Caracterització estructural i funcional de long non-coding RNAs (lncRNAs) en nematodes

Daniel Miquel Brossa

Directora: Dra. Maria Cinta Pegueroles Queralt

Contents

Resum

Els lncRNA són gens que al no codificar per proteïnes i estar poc conservats a nivell de seqüència, necessiten tècniques d'anàlisis diferents de les utilitzades per estudiar gens que sí que codifiquen proteïnes. Aquest fet suposa un repte i una oportunitat per impulsar l'ús de noves tècniques d'anàlisis i aprofundir en el coneixement de les funcions dels lncRNA[1]. A més, la seva funció és desconeguda per la majoria dels lncRNA anotats en l'actualitat, tot i que s'ha descrit la seva importància en el desenvolupament i la seva implicació en diverses malalties com ara càncer[2]. Per això, és important desenvolupar estratègies que ajudin a la seva caracterització funcional. En aquest treball he realitzat diferents estudis per tal d'avaluar aspectes estructurals i funcionals de 514 famílies de lncRNA's obtingudes de 4 espècies de nematodes. En els aspectes estructurals he observat que nombroses famílies presenten motifs conservats però en canvi la seva estructura secundària sembla poc conservada, només les famílies presents en les quatre espècies han presentat enriquiment en conservació estructural. En els aspectes funcionals, m'he enfocat en aquest grup de famílies i he anotat 252 interaccions amb 42 proteïnes diferents amb les quals he realitzat un estudi d'anotació funcional. He pogut observar, que un 74.68% de les proteïnes participen en processos relacionats amb la regulació de l'expressió genètica i que tots els gens dels lncRNAs amb els quals interaccionen es troben en coordenades allunyades.

Paraules clau: lncRNA, estructura secundària, interacció amb proteïnes, anotació funcional, nematodes.

Introducció

Des de mitjans del segle passat coneixem que quan es compara la quantitat de DNA present en el genoma d'un organisme amb la seva complexitat, no existeix una relació lineal. És a dir, la relació entre la complexitat d'un organisme i la mida del seu genoma és baixa (paradoxa del valor C). Després dels experiments d'hibridació de l'ADN-ARN que va realitzar Lewin B al 1980[14], es va determinar que els éssers humans no tindrien més de 20000-30000 gens que codificarien per proteïnes. Però des del 1970 es van començar a obtenir pistes que es transcrivia més genoma del que es podria atribuir a la part codificant[6].

En aquesta porció del genoma que s'expressa però no codificar per proteïnes trobem els lncRNA[1,3,4]. Els lncRNA es defineixen com transcrits de més de 200pb amb seqüències primaries poc conservades evolutivament i amb gran potencial de jugar un paper important en la regulació en l'expressió genètica[1,4,5]. Aquests últims anys amb el desenvolupament de noves tecnologies de seqüenciació a permès la identificació d'un gran nombre de lncRNA en molts organismes, però el coneixement d'aquests elements en la majoria d'espècies i la seva funcionalitat és molt limitat[2]. Molts estudis semblen indicar que l'expressió dels lncRNA és especific de certs teixits o etapes de desenvolupament i participaria en nombrosos processos cel · lulars i que fins hi tot podrien produir efectes a nivell d'organisme[7]. Per aquest fet és important la cerca de noves estratègies que ens permetin estudiar la seva caracterització funcional[1].

En aquest treball realitzarem un estudi a partir de 514 famílies de lncRNA obtingudes de quatre espècies de nematodes del genere *Caenorhabditis* (*C.briggsae*, *C.remanei*, *C.brenneri* i *C.elegans*) d'un dataset ja anotat[3]. Amb aquestes dades realitzarem diferents estudis per tal d'avaluar aspectes estructurals i funcionals d'aquestes famílies de lncRNA. Intentarem avaluar el grau de conservació de la seva estructura secundària, amb quines proteïnes interaccionen i en quins processos cel · lulars i teixits realitzarien la seva funció biològica.

Resultats i discusió

Conservació estructura secundària

En la majoria de transcrits que no codifiquen per proteïnes depenen en gran mesura de la seva estructura per realitzar les seves funcions biològiques. És de debat si en els lncRNA succeeix el mateix i l'estructura secundària té un paper important en la seva biologia[4]. Si la estructura secundària és tant necessària, segurament romandrà conservada per tal de poder mantenir la seva funcionalitat. Per contestar aquesta pregunta realitzarem l'avaluació de la conservació de l'estructura secundària dels lncRNA de la nostre base de dades. Per l'avaluació hem utilitzat un software desenvolupat recentment i especific per lncRNA anomenat CROSSalign[8]. Aquest programa, donada una seqüència, avalua la conservació de l'estructura secundària i dona un valor com a indicador del grau de conservació. Els creadors del programa consideren que un lncRNA té una estructura secundària conservada, quant s'obté un valor de distància estructural de menys de 0.095-0.01.

