p-values* a corregir muestra una cola pesada, una característica necesaria para los procedimientos de corrección de pruebas multiples (A). Los histogramas de *p-values* separados por niveles de covariable de confianza también tienen colas pesadas, por lo que se puede inferir que la covariable es independiente a la hipótesis nula (B). Las funciones de distribución acumulada de los *p-values* separados por los niveles de la covariable de confianza, confirman su utilidad para la corrección con IHW (C)

número de hipótesis rechazadas ya sea con *p-values* sin corregir, *p-values* corregidos con un procedimiento Benjamini-Hochberg (BH), *p-values* corregidos por IHW o *p-values* corregidos por ambos métodos (en todos los casos el valor de rechazo fue de 0.1). La corrección por IHW rechazó un número similar de hipótesis a comparación de los *p-values* crudos, aún luego de haberle dado prioridad a hipótesis provenientes de la evaluación de conjuntos funcionales de alta confianza (que son un número considerablemente menor que los conjuntos funcionales de baja confianza). En cambio, luego del control de la FDR por BH, IHW aumentó el número de hipótesis rechazadas, mostrando que la correcta inferencia de pesos de penalización puede incrementar el poder exploratorio de las pruebas múltiples. Esta aparente contradicción sucede porque un mayor número de pruebas son penalizadas por IHW que las que son favorecidas (los conjuntos funcionales de baja confianza son muchos más que los de alta confianza) , sin embargo, los *p-values* que fueron corregidos con pesos mayores a 1 se hicieron mucho más pequeños. Por lo anterior, muchos menos *p-values* son corregidos bruscamente en el control de la FDR.

Las condiciones observadas no forzosamente tienen que mantenerse en los resultados de los otros 2 métodos de enriquecimiento. Es por eso, que es necesario verificar las suposiciones para cada conjunto de resultados.

Confianza	No. total de pruebas	Pruebas rechazadas (PR)	PR luego de BH	PR luego de IHW	PR luego de BH +IHW
Baja	111761	24543	6805	21919	6346
Alta	23150	6404	2206	8646	2766
Total	134911	30947	9011	30565	9112

Tabla 5.1: Efectos de la corrección por IHW - Un mayor de número de hipótesis son rechazadas cuando el procedimiento BH es complementado con la corrección por IHW, aumentando el poder de detección sin dejar de controlar la FDR

