

# Algoritmos en Biología Computacional

## BCOM4006

### Tarea #5 – Coexpresión y medidas de centralidad

Estudiantes:

Michael Dylan Blanquicett Carvajal

Luisa Hernandez

Universidad de los Andes

24/10/2024

Cundinamarca, Bogotá, Bogotá D.C.

## Contenido

No se encontraron entradas de tabla de contenido.

## Parte 1

### Configuración R y CRAN

Para esta parte se descargo el Script de R y se ejecutó la sección de código que tiene todos los paquetes necesarios:

```
downloaded 3.1 MB
```

```
package 'fs' successfully unpacked and MD5 sums checked
package 'rappdirs' successfully unpacked and MD5 sums checked
package 'stringi' successfully unpacked and MD5 sums checked
package 'backports' successfully unpacked and MD5 sums checked
package 'sass' successfully unpacked and MD5 sums checked
package 'stringr' successfully unpacked and MD5 sums checked
package 'checkmate' successfully unpacked and MD5 sums checked
package 'htmlwidgets' successfully unpacked and MD5 sums checked
package 'rstudioapi' successfully unpacked and MD5 sums checked
package 'digest' successfully unpacked and MD5 sums checked
package 'bslib' successfully unpacked and MD5 sums checked
package 'evaluate' successfully unpacked and MD5 sums checked
package 'fontawesome' successfully unpacked and MD5 sums checked
package 'jquerylib' successfully unpacked and MD5 sums checked
package 'tinytex' successfully unpacked and MD5 sums checked
package 'xfun' successfully unpacked and MD5 sums checked
package 'yaml' successfully unpacked and MD5 sums checked
package 'highr' successfully unpacked and MD5 sums checked
package 'gridExtra' successfully unpacked and MD5 sums checked
package 'data.table' successfully unpacked and MD5 sums checked
package 'htmlTable' successfully unpacked and MD5 sums checked
package 'viridis' successfully unpacked and MD5 sums checked
package 'htmltools' successfully unpacked and MD5 sums checked
package 'base64enc' successfully unpacked and MD5 sums checked
package 'rmarkdown' successfully unpacked and MD5 sums checked
package 'knitr' successfully unpacked and MD5 sums checked
package 'Formula' successfully unpacked and MD5 sums checked
package 'iterators' successfully unpacked and MD5 sums checked
package 'dynamicTreeCut' successfully unpacked and MD5 sums checked
package 'fastcluster' successfully unpacked and MD5 sums checked
package 'Hmisc' successfully unpacked and MD5 sums checked
package 'impute' successfully unpacked and MD5 sums checked
package 'foreach' successfully unpacked and MD5 sums checked
package 'doParallel' successfully unpacked and MD5 sums checked
package 'preprocessCore' successfully unpacked and MD5 sums checked
package 'WGCNA' successfully unpacked and MD5 sums checked
```

### Análisis de los datos

Se puede encontrar información de los datos en el siguiente artículo:

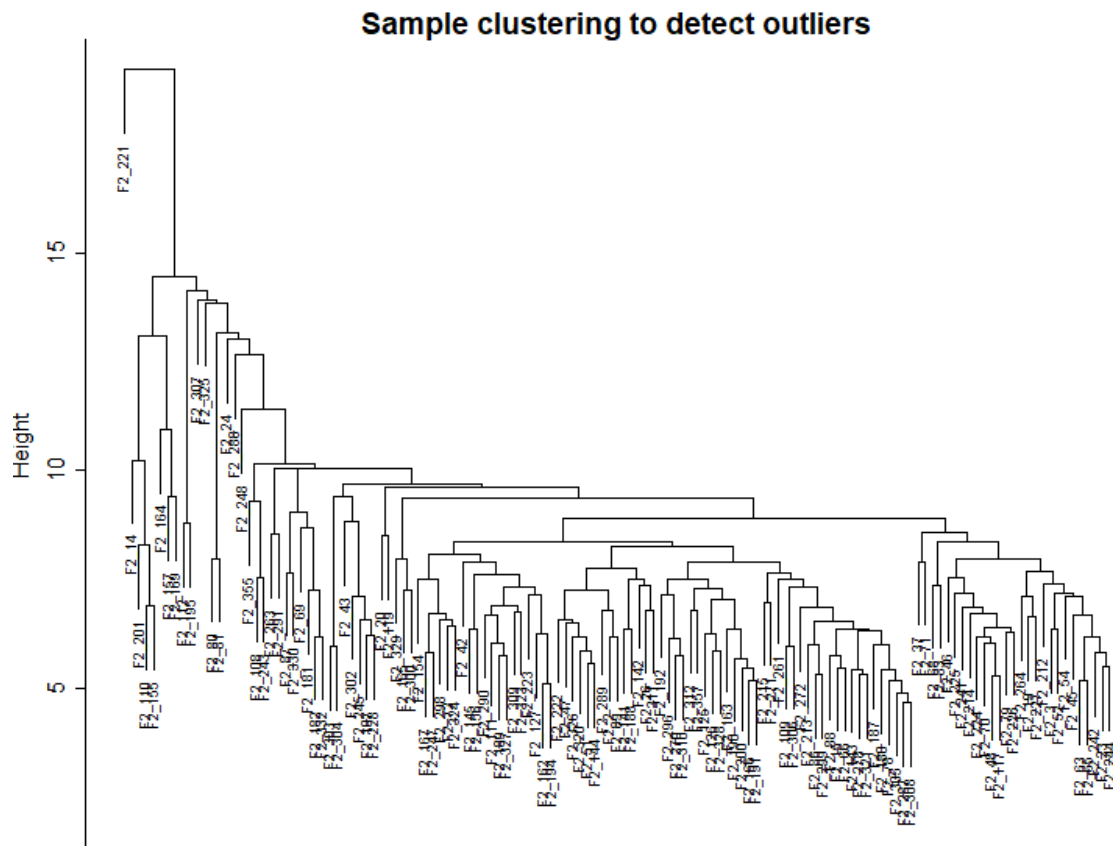
<https://pmc.ncbi.nlm.nih.gov/articles/PMC1998880/>

Aquí podemos entender que el archivo FemaleLiver3600.csv es un archivo con información sobre 3600 transcripts obtenidos del hígado de 135 ratones hembra, la