Mitjançant aquest programa hem obtingut els valors de conservació de l'estructura secundària de lncRNA de quatre espècies diferents de *Caenorhabdits* (*C.elegants, N2*; *C.briggsae, AF16*; *C.remanei, PB4641*; i *C.brenneri; PB2801*). Les seqüències d'aquests lncRNA van ser obtingudes per l'equip que va realitzar l'estudi **"Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus Caenorhabditis"**[3]. En total s'han realitzat i anotat 863 resultats dels anàlisis de conservació de l'estructura secundària. El programa a presentat un problemes a l'hora d'avaluar la conservació en les seqüències que presentaven codi de nucleòtid N i per aquest fet hem decidit de prescindir dels resultats obtinguts amb aquestes seqüències. Finalment hem utilitzat 850 resultats per a l'avaluació de la conservació de l'estructura secundària (**taula 1|** Representació dels resultats obtinguts amb CROSSalign, per lncRNA i per famílies.).

Conservació	lncRNA	percentatge	nº_famílies	percentatge.fam
Si	70	8.24	60	11.78
No	780	91.76	449	88.22
total	850	100.00	509	100.00

taula 1

Podem veure que amb els resultats obtinguts, que la gran majoria de lncRNA del nostre conjunt de dades (91.76%) no presenten una estructura secundària conservada. El 8.24% si que ha mostrat una distància estructural més petita o igual a 0.10 i per tant si que les podríem considerar semblants en termes d'estructura secundària. En conjunt els p-valors han sigut elevats, de les famílies que presenten conservació estructural, és a dir les que tenen una distància estructural <0.1, obtenim p-valors entre 0.0e+00 i 4.0e-03. A nivell de família veiem que el 11,78% de les famílies presentarien una estructura secundària conservada. També hem observat que en el cas de les famílies presents en les quatre espècies, el percentatge de famílies amb conservació d'estructura secundària és major que en el conjunt total de les famílies del dataset. Un 47.83% d'aquestes famílies obtenen alguna comparació amb una distància estructural igual o menys de 0.10.

Com hem comentat, hi ha un gran desconeixement sobre aquests elements i s'utilitzen noves tècniques per analitzar-los. Per això, em volgut comparar els resultats obtinguts amb el programa CROSSalign amb els resultats amb CMfinder[9], un altre programa que podem utilitzar per per analitzar la conservació de l'estructura secundària. CMFinder busca un patró de seqüència de nucleòtids amb importància biològica (motifs). El resultat és un arxiu amb format *stockholm* amb la seqüència i el nombre de motifs. Per la

comparació em considerat la presència de motifs com a un indicador de certa conservació de l'estructura secundària. S´ha obtingut els resultats d'un estudi anterior publicat a RNA Biology journal[3] i hem codificat els resultats de CMfinder com 1) presència de motifs i 0) no presència de motifs. Em observat que el software CMfinder a trobat motifs en 324 famílies de lncRNA, aquest fet suposaria que un 63.04% de les famílies dels lncRNA tindrien motius estructurals conservats. Hem realitzat la comparació dels resultats entre els dos programes a nivell de gens i em obtingut el següent resultat (**figura 1|** Diagrama de barres on podem observar-la diferència de resultats positius en conservació d'estructura secundària entre CROSSalign i CMfinder).