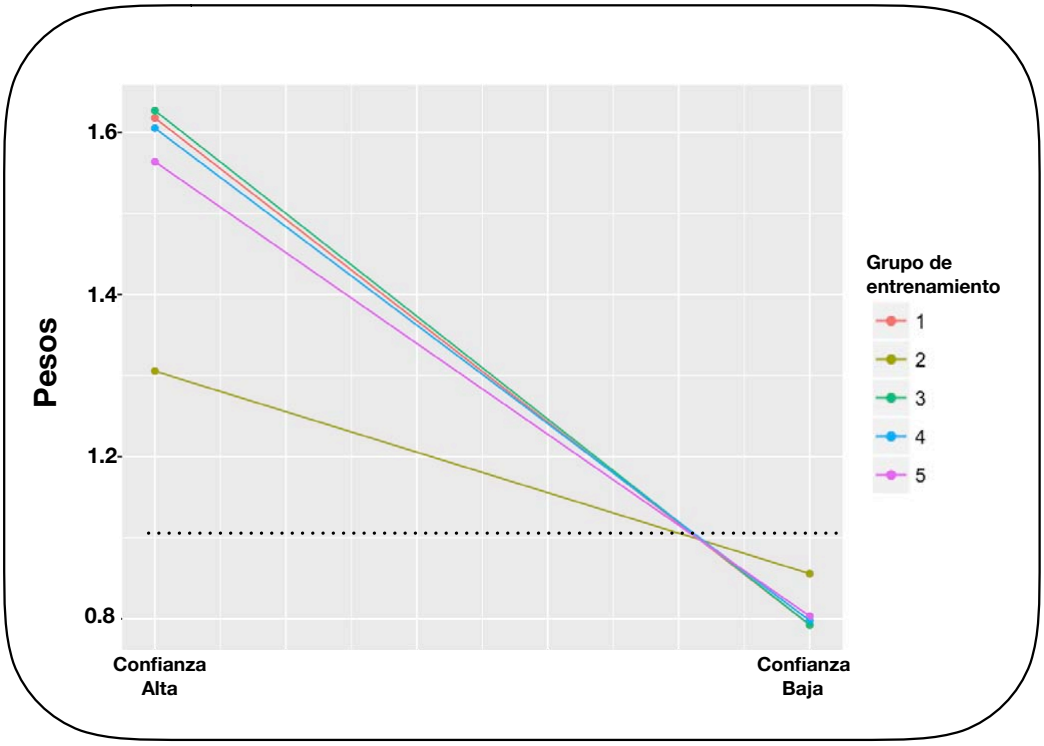

Figura 5.2: Estimación de pesos de corrección por IHW: Por cada grupo de entrenamiento se calculó un peso de corrección por nivel de la covariable de confianza. En este caso, las hipótesis etiquetadas con un valor de confianza alta serán prioridad.

Capítulo 6

Caso de Estudio

6.1. *eGSA* realiza análisis de enriquecimiento consenso

Al igual que PIANO [Varemo, et al. 2013], *eGSA* permite integrar los resultados provenientes de varios métodos de enriquecimiento en un análisis consenso de *p-values* y estadísticos globales. El procedimiento al igual que en PIANO, funciona de la siguiente manera:

1. Cada lista de resultados es ordenada por *p-values*, de tal manera que los conjuntos funcionales con *p-values* más pequeños estén en los primeros lugares de la lista.
2. Se recupera la posición en la que se ubican los conjuntos funcionales en cada lista de resultados y se calcula por cada conjunto funcional la posición media (mediana) y el *p-value* medio (mediano) del conjunto de resultados.
3. Se ordenan nuevamente los conjuntos funcionales, pero esta vez utilizando su posición media (mediana). Se seleccionan valores de corte de la lista ordenada por posición o *p-value* consenso.

Esta forma de realizar un análisis consenso permite identificar rápidamente conjuntos funcionales que fueron detectados por varios métodos a la vez. No obstante, esta aproximación carece de un estadístico el cuál pueda ser evaluado con una prueba estadística.

6.2. Identificación de actividades regulatorias genéticas en células cancerígenas

Para probar la practicidad y utilidad de *eGSA* como herramienta bioinformática en estudios de menor dimensión, se decidió analizar un conjunto de datos de expresión

genética de células cancerígenas, provistas por el grupo del Prof. Thorsten Cramer, investigador del Hospital de la *RWTH Aachen University*. Los datos estudian el efecto de la supresión de actividad del gen HIF-1α en diferentes condiciones ambientales, en las células modelo de cáncer de hígado *HEP-G2*. HIF-1α es un gen que controla las respuestas a ambientes hipóxicos, o sea, a condiciones de bajo nivel de oxígeno [Majmundar, et al., 2010]. Se ha mostrado que este gen también está relacionado a la resistencia del cáncer a diferentes tratamientos, por lo que es de utilidad conocer los procesos que controla este factor de transcripción [Majmundar, et al., 2010].

