Babelomics 5

Esta práctica incluye varios ejercicios para trabajar con los diferentes módulos de análisis que nos ofrece la suite Babelomics 5: http://babelomics.bioinfo.cipf.es/

Para comenzar, os sugerimos que os creéis un usuario en la herramienta, que nos permitirá almacenar los jobs que lancemos y visualizar los resultados en cualquier momento y desde cualquier ubicación.

Si tienes alguna duda: fgarcia@cipf.es

EJERCICIO 1.

Análisis de datos de RNA-Seq: clasificación no supervisada o clustering.

Objetivo.
Detectar grupos homogéneos de sujetos en función de su perfil transcriptómico.

Datos.
Disponemos de conteos no normalizados en 29.405 genes para un total de 10 individuos. Estos datos se obtuvieron tras aplicar un análisis primario que incluyó la evaluación de calidad de las secuencias, mapeo y cuantificación de la expresión a nivel de gen.

Plan de trabajo.

1. Abre el fichero de datos “rnaseq_clustering.txt” con una hoja de cálculo e inspecciona su contenido. Habrá tantas columnas como sujetos y tantas filas como genes.
2. Sube este fichero txt en Babelomics desde el menú “Upload”. Tendremos que indicar el tipo de dato que subimos: “Data matrix expression”. En este link se describen los diferentes tipos de datos que podemos utilizar en Babelomics: https://github.com/babelomics/babelomics/wiki/Data-types.
3. Tras la carga de los datos, el primer paso será la normalización. Desde “Processing / Normalization NGS: RNA-Seq” seleccionaremos nuestro fichero y elegiremos un método de normalización (empezaremos con TMM).
4. Una vez que los datos ya están normalizados, ya estamos preparados para realizar el clustering. Desde “Expression / Unsupervised analysis”, selecciona los datos.
5. A continuación, seleccionamos el clustering por muestras. Elegimos un método de clustering y la distancia (para empezar, los que hay por defecto). Asignamos un nombre al job y lo ejecutamos.

Cuestiones.

1. ¿Se observan grupos de individuos con un perfil transcriptómico similar? ¿Cuántos grupos aparecen?
2. ¿Hay alguna muestra que tenga un comportamiento anómalo respecto del resto de sujetos?

EJERCICIO 2.
Análisis de datos de RNA-Seq: diferencia de expresión entre grupos.

Objetivo.
Detectar genes diferencialmente expresados entre grupos de inviduos.

Datos.
Disponemos de conteos no normalizados en 29.405 genes para un total de 10 individuos. Estos datos se obtuvieron tras aplicar un análisis primario que incluyó la evaluación de calidad de las secuencias, mapeo y cuantificación de la expresión a nivel de gen.

Las cinco primeras muestras corresponden a RNA de riñón (k) y las 5 últimas a RNA de pulmón (l).

Plan de trabajo.

1. Abre el fichero de datos “rnaseq_difexp.txt” con una hoja de cálculo e inspeccion su contenido. Habrá tantas columnas como sujetos y tantas filas como metabolitos que queremos valorar.
2. Sube este fichero txt en Babelomics desde el menú “Upload”. Tendremos que indicar el tipo de dato que subimos: “Data matrix expression”. En este link se describen los diferentes tipos de datos que podemos utilizar en Babelomics: https://github.com/babelomics/babelomics/wiki/Data-types
3. Tenemos que informar a la herramienta que las 5 primeras muestras son de riñón (K) y las 5 últimas son pulmón (L) Esta información la incluiremos desde “Processing / Edit your uploaded data”, siguiendo el asistente que aparece en esta opción. Para ello generamos una variable que llamaremos “GRUPO” y utilizaremos los valores “K” y “L” para etiquetar cada uno de los valores. Tras asignar un valor a cada sujeto, guardaremos los cambios.
4. Tras la carga y etiquetado de los datos, el siguiente paso incluye la normalización y la expresión diferencial simultáneamente. . Desde “Expression / RNA-Seq, class comparison” seleccionaremos nuestro fichero, la variable que establece los grupos que queremos comparar, un método de normalización (TMM) y un procedimiento para la corrección de test múltiple.
5. Por último, damos un nombre al job y lo ejecutamos.

Cuestiones.

1. ¿Cuántos genes diferencialmente expresados hemos obtenido en el análisis?
2. ¿Cuál es la interpretación de los gráficos que aparecen en los resultados?
3. Realiza un análisis de enriquecimiento sobre los genes significativos que tienen un nivel de expresión mayor en “K” que en “L”.
4. Descarga el ficheros con resultados significativos y comenta los valores obtenidos para un que nos gusta especialmente: MSRA.

EJERCICIO 3.
Diferencias en niveles de metabolitos en cáncer de pulmón.

Objetivo.
Detectar marcadores metabolómicos en pacientes con cáncer de pulmón.

Datos.
Disponemos de 2 grupos de sujetos: 5 pacientes con cáncer de pulmón (LC) y 5 pacientes sanos (CONTROL) en los que se han evaluado el nivel de 22 metabolitos de interés.

Los datos se obtuvieron mediante resonancia magnética y previamente se realizó un procesamiento que incluyó la integración de los espectros correspondientes a cada metabolito y su posterior normalización conjunta.

