

Métodos computacionales para la caracterización de dianas de microRNA en datos CLIP-seq

José Antonio Piqueras Mellado
Consultor: Albert Pla Planas

Máster universitari de Bioinformàtica y Bioestadística
Universitat Oberta de Catalunya
Universitat de Barcelona

Junio - 2018

Contexto y justificación del trabajo

microRNAs

- Pequeños ARNs no codificantes que intervienen en la regulación de la expresión génica inhibiendo la traducción de los mRNAs con los que interaccionan.

Interacciones miRNA-mRNA

- Especificidad de las interacciones.
- Región semilla.
- Dianas canónicas y no canónicas.

Contexto

- Tecnología CLIP-seq.
- Rol desconocido de las dianas no canónicas.
- Elevado coste de la experimentación biológica.

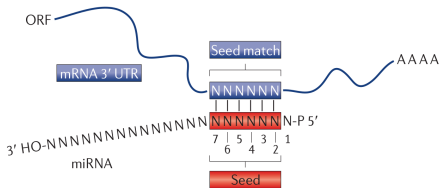


Imagen tomada de la publicación de Huntzinger e Izaurralde [1].

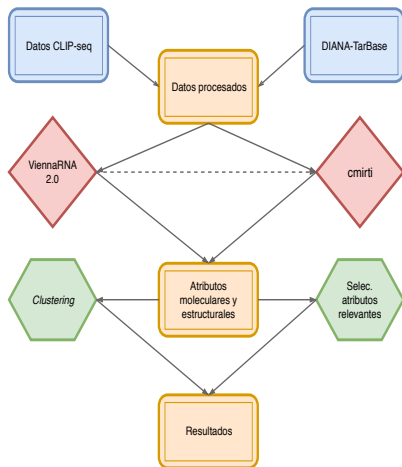
Objetivos y plan de trabajo

Objetivos del proyecto

- Construir un conjunto de datos y herramientas adecuadas a la caracterización de interacciones miRNA-mRNA.
- Identificar características frecuentes o distintivas de las dianas miRNA no canónicas.
- Interpretar las posibles implicaciones biológicas de los resultados obtenidos.

Plan de trabajo

- Recopilación y procesamiento de los datos.
- Caracterización de las interacciones miRNA-mRNA.
- Análisis de *clustering* y selección de atributos relevantes.



Recopilación y procesamiento de los datos

Datos

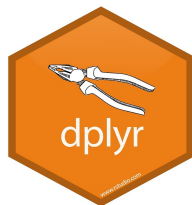
- CLIP-seq [2-4].
- DIANA-TarBase [5].

Tareas

- Recuperación de secuencias nucleótidas.
- Obtención de distintos identificadores.
- Validación funcional de las interacciones.

Especies

- Humano (*Homo sapiens*).
- Ratón (*Mus musculus*).



Caracterización de interacciones miRNA-mRNA

Paquete ViennaRNA 2.0 [6]

- Predicción estructura secundaria (*dot-bracket*).
- Estabilidad termodinámica de los dímeros.
- Accesibilidad a la diana.

Paquete cmirti

- Notación modificada.
- Clasificación de las dianas.
- Extracción de atributos moleculares y estructurales.

A

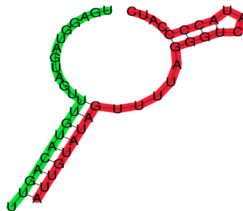
ViennaRNA

```
UGAGGUAGUAGUUUGUACAGUU&AUUGUAGUAGUUUAGGGUCAUACCCCAUC  
.....(((((((.&))))))))......((((...))).....
```

cmirti

```
UGAGGUAGUAGUUUGUACAGUU  
.....!|!|!|!|!|!  
CUACCCCAUACUGGGAUUUUGAUUGUUA  
...[[[...]]].....!|!|!|!|!|!
```

B



Caracterización de interacciones miRNA-mRNA

Dianas canónicas [7-9]

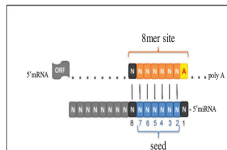
- Canónicas (offset-6mer, 6mer, 7mer-A1, 7mer-m8, 8mer).
- Canónicas con unión suplementaria en 3' (offset-6mer-supp, ..., 8mer-supp).