El objetivo de este estudio es la identificación de actividades de reguladores genéticos con un análisis consenso en los datos de expresión genética mencionados arriba. Como estadísticos locales se utilizaron 19,722 estadísticos t moderados provenientes del análisis diferencial de expresión de genes de células *HEP-G2* en condiciones normales y células *HEP-G2* con una represión de la actividad de HIF-1α. Los conjuntos funcionales recuperados, representan la colección de regulones de *RegNetwork* [Liu, et al., 2015], separados por conjuntos de alta y baja confianza, como se hizo con anterioridad. De los 986 regulones recuperados, 823 fueron de baja confianza y 163 de alta confianza.

eGSA se utilizó para calcular los scores de enriquecimiento y su significancia (1000 permutaciones), con los 3 métodos implementados: Estadístico de la media, *GSEA* y *aREA*. A excepción de *aREA*, todos los métodos fueron no direccionales (se utilizó el valor absoluto de los estadísticos locales), debido a que la información de dirección de interacción no está disponible en *RegNetwork*. aREA es un método robusto a la ausencia de dirección por lo que, en este caso, no fue necesario modificar los estadísticos globales.

Por cada método se generaron los histogramas y los gráficos de las funciones de distribución acumulada de los *p-values* separados por los niveles de confianza, para evaluar los supuestos de IHW (Fig. 6.1). En todos los casos, los supuestos se cumplen, por lo que se corrigieron los p-values utilizando como covariable de corrección las etiquetas de confianza de interacción. En este caso se decidió no controlar la FDR, debido a que el análisis interpretativo de consenso que se hará de los resultados no necesariamente requiere del rechazo de hipótesis. Si se fueran a seleccionar conjuntos funcionales significativos de cada método de manera individual, entonces se sugiere corregir utilizando un procedimiento BH, aunque también es necesario prestarle atención a la violación del supuesto de independencia en algunas situaciones en las que los conjuntos funcionales no esten definidos de la mejor forma. Lo anterior se discutirá en secciones posteriores.

Se realizó un análisis consenso del resultado de los 3 métodos de enriquecimiento, utilizando los *p-values* corregido por *IHW* como variable de ordenamiento, y el valor de la media de la posición y *p-value* consenso como resultados de salida. En la figura 6.2 puede observarse el resultado de seleccionar a los 50 primeros conjuntos funcionales de este ordenamiento, los cuáles tienen un *p-value* medio menor a 0.01. El identificar la actividad de los conjuntos .^{EP}AS1ƨ "HIF1A", nos permite corroborar que el método

