



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

III. IKERGAZTE NAZIOARTEKO IKERKETA EUSKARAZ

2019ko maiatzaren 27, 28 eta 29
Baiona, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

OSASUN ZIENTZIAK

**RNA ez-kodetzaile luzeak
(lncRNAk) etxabereetan: RNA-seq
datuetan oinarritutako analisi
transkriptomikoa ardian**

*Martin Bilbao-Arribas, Endika Varela-
Martínez, Naiara Abendaño
eta Begoña M. Jugo*

31-37 or.
<https://dx.doi.org/10.26876/ikergazte.iii.04.04>



RNA ez-kodetzaile luzeak (lncRNAk) etxabereetan: RNA-seq datuetan oinarritutako analisi transkriptomikoa ardian

Bilbao-Arribas, Martin; Varela-Martínez, Endika; Abendaño, Naiara eta Jugo, Begoña M.

*Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila,
Zientzia eta Teknologia Fakultatea
martin.bilbao@ehu.eus, begonamarina.jugo@ehu.eus*

Laburpena

RNA ez-kodetzaile luzeak (lncRNAk), proteinak kodetzen ez dituzten eta 200 nukleotido baino luzeagoak diren transkriptoak, oso gutxi ikertu dira etxabereetan gizakiarekin konparatuz, eta oso mugatua da espezie batetik bestera eratorri daitekeen gene hauen egitura edo funtzioari buruzko informazioa. RNA-seq datu-multzo bat baliatuz ardi lncRNAk identifikatu, kontserbazioa aztertu eta aluminioidun txertoen aurrean erakusten duten jokaera aztertu dugu. Bi ehun desberdinetan fidagarritzat eman dira 2284 eta 3004 lncRNA transkripto. Horietako batzuk posizio genomikoa mantendu dute gertuko espezieekiko, sekuentzia mailako kontserbazio maila altua transkripto gutxitan ikusi bada ere. Gainera, diferentzialki adierazita daude hainbat lncRNA aluminioidun txertoak jaso zituzten animalietan. Analisi gehiago beharrezkoak dira mRNA eta miRNA geneekin dituzten erlazioak argitzeko.

Hitz gakoak: lncRNA, RNA-seq, adierazpen diferentziala, aluminioa, txerto-laguntzaileak

Abstract

Long non-coding RNAs (lncRNAs), transcripts without coding capability and longer than 200 nucleotides, have been poorly studied in livestock comparing with other major model organisms, and the information about structure and function that can be inferred from other species is limited. Using a sheep RNA-seq dataset we identified, performed conservation analyses and profiled the response to aluminium adjuvants of lncRNAs. 2284 and 3004 transcripts have been confidently classified as lncRNAs. Some of them show conserved genomic position relative to close species, even if few had a high level of sequence conservation. Besides, several lncRNAs are differentially expressed in animals that received aluminium-containing vaccines. More analyses are needed to shed light on the relations with mRNAs and miRNAs.

Key words: lncRNA, RNA-seq, differential expression, aluminium, adjuvants

1. Sarrera eta motibazioa

RNAk ez du funtzio bakar gisa DNA eta proteinen arteko mezularia izatea, gene-adierazpenaren eta genomaren antolaketaren doitzaila ere bada. Ugaztunetan ia genoma guztia nonahi transkribatzen da, ez bakarrik mRNAk sortzeko, mota askotako RNA ez-kodetzaileak sortzen dira ere, gene intergenikoetan, gene kodetzaileen kontrako harizpian edo intronetan esaterako (Morris & Mattick, 2014). RNA ez-kodetzaile luzeak (lncRNAk) proteinak kodetzen ez dituzten eta 200 nukleotido baino luzeagoak diren transkriptoak dira. Egitura aldetik askok RNA mezularien antza izan arren, hau da, poliadenilatu egiten dira, 5' txanoa dute eta moztitsasketa jasaten dute (Deveson et al., 2017), lncRNAk adierazpen maila baxuagoa dute, adierazpen ehun- edo garapen-espezifikoa dute eta sekuentzia mailan kontserbazio txikia erakusten dute, salbuespenak badaude ere.

Kontserbazioari dagokionez, lncRNAen sekuentzien eboluzio azkarrak zaildu egiten du lncRNAekin lan egitea eredu ez diren espezieetan, izan ere espezieen arteko ohiko konparaketa filogenetikoak ez dira erabilgarriak (Ulitsky, 2016). Gizakian bilduma zabal eta fidagarri bat dagoen bitartean, abereetan balidatutako gene-eredu gutxi daude. Arrazoi hauek direla-eta, bestelako analisiak erabiltzea beharrezkoa da, adibidez, gene-ordenaren (sintenia), funtziodun azpisekuentzien, adierazpenaren edo RNA molekulen egituren kontserbazioa (Hezroni et al., 2015).