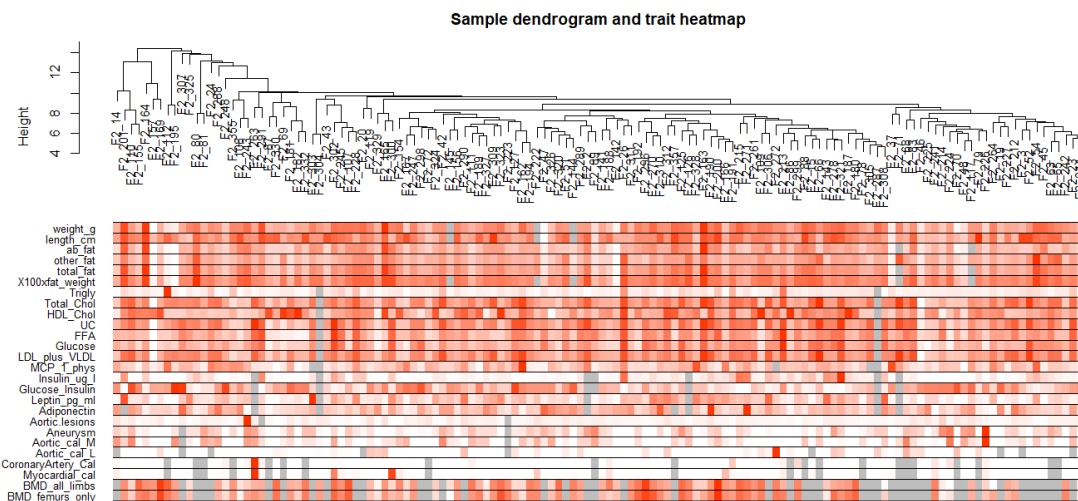
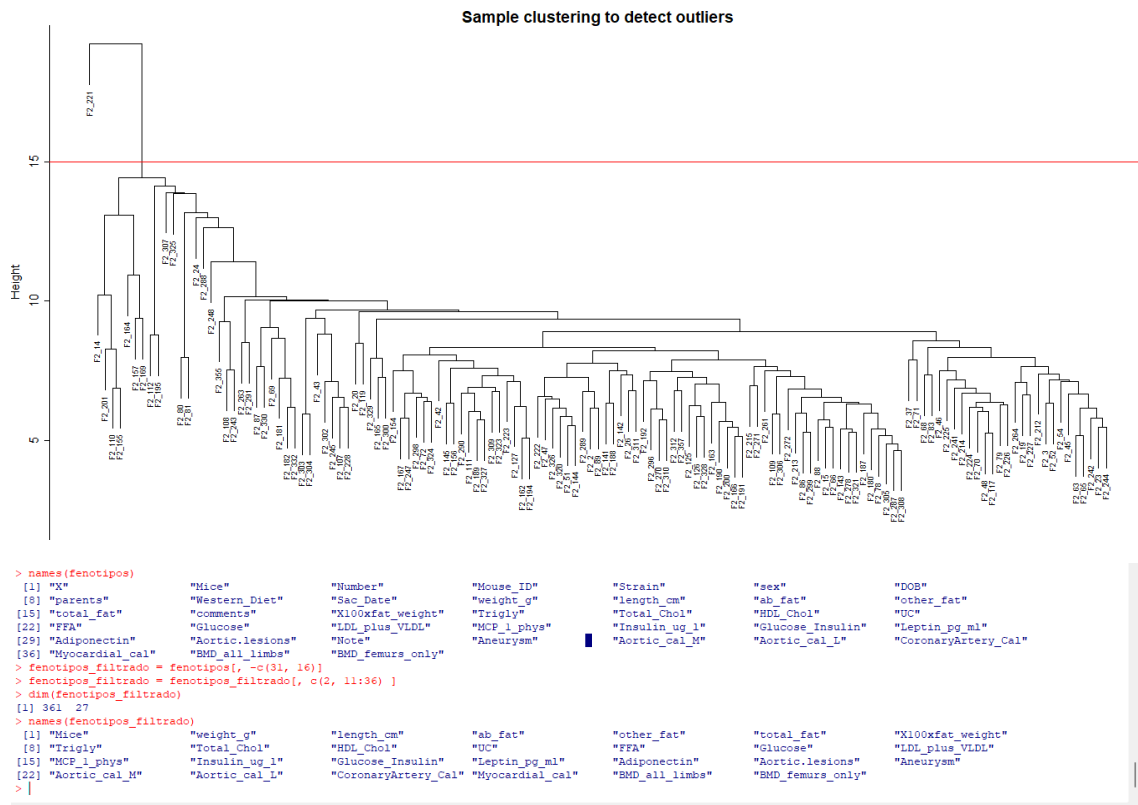
muestra conocida como BxH el cual es el cruce entre las familias C3H/HeJ y C57BL/6J, también están las muestras del hígado de 113 ratones que son el cruce entre C57BL/6J y DBA/2J. Este será el archivo más importante para el análisis por peso de los genes  
Análisis de red de co-expresión genética ponderada.

EL archivo ClinicalTraits.csv sirve para poder categorizar todas las características fenotípicas de los ratones. Sirve para poder asociar las diferentes características de los ratones con las muestras.

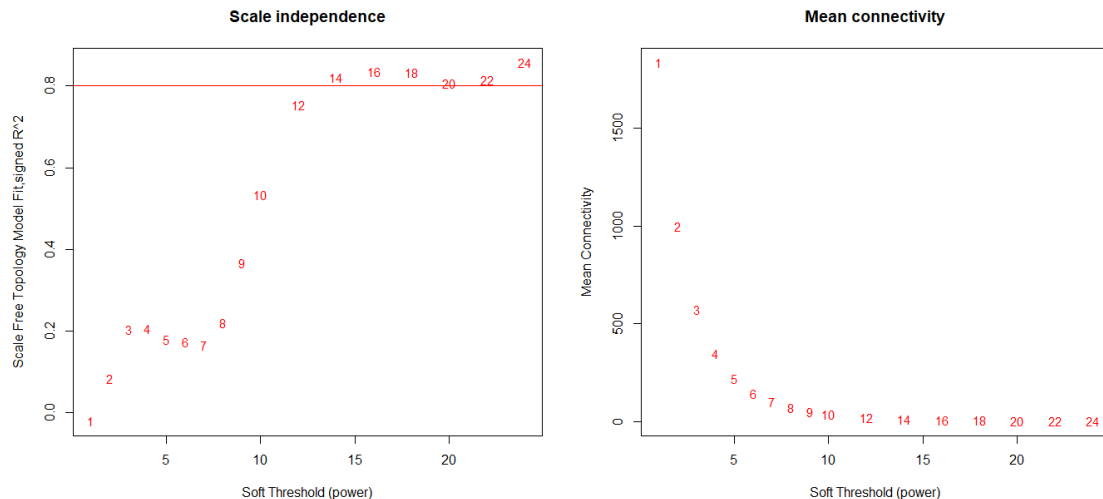
## Punto 1



Podemos observar en esta grafica obtenida de realizar una gráfica de tipo árbol, se puede observar que el corte de la altura se puede dar entre la muestra F2\_221 y las demás, siendo esta fuera del corte, por ejemplo, en la altura 15.