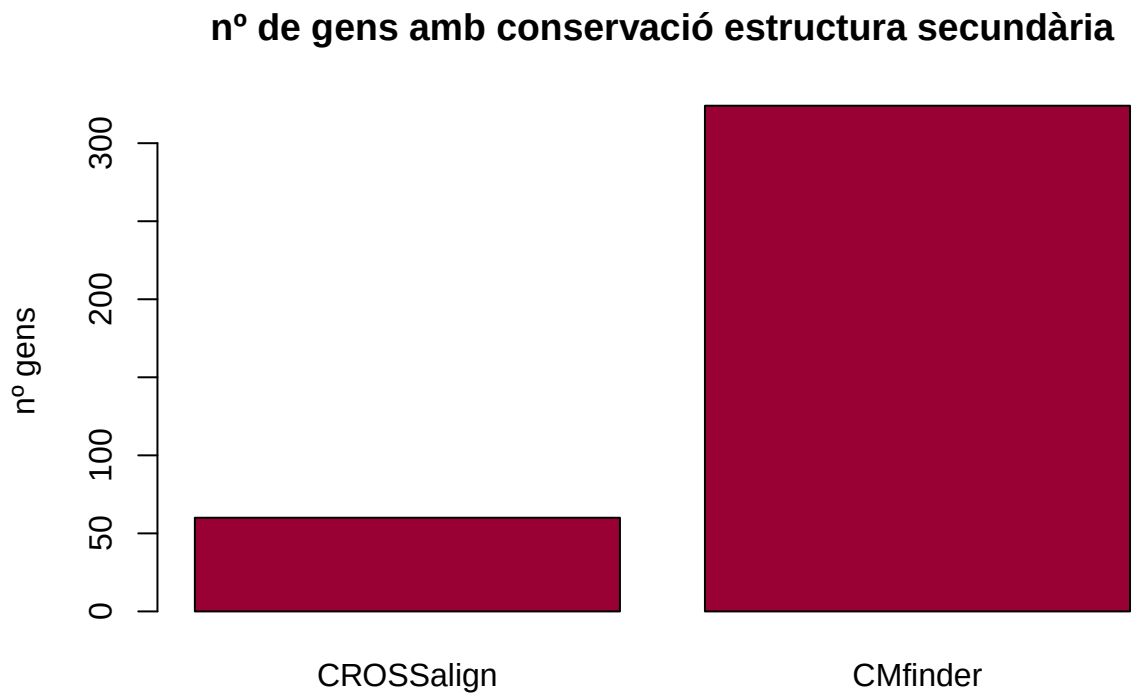

figura 1

Quant comparem els dos programes observem que en 50 de les 509 famílies (un 9.82%) els dos programes coincideixen en determinar que hi ha conservació de l'estructura secundària. D'aquestes, el 22,00% són famílies que són presents en les quatre espècies. Sabem que en el conjunt de dades utilitzat, només el 4.52% de les famílies es troben representades en les quatre espècies, per tant aquestes famílies presenten enriquiment en conservació estructural.

Interacció amb proteïnes

Atès que els pocs lncRNA que hi ha caracteritzats en l'actualitat tenen funció reguladora[1,4,5], em cregut necessari realitzar prediccions d'interacció amb proteïnes per obtenir una visió de la caracterització funcional dels lncRNA en els nematodes. Per avaluar aquesta interacció em utilitzat un altre programa específic, el catRAPID omics[10]. Aquest programa donada un seqüència d'un RNA, retorna una sèrie de valors i un puntuació per ajudar a classificar els resultats. Aquesta puntuació és la suma de tres valors obtinguts en l'anàlisi. Un valor prové de la propensió normalitzada de catRAPID, un altre de presència de dominis d'unió a RNA/DNA i regions desordenades i l'últim de la presència o no de motifs d'unió a RNA. D'aquesta suma es genera un sistema de qualificació que va del 0(mínim) a 3(màxim).

Les seqüències dels lncRNA són provinents del mateix conjunt de dades utilitzat en l'avaluació de conservació d'estructura secondària[3]. Ens vam enfocar en les famílies que eren presents en les quatre espècies i varem anotar aquelles proteïnes que tenen una qualificació de 2 o més (puntuació hight) en el sistema indicat per catRAPID. En total varem realitzar 96 prediccions amb resultat de detecció de 252 interaccions amb 41 proteïnes diferents (**figura 2|** Diagrama de barres on podem observar per cada família el nºde proteïnes amb les quals interaccionen; **figura 3|** Nom de totes les proteïnes que interaccionen amb els lncRNA de les famílies presents en les quatre espècies; **figura 4|** diagrama de barres on podem observar per cada proteïna el nºd'interaccions amb algun membre de les famílies de lncRNA presents en les quatre espècies).