Dianas no canónicas [7-9]

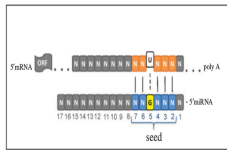
- No canónicas con unión compensatoria en 3'.
- No canónicas con unión centrada.
- No canónicas (NonCa-1mism, ..., NonCa-6mism).

Imagen tomada de la publicación de Moszyńska *et al.* [9].

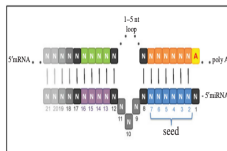
(e) 8-mer



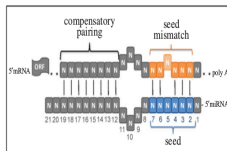
(f) GU wobble pairing



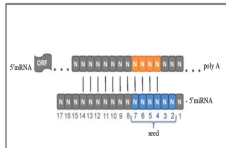
(g) productive 3'-pairing



(h) compensatory site



(i) centred site



Caracterización de interacciones miRNA-mRNA

Extracción de atributos

- Cantidad, tipo y disposición de los pares de bases entre miRNA y diana.
- Cantidad y tamaño de distintos motivos en la estructura secundaria de los dímeros.
- Composición nucleotídica de los miRNAs.

Composición nucleotídica de los miRNAs

- Proporción de nucleótidos (A, C, G, U).
- Proporción de dinucleótidos (AA, ..., UU).
- Contenido GC y contenido AU.
- Proporción de bases púricas y pirimidínicas.

Caracterización de interacciones miRNA-mRNA

Pares de bases miRNA-diana

- Número de pares de bases (Watons-Crick y G-U *wobble*).
- Número de pares de bases intramoleculares (horquillas).
- Distancia mínima, media y máxima entre pares de bases correlativos.
- Número de regiones de pares de bases consecutivas.
- Longitud mínima, media y máxima de las regiones de pares de bases consecutivas.

Mediciones por duplicado

- Interacción: Watson-Crick + G-U *wobbles*
- Interacción: Watson-Crick

miRNA Interaction Map

UAAAGUGCUGACAGUGCAGAU

.. |||!!||| . ||!|| ...

.. |||!!||| . ||!|| ...

Caracterización de interacciones miRNA-mRNA

Estructura secundaria del dímero

- Número de motivos en el dímero.
- Longitud o tamaño medio de los motivos.
- Longitud o tamaño del motivo de menor y mayor tamaño.
- Número total de nucleótidos que participan en motivos de una misma clase.

Motivos estructurales

- A Protuberancia o *bulge*.
- B Bucle o *loop* simétrico.
- B Bucle o *loop* asimétrico.
- C *Three-way junction*.
- D *Multi-way junction*.

A



B



C



D



Análisis de los datos

Transformación de los datos

- Normalización de las variables.
- Estandarización (*z-score*) de las variables.
- Análisis de componentes principales (PCA).
- Análisis de componentes independientes (ICA).

Análisis de clustering

- CLARA:
 - ▶ Coeficiente de la silueta.
- *Expectation-Maximization* (EM):
 - ▶ Criterio de información bayesiano (BIC).
- OPTICS:
 - ▶ Distancia a los k vecinos más cercanos.

Selección de atributos relevantes

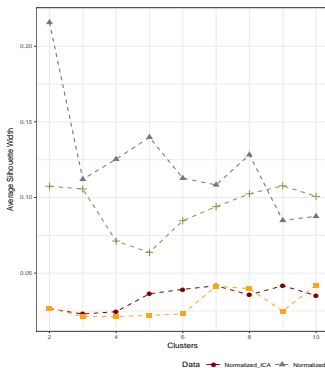
- Previo a la selección:
 - ▶ Prueba t de Student.
 - ▶ Medida d de Cohen.
- Método de filtrado:
 - ▶ GAM \rightarrow *Random Forest*.