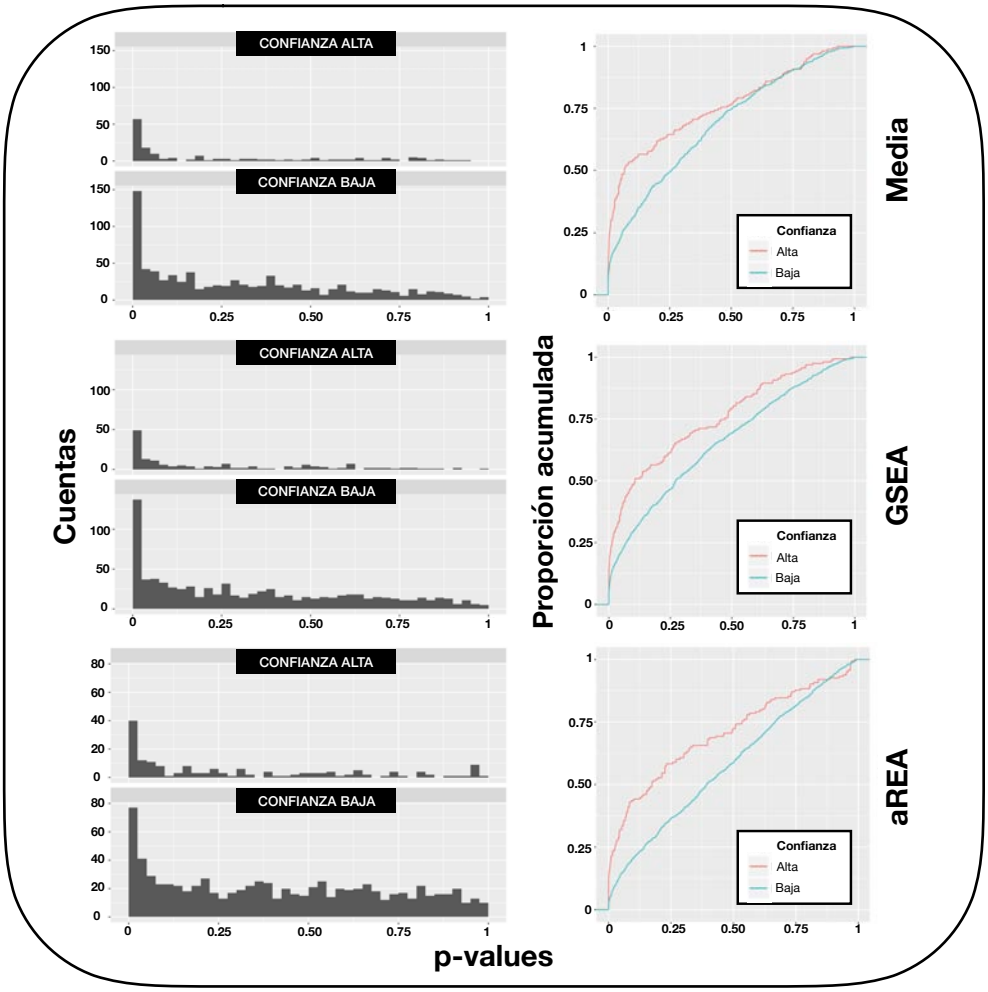

Figura 6.1: Prueba de suposiciones de la covarible de confianza para los 3 métodos: Los histogramas de *p-values*, separados por los niveles de la covariable de confianza, tienen colas pesadas, por lo que se puede inferir que la covariable es independiente a la hipótesis nula en todos los métodos. Las funciones de distribución acumulada de los *p-values*, separados por los niveles de la covariable de confianza, confirman su utilidad para la corrección con *IHW* al mantenerse separadas en casi todos los valores de los *p-values*