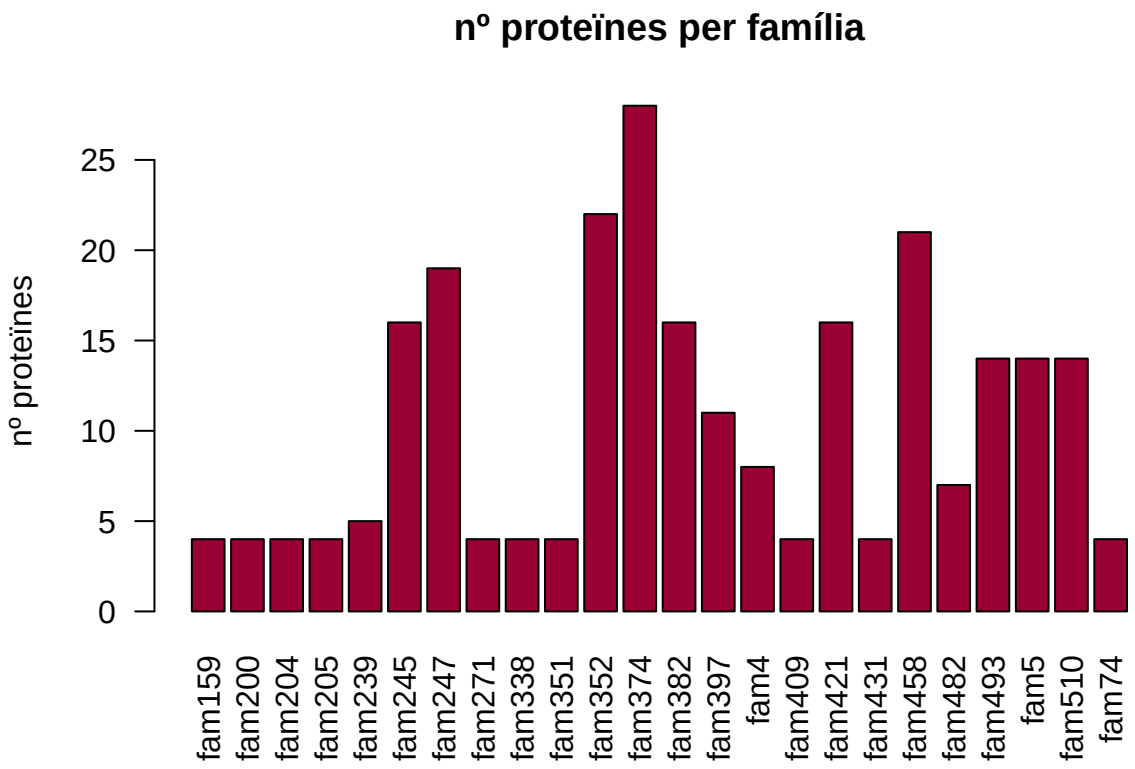

figura 2

Proteïnes obtingudes en l'anàlisi d'interacció

```
##  [1] "SWAP_CAEEL"  "YK27_CAEEL"  "CWC15_CAEEL" "YQOA_CAEEL"  "RSP7_CAEEL"
##  [6] "FBF2_CAEEL"  "YQ4B_CAEEL"  "EIF3G_CAEEL" "ROA1_CAEEL"  "YOT2_CAEEL"
## [11] "NH2L1_CAEEL" "R23A1_CAEEL" "GLH3_CAEEL"  "DIM1_CAEEL"  "RT07_CAEEL"
## [16] "DKC1_CAEEL"  "MCES_CAEEL"  "RL5_CAEEL"   "YLPD_CAEEL"  "CPB3_CAEEL"
## [21] "CPB2_CAEEL"  "RT05_CAEEL"  "IF5A1_CAEEL" "SRP19_CAEEL" "PM14_CAEEL"
## [26] "TRMB_CAEEL"  "DIMQ_CAEEL"  "PGL1_CAEEL"  "GAR1_CAEEL"  "RSP6_CAEEL"
## [31] "NFX1_CAEEL"  "MBL_CAEEL"   "TRM1_CAEEL"  "CTU1_CAEEL"  "RS4_CAEEL"
## [36] "IF4E3_CAEEL" "YS4L_CAEEL"  "RO60_CAEEL"  "EIF3B_CAEEL" "SBDS_CAEEL"
## [41] "GLD1_CAEEL"
```