Resultados: CLARA

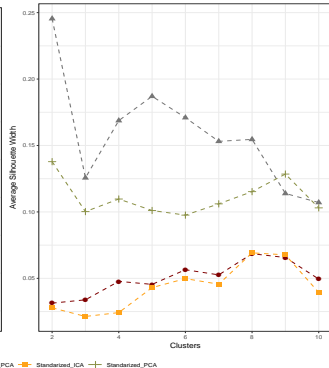
Puntos a comentar

- Normalización y PCA.
- Se descarta $k = 2$.
- Máximo: $k = 5$.
- Humano: $k = 8$.
- Ratón: $k = 6$.

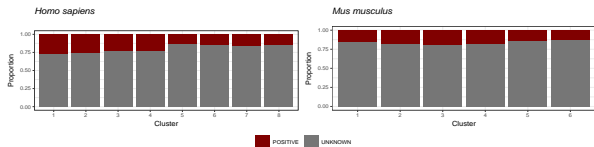
Homo sapiens



Mus musculus



Resultados: CLARA

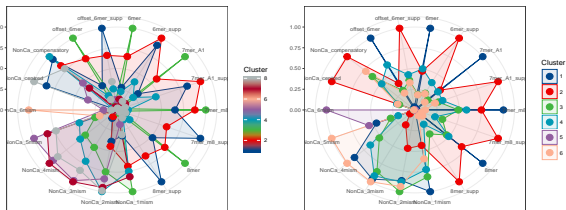
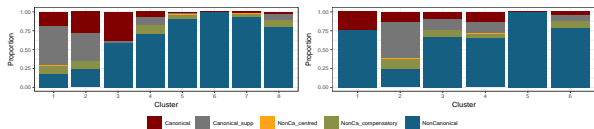


Funcionalidad

● Humano \neq Ratón.

Tipo de interacción

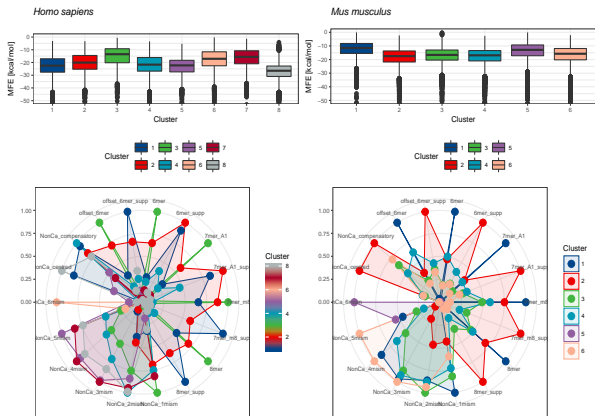
● Humano \approx Ratón.



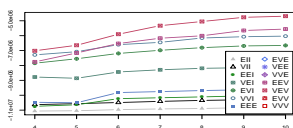
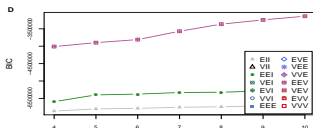
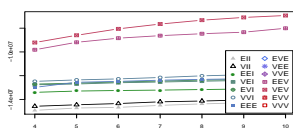
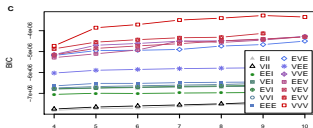
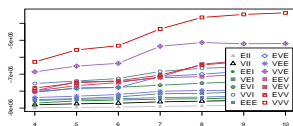
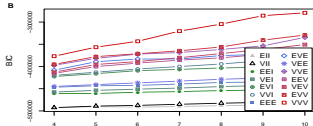
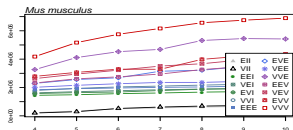
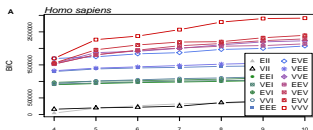
Resultados: CLARA

Estabilidad de las interacciones

- Dianas canónicas menos estables (resultado llamativo).
- Dianas sin interacción en la región semilla menos estables.