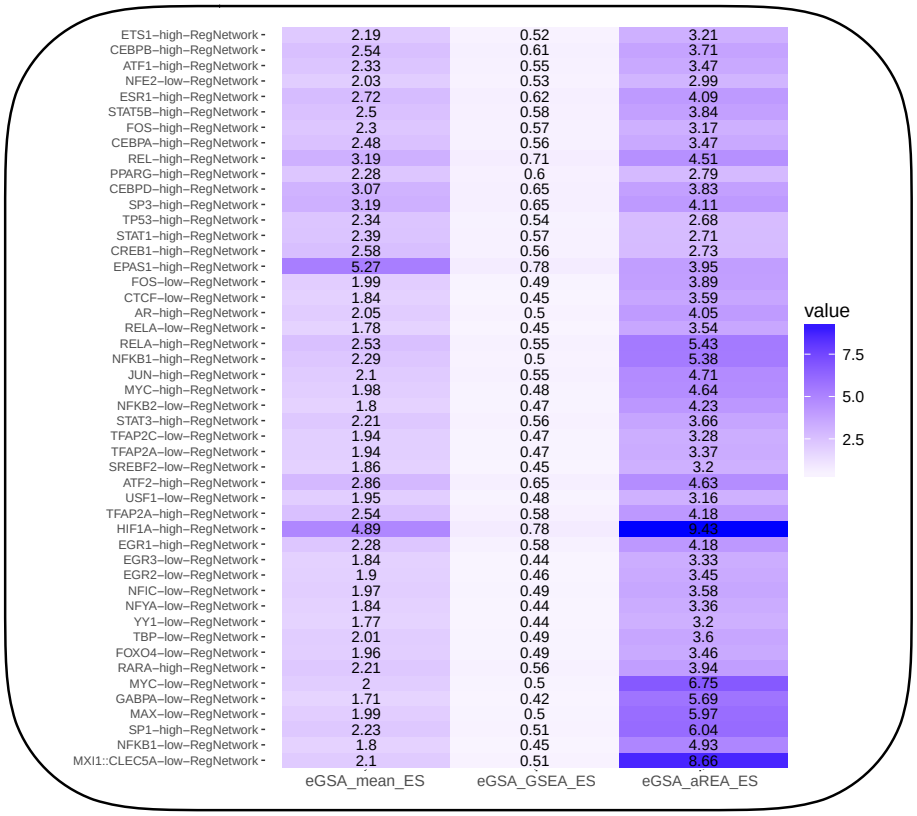

Figura 6.2: Mejores 50 conjuntos funcionales determinados por el análisis consenso: El valor de cada columna es el ES obtenido en cada método

es efectivo, debido a que estos conjuntos son los verdaderos positivos del experimento de perturbación. En términos generales, los procesos identificados coinciden con lo reportado en la literatura [Majmundar, et al., 2010].

No es objetivo de este trabajo ahondar en los resultados observados, sin embargo, sí se mencionará que el análisis de enriquecimiento consenso auxilia al investigador en la generación de hipótesis. La meta principal de esta herramienta es generar estudios exploratorios de auxilio que de cierta forma resuman la cantidad de información contenida en miles de componentes en grupos de procesos celulares que pueden interpretarse de manera más directa.

Referencias

[1] Alhamdoosh, M., Ng, M., Wilson, N. J., Sheridan, J. M., Huynh, H., Wilson, M. J., and Ritchie, M. E. (2017). Combining multiple tools outperforms individual methods in gene set enrichment analyses. *Bioinformatics (Oxford, England)*, 33(3):414–424.

[2] Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Hilda Ye, B., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature Genetics*, 48(8):838–847.

[3] Barry, W. T., Nobel, A. B., and Wright, F. A. (2005). Significance analysis of functional categories in gene expression studies: A structured permutation approach. *Bioinformatics*, 21(9):1943–1949.

[4] Benjamini, Yoav; Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

[5] Dai, H., Leeder, J. S., and Cui, Y. (2014). A modified generalized fisher method for combining probabilities from dependent tests. *Frontiers in Genetics*, 5(FEB):1–10.

[6] Garcia-Alonso, L., Iorio, F., Matchan, A., Fonseca, N., Jaaks, P., Peat, G., Pignatelli, M., Falcone, F., Benes, C. H., Dunham, I., Bignell, G., McDade, S. S., Garnett, M. J., and Saez-Rodriguez, J. (2018). Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Research*, 78(3):769–780.

[7] Hernandez-Armenta, C., Ochoa, D., Gonçalves, E., Saez-Rodriguez, J., and Beltrao, P. (2017). Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics*, 33(12):1845–1851.

[8] Ibarra-Arellano, M. A., Campos-González, A. I., Treviño-Quintanilla, L. G., Tauch, A., and Freyre-González, J. A. (2016). Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria. *Database : the journal of biological databases and curation*, 2016:1–16.

[9] Ignatiadis, N., Klaus, B., Zaugg, J. B., and Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, 13(7):577–580.

[10] Joyce, A. R. and Palsson, B. Ø. (2006). The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology*, 7(3):198–210.

[11] Liu, Z.-P., Wu, C., Miao, H., and Wu, H. (2015). RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database*, 2015:bav095.

[12] Majmundar, A. J., Wong, W. J., and Simon, M. C. (2010). Hypoxia-Inducible Factors and the Response to Hypoxic Stress. *Molecular Cell*, 40(2):294–309.

[13] Roadmap Epigenomics Consortium, Kundaje, A., and Meuleman, Wouter, e. a. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–329.

[14] Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M. J., Blüthgen, N., and Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications*, 9(1).

[15] Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: Guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods*, 13(12):966–967.

[16] Väremo, L., Nielsen, J., and Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, 41(8):4378–4391.