figura 3

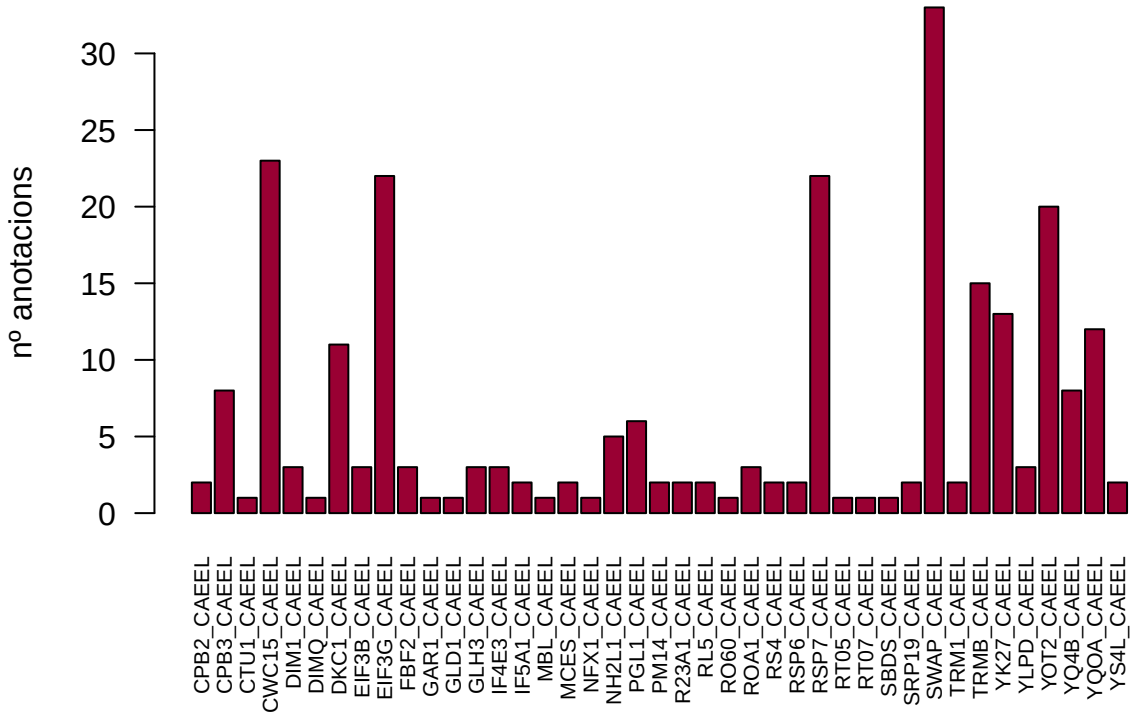

figura 4

Anotació funcional

Per continuar en la caracterització funcional dels nostres lncRNA i amb les proteïnes obtingudes en l'anàlisi d'interacció, varem realitzat un estudi d'anotació funcional. En aquest estudi es van obtenir 402 termes GO relacionats amb les 96 proteïnes que segons catRAPIDomics interaccionen (amb nivell alt) amb els lncRNA del nostre conjunt de dades. En total vàrem anotar uns 402 termes GO, dels quals 154 estan relacionats amb processos biològics (38,3%), 137 relacionats amb funcions moleculars (34,08%) i 111 relacionats amb components cel · lulars (27,61%).

A partir de les anotacions en l'estudi d'anotació funcional podem observar la distribució dels tipus de relacions entre la proteïna i el terme GO obtinguts pel nostre conjunt de dades (**figura 5|** Diagrama de barres on podem observar la distribució segons el tipus de relació entre la proteïna i la funció relacionada amb els GOterms.).

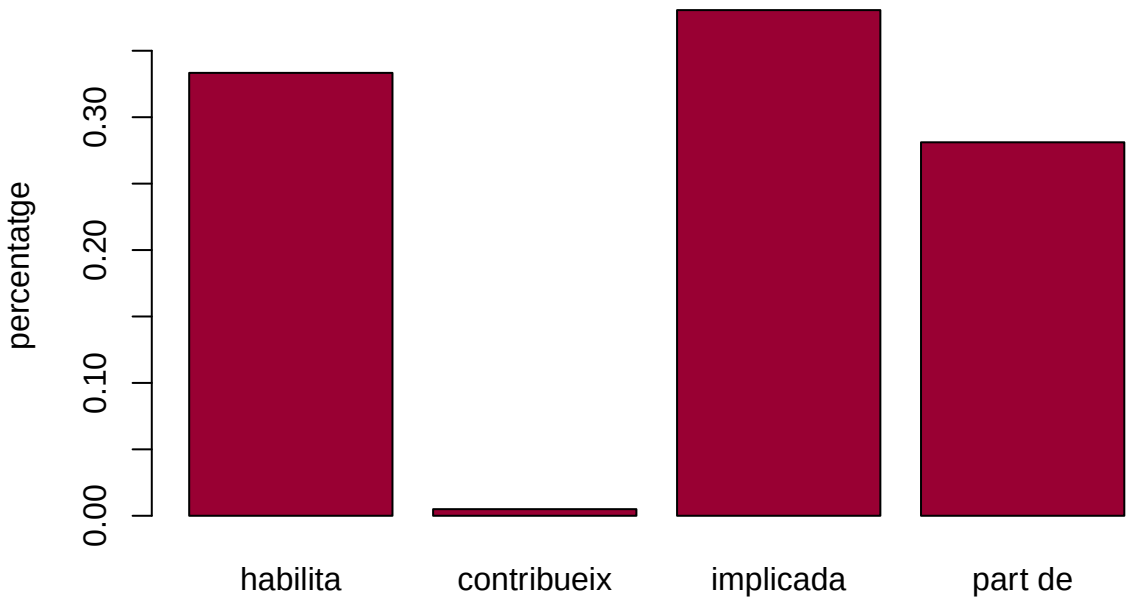

figura 5

Pel nostre estudi ens hem fixat més atentament en els processos biològics en què participen les proteïnes que en les seves funcions moleculars i la seva participació com a components cel · lulars que se'n deriven. Podem veure que hi ha una gran diversitat funcional en les proteïnes amb 85 activitats diferents. Com a resum, en la taula següent (**taula 2|** Resum de les funcions de les proteïnes amb freqüències =>4) podem observar les funcions que apareixen en 4 o més ocasions.

funció	freqüència	prop
translation	11	7.142857
mRNA processing	7	4.545454
mRNA splicing, via spliceosoma	6	3.896104
RNA splicing	6	3.896104
rRNA processing	6	3.896104
methylation	5	3.246753
cell differentiation	4	2.597403
multicellular organism development	4	2.597403
regulation of translation	4	2.597403
ribosome biogenesis	4	2.597403
tRNA processing	4	2.597403

taula 2

Hem observat que un 74.68% de les proteïnes participen en processos relacionats amb la traducció, la regulació de l'expressió gènica i/o la modificació de diferents RNA's, fet que corroboraria amb els que sabem de les

funcions dels lncRNA (**figura 6|** Diagrama de barres don podem veure el nº de proteïnes segons els processos biològics en què participen. Per una millor comprensió s'han agrupat les funcions en grups més funcionals més grans. Traducció: funcions relacionades amb el procés de traducció i biogènesis del ribosoma. Modificació de RNA's: funcions relacionades amb la modificació de tota mena de RNA. Desenvolupament: funcions relacionades amb organogènesis, maduració i processos cel · lulars. Transcripció: funcions relacionades amb el procés de transcripció. Modificació de proteïnes: funcions relacionades amb la modificació de proteïnes).