Para entender mejor cómo se relacionan los genes, utilizamos potenciadores de co-expresión con soft-thresholds. Esto permite medir la intensidad de la conexión. Esto crea una red genética que refleja con mayor precisión las interacciones reales entre los genes, proporcionando una visión más detallada de cómo trabajan juntos.



Signed  $R^2$  es una métrica que evalúa qué tan bien la topología de la red genética sigue un modelo de red libre de escala. Un valor alto indica que la red tiene una estructura de red libre de escala, lo cual es deseable para análisis genéticos al sugerir que la red es robusta.

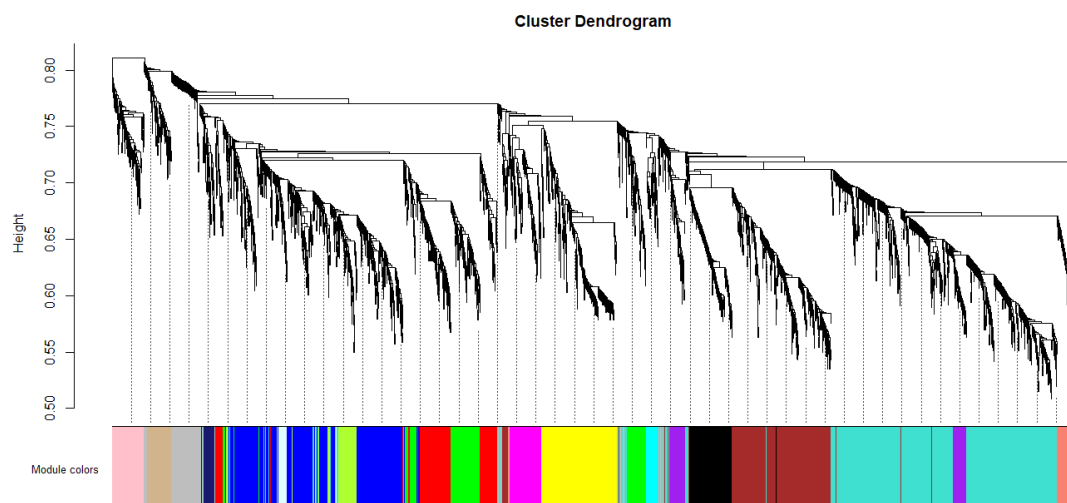
Para seleccionar un potenciador en el rango de 1 a 30, buscamos un valor que mantenga o supere este corte de 0.8 en la gráfica de  $R^2$  firmado. Este valor garantiza que la estructura de la red preserve las propiedades de una red libre de escala.

Módulos y genes por módulos después de realizar clustering con potenciador 1:

```
> table(red$colors)
```

```

 0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
179 892 494 415 282 232 223 162 120 117 111  95  88  74  43  40  33
> |
```



Solo como ejemplo podemos observar este dendrograma y los diferentes clusters que salen al usar un potenciador 1, ahora, para poder encontrar el potenciador que nos interesa, podemos ir cambiándolo y verificando, o podemos hacer uso de `picksSoftThreshold` y graficar el resultado con todos los enteros del 1 al 30:

```

> powers = c(1:30)
> pickSoftThreshold(df_expr_filtrado, powerVector = powers, verbose = 5)
pickSoftThreshold: will use block size 3600.
pickSoftThreshold: calculating connectivity for given powers...
..working on genes 1 through 3600 of 3600