Plan de trabajo.

1. Abre el fichero de datos “metabolitos.txt” con una hoja de cálculo e inspecciona su contenido. Habrá tantas columnas como sujetos y tantas filas como metabolitos que queremos valorar.
2. Sube este fichero txt en Babelomics desde el menú “Upload”. Tendremos que indicar el tipo de dato que subimos: “Data matrix expression”. En este link se describen los diferentes tipos de datos que podemos utilizar en Babelomics: https://github.com/babelomics/babelomics/wiki/Data-types
3. Tenemos que informar a la herramienta que los primeros 5 sujetos pertenecen al grupo “lung cancer” y los 5 siguientes son “controles”. Esta información la incluiremos desde “Processing / Edit your uploaded data”, siguiendo las indicaciones del asistente de este editor. Para ello generamos una variable que llamaremos “GRUPO” y utilizaremos los valores “LC” y “CONTROL” para etiquetar cada uno de los valores. Tras asignar un valor a cada sujeto, guardaremos los cambios.
4. Ahora Babelomics ya conoce el grupo al que pertenece cada individuo y seguimos con el siguiente paso: la detección de diferencias de metabolitos entre ambos grupos.
5. Si queremos conocer las diferencias de los niveles de metabolitos entre ambos grupos, utilizaremos la opción “Differential Expression / Microarray /Class Comparison”:
6. Select your data: indicamos a Babelomics que queremos trabajar con el fichero “metabolitos”.
7. Seleccionamos la variable que establece los grupos (a veces podemos disponer de varias variables) e indicamos los grupos que queremos comparar. Me interesa LC vs CONTROL.
8. Escogemos un método de comparación que nos guste (empezaremos con el t-test) y un procedimiento de ajuste de p-valores (de momento el que hay por defecto).
9. Asignamos un nombre al job y lo ejecutamos!

Cuestiones.

1. ¿Cuántos marcadores significativos encontramos?
2. ¿Cuáles tienen un nivel más alto en Lung Cancer que en Control?
3. ¿Y qué marcadores presentan niveles menores en Lung Cancer que en Control?
4. ¿Cómo interpretamos el heatmap que nos proporciona Babelomics?
5. Nos hubiera gustado que la "Creatina" hubiese sido un marcador de interés. ¿Puedes bajar desde Babelomics todos los resultados e inspeccionar los resultados en este metabolito?

Algunos comentarios.

La expresión diferencial, clasificación no supervisada (clustering) y supervisada (predictores) son procedimientos implementados en Babelomics inicialmente para datos de expresión procedentes de microarrays o RNA-Seq. Además se pueden utilizar con cualquier set de datos que presenten una estructura matricial donde las columnas correspondan a sujetos que participen en el estudio y cada fila represente una unidad biológica que queramos evaluar (gen, proteína, metabolito, microRNA...)

EJERCICIO 4.

Caracterización funcional de genes que incluyen variantes de interés en diferentes análisis de exoma completo.

Objetivo.

Caracterizar funcionalmente un grupo de genes de interés procedente de un estudio genómico.

La secuenciación de exomas es últil en el descubrimiento de nuevas variantes de interés. Tras valorar un determinado número de sujetos, es posible detectar la presencia repetida de algunas de estas variantes en determinadas enfermedades. Estos genes serían unos buenos candidatos en el diseño de un panel de genes orientado al diagnóstico de estas patologías.

Antes de su inclusión en el diseño, nos gustaría realizar una selección lo más certera posible y para ello nos interesa disponer de diversos criterios de evaluación de estos genes.

El enriquecimiento funcional de estos genes nos proporciona una información adicional que permitirá una mejor comprensión de los procesos en los que están participando.

Datos.

Disponemos de 73 genes en los que hemos detectado variantes en diferentes sujetos con distrofias de retina. Están incluidos en el fichero de texto llamado "distrofias".

Plan de trabajo.

1. Abre el fichero de datos “distrofias.txt” con un bloc de notas o similar e inspecciona su contenido.
2. Sube este fichero txt en Babelomics desde el menú “Upload”. Tendremos que indicar el tipo de dato: “Id list (Id)”.
3. Como queremos realizar un análisis de enriquecimiento funcional, vamos al menú “Functional / Single Enrichment: FatiGO”.
4. Para empezar, compararemos nuestra lista de genes frente el resto del genoma. Seleccionamos los datos con los que trabajaremos y el organismo (en este caso humano).
5. Es posible realizar simultáneamente varios análisis de enriquecimiento utilizando diferentes bases de datos. Seleccionamos los procesos biológicos, funciones moleculares y componentes celulares de la Gene Ontology.
6. Damos un nombre al job y lo ejectuamos. Puede tardar unos 20 minutos la finalización de este job.

Cuestiones.

1. ¿Cuántas términos GO son significativos? (Especificando por procesos biológicos, funciones moleculares y componentes celulares?
2. Nos interesa especialmente el término “retinal binding” (función molecular). ¿Qué genes de nuestra lista de interés, están enriquecidos con esta función?
3. ¿Qué nos indican los valores estadísticos que aparecen en este término GO en cada una de las columnas de la tabla?
4. ¿Cómo interpretamos el gráfico que aparece bajo la tabla de resultados significativos?