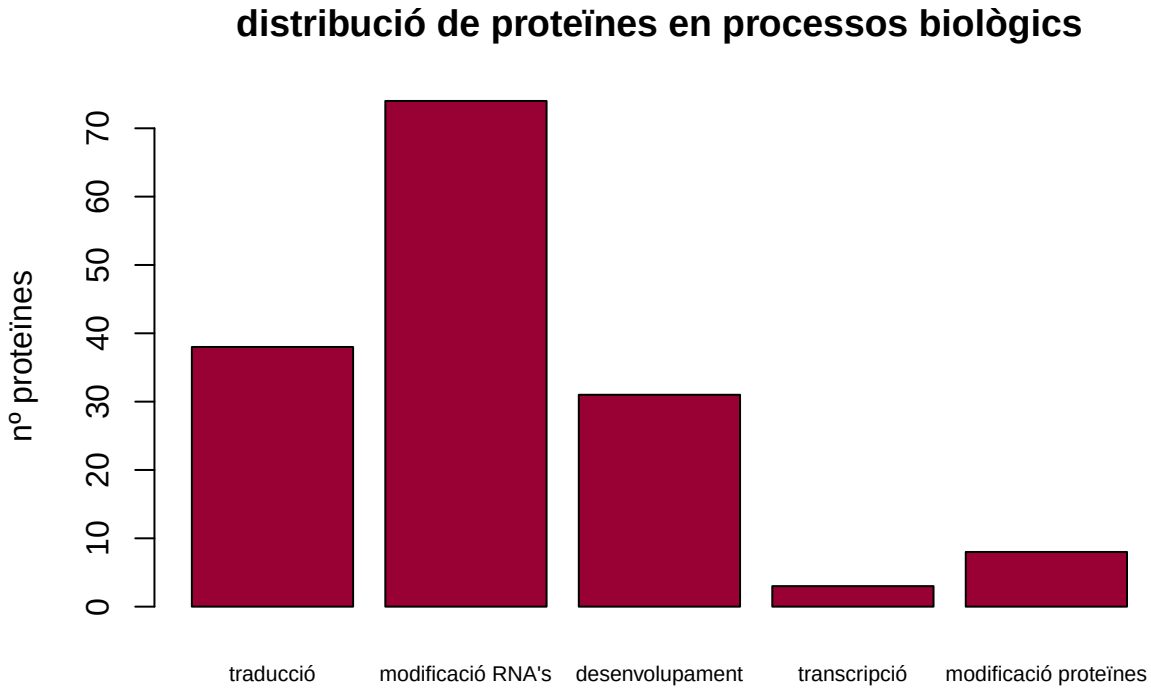

figura 6

Com hem vist els lncRNA que hem estudiat interactuen amb proteïnes que majoritàriament tenen una funció de regulació de l'expressió genètica, ja sigui mitjançant la modificació de RNA's o directament afectant la transcripció i la traducció. Seria lògic que els lncRNA i les proteïnes amb les quals interactuen tinguin posicions properes dins del genoma (actuarien en cis). Per comprovar aquesta afirmació hem comprovat la posició dels nostres lncRNA i de les proteïnes que interactuen per observar si realment es troben localitzats en un mateix segment. En l'anàlisi d'interacció amb proteïnes s'ha utilitzat com a referència per a calcular la interacció l'espècie *Caenorhabditis elegans*, per tant hem decidit realitzar aquest estudi de posició només amb els lncRNA d'aquesta espècie. Hem utilitzat el servidor e!Ensembl[12] amb la versió del genoma *Caenorhabditis elegans WBcel235*. Al finalitzar hem pogut observar que a diferència de la nostra afirmació inicial, la gran majoria de lncRNA actuarien trans. Hem detectat 53 interaccions dels lncRNA d'aquesta espècie amb proteïnes però només un, el **TCONS_00010265cele**, es troba situat al mateix cromosoma del gen que codifica per una de les proteïnes amb les que interactua, **la EIF3G_CAEEL**, però a unes 641209pb de distància (**taula 3|** coordenades dels transcrits *TCONS_00010265cele* i *la EIF3G_CAEEL*).