```

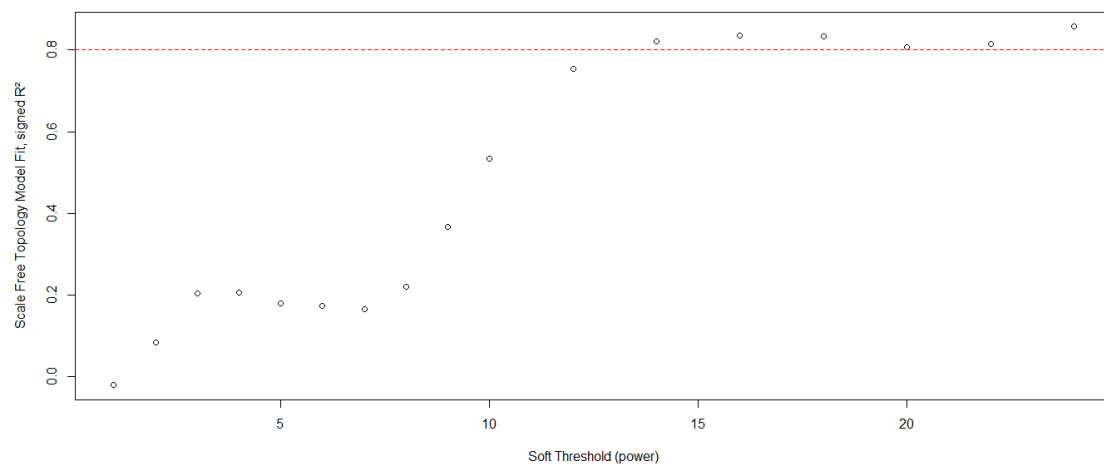
	Power	SFT.R.sq	slope	truncated.R.sq	mean.k.	median.k.	max.k.
1	1	0.0278	0.345	0.456	747.000	7.62e+02	1210.0
2	2	0.1260	-0.597	0.843	254.000	2.51e+02	574.0
3	3	0.3400	-1.030	0.972	111.000	1.02e+02	324.0
4	4	0.5060	-1.420	0.973	56.500	4.72e+01	202.0
5	5	0.6810	-1.720	0.940	32.200	2.51e+01	134.0
6	6	0.9020	-1.500	0.962	19.900	1.45e+01	94.8
7	7	0.9210	-1.670	0.917	13.200	8.68e+00	84.1
8	8	0.9040	-1.720	0.876	9.250	5.39e+00	76.3
9	9	0.8590	-1.700	0.836	6.800	3.56e+00	70.5
10	10	0.8330	-1.660	0.831	5.190	2.38e+00	65.8
11	11	0.8640	-1.570	0.894	4.100	1.65e+00	61.7
12	12	0.8530	-1.480	0.911	3.330	1.15e+00	58.1
13	13	0.8610	-1.430	0.933	2.770	7.99e-01	54.8
14	14	0.8760	-1.380	0.949	2.350	5.74e-01	51.9
15	15	0.9010	-1.340	0.961	2.020	4.22e-01	49.3
16	16	0.9070	-1.300	0.970	1.770	3.09e-01	46.8
17	17	0.9050	-1.260	0.969	1.560	2.25e-01	44.6
18	18	0.9120	-1.240	0.973	1.390	1.67e-01	42.5
19	19	0.9190	-1.230	0.970	1.250	1.25e-01	40.5
20	20	0.9310	-1.210	0.977	1.140	9.51e-02	38.7
21	21	0.9400	-1.190	0.976	1.030	7.27e-02	37.0
22	22	0.9450	-1.170	0.975	0.947	5.46e-02	35.4
23	23	0.9510	-1.160	0.979	0.871	4.16e-02	33.9
24	24	0.9500	-1.150	0.978	0.804	3.15e-02	32.4
25	25	0.9580	-1.140	0.980	0.745	2.44e-02	31.1
26	26	0.9690	-1.140	0.987	0.692	1.89e-02	29.8
27	27	0.9760	-1.130	0.990	0.645	1.47e-02	28.7
28	28	0.9690	-1.130	0.974	0.602	1.13e-02	27.5
29	29	0.9600	-1.130	0.959	0.564	8.68e-03	26.5
30	30	0.9750	-1.120	0.974	0.528	6.72e-03	25.4

```

$powerEstimate

```

De esta forma realizamos nuestro propio proceso para encontrar el potenciador, el cual coincide con el arrojado por la primera gráfica:

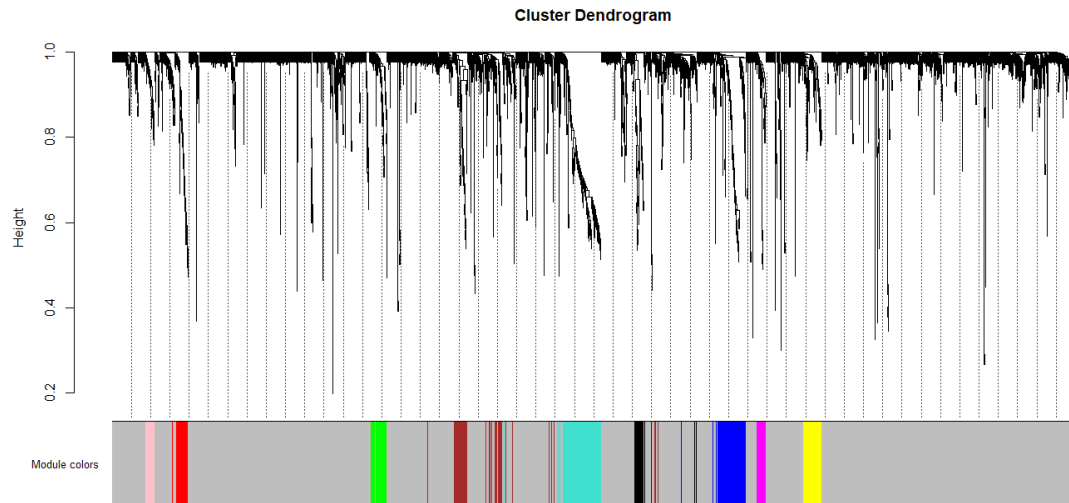


Con potenciador de 20:

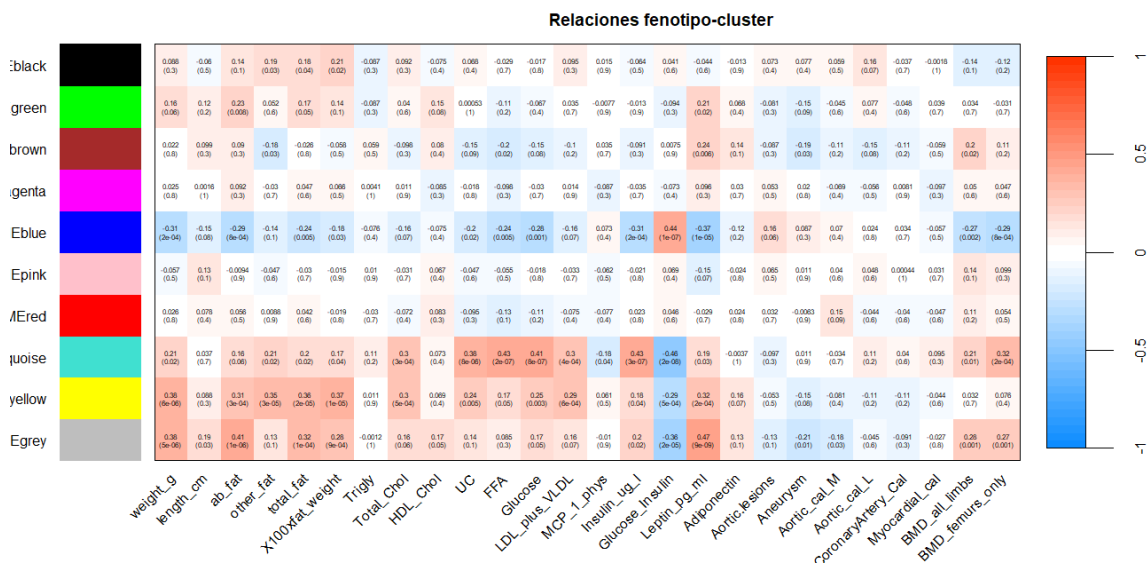
```
> table(red$colors)

 0    1    2    3    4    5    6    7    8    9
2950 147 120 101  72  54  47  42  34  33
> |
```

También obtenemos el siguiente dendrograma de clusters:



Gracias a la selección del potenciador, obtenemos el siguiente diagrama de relaciones entre los diferentes fenotipos:

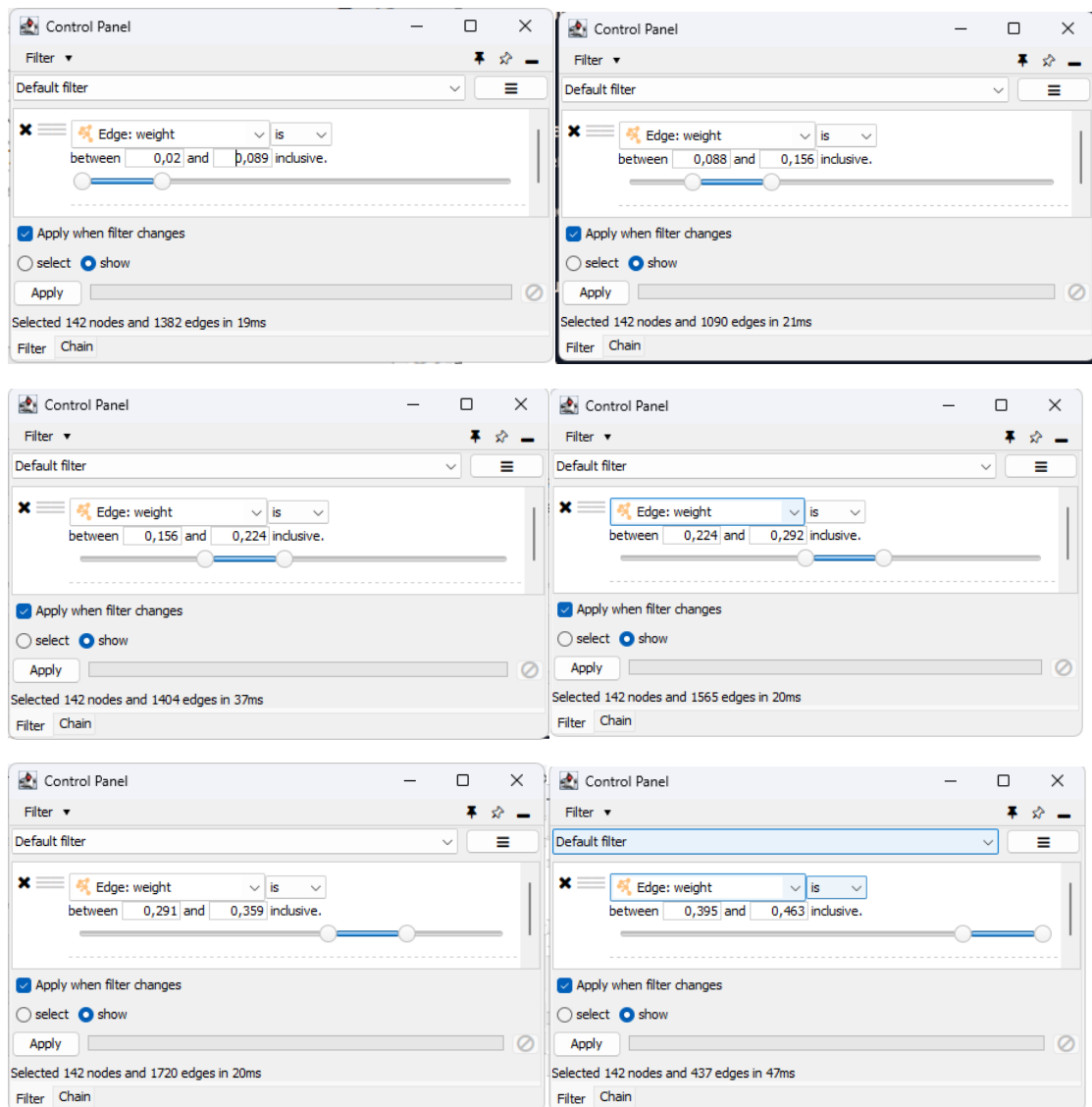


### Punto 3

Para realizar este punto tomaremos como referencia la grafica de relaciones de fenotipos del punto anterior, el cual obtuvimos al usar un potenciador de 20. Solo por interés en ver la distribución de las lecturas, se escogerá Glucose\_Insulin al ser el fenotipo con 3 clusters con un valor significativo de relaciones inversas (negativas), siendo el módulo de mayor interés el turques:







Se puede analizar que con el grafo con muchos nodos fuertemente conectados sugerimos que el grupo de genes que trabajan sobre el metabolismo de la glucosa e insulina lo hacen en conjunto.

Todos los archivos se encuentran en la carpeta /data.