Resultados: EM



Punto a comentar

- Número de grupos.

Transformación de los datos

- A. Normalización y PCA.
- B. Normalización e ICA.
- C. Estandarización y PCA.
- D. Estandarización e ICA.

Resultados: EM

Punto a comentar

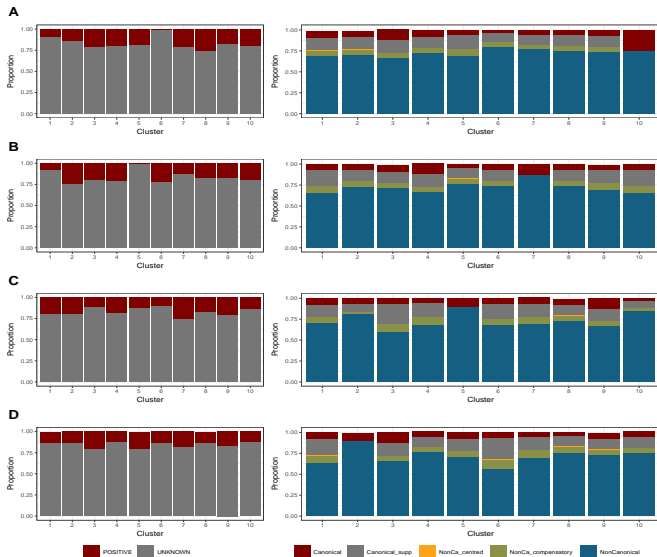
- Homogeneidad de los grupos.

Transformación de los datos

- A. Normalización y PCA.
- B. Normalización e ICA.
- C. Estandarización y PCA.
- D. Estandarización e ICA.

Especie

- *Mus musculus*.



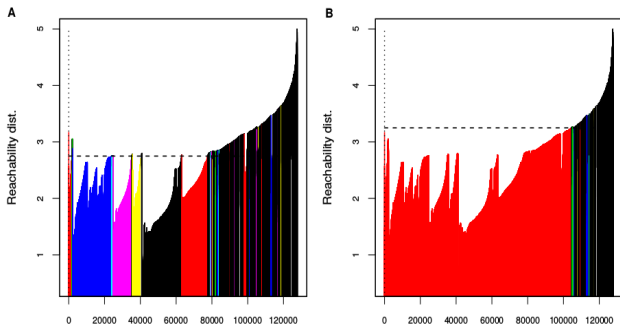
Resultados: OPTICS

Punto a comentar

- Imposibilidad de encontrar un punto de corte para el parámetro ϵ que permitiera obtener un agrupamiento razonable dada la naturaleza de los datos y del problema a tratar.

Ejemplo

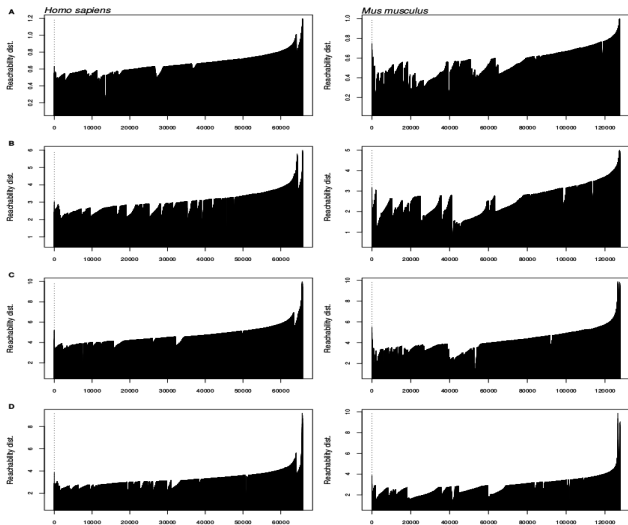
- *Mus musculus*.
- Normalización e ICA.