nom transcrit	tipus	coordenades

nom transcrit	tipus	coordenades
TCONS_00010265cele	lncRNA	II:7801280-7801556
EIF3G_CAEEL	proteïna	II:8442765-8444239

taula 3

Conclusions

Els articles d'Igor Ulitsky[4] i Anne Nitsche/Peter F.Stadler[7] anvers el poc nivell de conservació de les seqüències en la gran majoria de lncRNA, ens obren a la qüestió de la importància i conservació de la seva estructura secondària. En l'article d'Anne Nitsche i Peter F.Stadler[7] podem llegir que la majoria de ncRNA presenten bona conservació de l'estructura secondària, aquest fet xoca amb els resultats que hem obtingut en l'anàlisi amb CROSSalign. Hem obtingut que la gran majoria dels lncRNA de les espècies de nematodes estudiades no seguiria aquest patró i no presentarien una conservació de l'estructura secondària. Però en canvi, en la comparació dels resultats amb CMfinder veiem grans diferències entre els dos programes. CMfinder és un software usat per l'estudi del qual hem obtingut les dades[3] i citat en l'article d'Anne Nitsche i Peter F.Stadler[7], busca patrons de seqüència de nucleòtids amb importància biològica, és a dir busca *motifs*. Per l'article de Nitsche i Stadler[7] sabem que hi ha evidència que en els lncRNA, les regions on l'estructura secondària és important per conservar la funció biològica ocupen una fracció mes petita en la seqüència que en els altres ncRNA. Les nostres divergències, podrien ser degut a aquests fet, només els petits fragments importants per la funció biològica conservarien la seva estructura secondària i els altres fragments haurien divergit entre les diferents espècies de nematodes del nostre conjunt de dades.

Tenim evidències per diversos estudis, que els lncRNA tindrien importància en la regulació de l'expressió genètica. En el nostre estudi hem volgut esbrinar si en el cas dels nematodes, els lncRNA podrien també ser importants en aquest aspecte. L'estudi d'anotació funcional sembla indicar que, en els nematodes estudiats, els lncRNA si que participarien en la regulació de l'expressió. Una gran majoria de les proteïnes amb les quals interacciones actuen o bé directament en el procés de traducció o participarien en els processos de modificació de diferents RNA's. Fins i tot una petita part de lncRNA interaccionaria amb proteïnes que actuarien en la transcripció. També hem vist que una petita part, però no menyspreable, participarien en diferents processos de desenvolupament cel · lular i de teixits, tal com s'havia suggerit en alguns estudis.

En aquest estudi hem vist que la distància dels lcnRNA de l'espècie *Caenorhabditis elegans* amb els gens que codifiquen per les proteïnes amb les quals interacciones és molt elevada. La majoria es troben en cromosomes diferents, només un comparteix posició dins del mateix cromosoma però a una distància molt gran. Aquest fet és contrari a la idea inicial amb la qual hem començat aquest apartat. Aquest fet és contrari a la idea inicial amb la qual hem començat aquest apartat, però ja han sigut descrits exemples de lncRNA que actuarien en trans[13]. Segurament s'hauria d'aprofundir i ampliar l'estudi a les altres 3 espècies.

Materials i mètodes

Avaluació estructura secundària

Hem utilitzat el programa CROSSalign per avaluar la conservació de l'estructura secundària. Aquest programa combina dos algoritmes per predir la similitud estructural entre dos RNA de llargada diferent. El primer d'ells és el CROSS (computacional recognition of secondary structure), un algoritme entrenat mitjançant dades experimentals i amb una resolució de nucleòtid únic, que prediu l'estructura secundària d'un RNA. El segon és un algoritme tipus DTW(Dinamic Time Warping) que permet avaluar la similitud entre dos fragments sense importar les diferències de longituds.

Els creadors van validar el software mitjançant un test amb un conjunt de dades de 22 estructures cristal · logràfiques disponible a NCBI (Lorentz et al., 2016)[15]. Van obtenir les distàncies estructurals entre totes les parelles del dataset i van comparar els seus resultats amb els perfils cristal · logràfics i amb els del programa SHAPE. En els dos casos van obtenir una major correlació.

En aquest estudi per tal d'avaluar la conservació de l'estructura secundària dels lncRNA del nostre conjunt de dades, hem utilitzat aquest programa en la versió local. Hem realitzat comparacions entre totes les seqüències de lncRNA de cada família i hem anotat per cada comparació la distància estructural i el p-valor. Els creadors del programa consideren que un lncRNA té una estructura secundària conservada, quan s'obté un valor de conservació de més de 0.095-0.01.

Interacció amb proteïnes

Per avaluar la interacció amb proteïnes dels lncRNA del nostre conjunt de dades, hem utilitzat un altre programa específic, el catRAPID omics. En aquesta ocasió vam utilitzar la versió online. Aquest programa mitjançant el càlcul de l'estructura secondària, els enllaços d'hidrogen i les forces de van der Walls prediu la interacció RNA-proteïna en models d'organismes concrets. En el nostre estudi es van realitzar totes les anàlisis en l'organisme *Caenorhabditis elegants* que és una de les espècies contribuïdores en el dataset utilitzat.

Aquest programa donada una seqüència d'un RNA, retorna una sèrie de valors i una puntuació per ajudar a classificar els resultats. Aquesta puntuació és la suma de tres valors obtinguts en l'anàlisi. Un valor prové de la propensió normalitzada de catRAPID, un altre de presència de dominis d'unió a RNA/DNA i regions desordenades i l'últim de la presència o no de motifs d'unió a RNA. D'aquesta suma es genera un sistema de qualificació que va del 0(mínim) a 3(màxim).

Les dades que hem utilitzat són les seqüències dels lncRNA presents en les quatre espècies. Per cada lncRNA hem anotat per les proteïnes amb una qualificació alta (2-3), el nom del gen, la família, el nom de la proteïna, el z-score, el *ranking* i un link a la pàgina Uniprot.

Anotació funcional

En la taula dels resultats de catRAPID omics hi ha un link a Uniprot[11] amb la fitxa de cada proteïna. En aquesta fitxa es pot obtenir el llistat de termes GO anotats per la proteïna. Es va anotar per a cada proteïna la idGO, el tipus de procés, el tipus de de participació i una petita descripció de la seva funció.

GOterm

tipus de procés on participa

- biològic
- component cel · lular
- metabòlic

tipus de participació (codificada de 0-3)

- 0 enables (habilita)
- 1 contributes to (contribueix)
- 2 involved in (implicat)
- 3 part of (part de)

breu descripció de la seva actuació

Pel que fa a l'arxiu per comprovar la distància entre els lncRNA i els gens codificadors de les proteïnes hem utilitzat el navegador genòmic de e!Ensembl de l'espècie *Caenorhabditis elegans WBcel235*. Hem visualitzat la situació i anotat les coordenades de les proteïnes i les hem comparat amb la posició dels lncRNA. Els resultats estan codificats de la següent manera:

- 00 -> coordenades molt llunyanes i cromosoma diferent.
- 01 -> coordenades molt llunyanes i mateix cromosoma.
- 10 -> coordenades a distància mitjana i cromosoma diferent.
- 11 -> coordenades a distància mitjana i mateix cromosoma.

Referències

1. Per Johnsson, Leonard Lipovich, Dan Grandér, Kevin V. Morris. Evolutionary conservation of long non-coding RNAs; sequence, 3 structure, function. Biochimica et Biophysica Acta. 2014, 1840;1063-1071.sciencedirect

2. Antonin Morillon. Long Non-coding RNA, The Dark Side of the Genome.sciencedirect

3. Cinta Pegueroles, Susanna Iraola-Guzmán, Uciel Chorostecki, Ewa Ksiezopolska, Ester Saus & Toni Gobaldon. Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus Caenorhabditis RNA Biology. 2019;16:320-329. RNA Biology

4. Igor Ulitsky. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nature Reviews Genetics. 2016, 17;601–614.naturereviews

5. Paschold, A. and Jia, Y. and Marcon, C. and Lund, S. and Larson, N. B. and Yeh, C. T. and Ossowski, S. and Lanz, C. and Nettleton, D. and Schnable, P. S. and Hochholdinger, F. Complementation contributes to transcriptome complexity in maize (Zea mays L.) hybrids relative to their inbred parents. Genome Biolog. 2014;15:R40.RNABiology

6. Jonhy T. Y. Kung, David Colognori, Jeannie T.Lee. Long Noncoding RNAs: Past, Present, and Future. Genetics. 2013 Mar; 193(3): 651-669. ncbi

7. Anne Nitsche, Peter F. Stadler. Evolutionary clues in lncRNA. WIREs RNA 2016.doi: 10.1002/wrna.1376.researchgate

8. Delli Ponti R, Armaos A, Marti S, Tartaglia GG. A Method for RNA Structure Prediction Shows Evidence for Structure in lncRNAs. Front Mol Biosci 2018; 5:1–14.CROSSalign

9. Yao Z, Weinberg Z, Ruzzo WL. CMfinder–a covariance model based RNA motif finding algorithm. Journal Bioinformatics. Volume 22 Issue 4, February 2006:445-452.Bioinformatics

10. Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG. catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. Bioinformatics. 2013, 29;2928-2930. ncbi

11. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D506–D515.academic

12. Sarah E Hunt, William McLaren, Laurent Gil, Anja Thormann, Helen Schuilenburg, Dan Sheppard, Andrew Parton, Irina M Armean, Stephen J Trevanion, Paul Flicek, Fiona Cunningham. Ensembl variation resources Database; Volume 2018.academic

13. Aleksandra E Kornienko, Philipp M Guenz, Denise P Barlow, Florian M Pauler. Gene regulation by the act of long non-coding RNA transcription. BMC Biol.2013;11:59.ncbi

14. Lewin B., 1980. Gene Expression, 2: Eucaryotic Chromosomes John Wiley & Sons, New York. Google Scholar

15. Ronny Lorenz, Dominik Luntzer, Ivo L.Hofacker, Peter F. Stadler, Michael T.Wolfinger. SHAPE directed RNA folding. Bioinformatics. 2016 Jan 1;32(1):145-147.ncbi