Resultados: OPTICS

Transformación de los datos

- A. Normalización y PCA.
- B. Normalización e ICA.
- C. Estandarización y PCA.
- D. Estandarización e ICA.



Resultados: Selección de atributos

Selección de atributos relevantes

- Previo a la selección:
 - ▶ Prueba t de Student.
 - ▶ Medida d de Cohen.
- Método de filtrado:
 - ▶ GAM → *Random Forest*.

Previo a la selección

- Diferencias significativas en la mayor parte de las variables.
- Tamaño del efecto despreciable en la mayor parte de los casos.
- Efecto pequeño en algunas variables en humano.

Selección de atributos

- Saldo: 55 variables en humano y 56 en ratón.
- Escasa capacidad de predicción de la funcionalidad de las interacciones.

Discusión

Resultados

- Patrones consistentes en ambas especies.
- La mayor parte de las dianas funcionales son canónicas.
- La complementariedad de bases en la región 3' del miRNA mejora la estabilidad y la efectividad de las interacciones [10].

Limitaciones

- Desconocimiento sobre la funcionalidad de la mayor parte de las interacciones miRNA-diana.
- Desconocimiento de la estructura secundaria de las interacciones *in vivo*.
- Reducción del rendimiento de los algoritmos de *clustering* cuando la dimensionalidad de los datos es alta [11].

Futuro

- Supercomputadores o *clusters*.
- Nuevos algoritmos de *clustering*.
- Estudio de otras especies.

Conclusiones

Conclusiones

- Objetivos cumplidos aunque con matices.
 - ▶ Procesamiento de los datos y extracción de atributos a partir de las interacciones miRNA-diana.
 - ▶ **Caracterización e identificación de atributos relevantes en las dianas no canónicas**
- Se ha respetado la planificación del proyecto.
- Pese a las limitaciones, la metodología seguida era adecuada a los objetivos propuestos.

Bibliografía

- [1] Huntzinger E, Izaurralde E. 2011 Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nature Reviews. Genetics* 12, 99–110. (doi:10.1038/nrg2936)
- [2] Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013 Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* 153, 654–665. (doi:10.1016/j.cell.2013.03.043)
- [3] Grosswendt S, Filipchuk A, Manzano M, Klironomos F, Schilling M, Herzog M, Gottwein E, Rajewsky N. 2014 Unambiguous Identification of miRNA:target site Interactions by Different Types of Ligation Reactions. *Molecular cell* 54, 1042–1054. (doi:10.1016/j.molcel.2014.03.049)
- [4] Moore MJ, Scheel TKH, Luna JM, Park CY, Fak JJ, Nishiuchi E, Rice CM, Darnell RB. 2015 miRNA–target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nature Communications* 6. (doi:10.1038/ncomms9864)
- [5] Vlachos IS et al. 2015 DIANA-TarBase v7.0: Indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research* 43, D153–D159. (doi:10.1093/nar/gku1215)
- [6] Lorenz R, Bernhart SH, Höner zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011 ViennaRNA Package 2.0. *Algorithms for Molecular Biology* 6, 26. (doi:10.1186/1748-7188-6-26)
- [7] Bartel DP. 2009 MicroRNAs: Target Recognition and Regulatory Functions. *Cell* 136, 215–233. (doi:10.1016/j.cell.2009.01.002)
- [8] Friedman RC, Farh KK-H, Burge CB, Bartel DP. 2009 Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* 19, 92–105. (doi:10.1101/gr.082701.108)
- [9] Moszyńska A, Gebert M, Collawn JF, Bartoszewski R. 2017 SNPs in microRNA target sites and their potential role in human disease. *Open Biology* 7, 170019. (doi:10.1098/rsob.170019)
- [10] Agarwal V, Bell GW, Nam J-W, Bartel DP. 2015 Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4, e05005. (doi:10.7554/eLife.05005)
- [11] Xu R, Wunsch DC. 2010 Clustering Algorithms in Biomedical Research: A Review. *IEEE Reviews in Biomedical Engineering* 3, 120–154. (doi:10.1109/RBME.2010.2083647)