Transkripto hauen funtzioak, oraingoz gutxi batzuk funtzionalki karakterizatu direnak, burutu diren ikerketen arabera lotura izan dezakete gene-adierazpenaren erregulazioaren maila desberdinetan: transkripzioan, transkripzioaren ostekoan, edo kromatinaren antolaketan (Dykes & Emanueli, 2017). RNAk hainbeste egitura desberdin hartzen ahal ditu, ezen, lncRNAk molekula mota askorekin elkarrekintzak izaten ahal baitituzte, eta hauen bitartez euren zereginak gauzatu. Besteak beste, DNA, proteinak edo bestelako RNA motak (miRNAk edo mRNAk kasu) dira RNA ez-kodetzaile hauen ituak. Karakterizazio funtzionalerako, teknika esperimentalez gain, lan-fluxu bioinformatikoak garatu dira. Hauek gene-adierazpenean (adierazpen diferentziala edo ko-adierazpena) eta elkarrekintzen auresatean oinarritzen dira (Cao et al., 2018; Signal et al., 2016).

Oinarritzko sailkapen argi ez-anbiguo baten faltak erronka desberdinak dakartza lncRNAk ongi anotatzeko eta datu transkriptomikoak ulertzeko (Laurent et al., 2016). Azkenaldian lncRNAk klaseetan banatzeko gailendu den hurbilketa beste geneekiko zein kokapen kromosomiko duten aztertzea da, izan ere, kokapenak ekintza-mekanismoei buruzko orokortasun batzuk egitea ahalbidetzen duela ikusi da. Horrela, gene kodetzaileen artean kokatzen direnak lincRNA (RNA ez-kodetzaile intergeniko luzeak) bezala ezagutzen dira. Berriz, gene kodetzaile batetik gertu kokatzen direnak talde propioa osatzen dute. Horietatik promotorea elkarbanatzen dutenei, esaterako, lncRNA dibergente esaten zaie (Luo et al., 2016). Gene kodetzaile baten kontrako harizpian kokatzen direnak kontrako harizpiko lncRNA (*antisense* lncRNA) deitzen dira eta intronetatik eratortzen direnak lncRNA intronikoak dira.

2. Arloko egoera eta ikerketaren helburuak

Etxabereen hainbat ehunetatik lortutako RNA-seq liburutegiak genoma osoko lncRNA bildumak lortzeko erabili izan dira behia, txerria, zaldia edo ardia bezalako espezieetan (1. taula). Bilduma hauetan, milaka lncRNA deskribatzen diren arren, gene bakoitzaren funtzioari edo gaixotasunekin dituzten asoziazioei buruzko informazio gutxi dago oraindik. Lan honetan ardietan egindako esperimentu batetik lortutako RNA-seq datu-multzo bi erabili dira lncRNA berriak identifikatzeko eta tratamendu batekin erlazioa izaten ahal dutenak aztertzeko, lanabes bioinformatikoak erabiliz.

1. taula. Transkriptoma osoko sekuentziazioaren bidez etxabereen lncRNAk bilatzeko egin diren entseguak. Ensembl-ek edo NCBIk informatikoki auresandako geneak ez dira kontutan hartu.

Espeziea	Ehunak	LncRNA kopurua	Egilea
Ardia (<i>Ovis aries</i>)	50+	12296	(Bush et al., 2018)
Ahuntza (<i>Capra hircus</i>)	9	2657	(Bush et al., 2018)
	3	13864	(Foissac et al., preprint)
Behia (<i>Bos taurus</i>)	18	9778	(Koufariotis et al., 2015)
	3	22724	(Foissac et al., preprint)
Zaldia (<i>Equus caballus</i>)	8	20800	(Scott et al., 2017)
Txerria (<i>Sus scrofa</i>)	3	12587	(Foissac et al., preprint)
Oiloa (<i>Gallus gallus</i>)	2	20516	(Kuo et al., 2017)
	2	2193	(Muret et al., 2017)
	3	7502	(Foissac et al., preprint)

Datuak gure taldeak burutu duen ikerketa batetik datoz, zeinetan RNA-seq eta miRNA-seq bidez karakterizatu egin baita aluminio hidroxidoaren eragina txertoen laguntzaile bezala eta bere ondorioak txertaketaren osteko erantzun immunean. Oraingoz sekuentziatu diren ehunak odol-zelula mononuklearrak (Varela-Martínez et al., 2018) eta entzefaloa (artikulu prestakuntzan) izan dira. Laguntzaileen erabilerak historia luzea du txertoen konposizioan, antigorputz erantzunak bultzatuz babes hobea lortzeko. Aluminio konposatuak gehien erabiltzen

diren laguntzaileak dira, dena den, erantzun immunea pizteko darabilten mekanismoa ez da ongi ezagutzen oraindik eta ardietan albo-kalteak ikusi izan dira (Asín et al., 2019). Oso ikerketa gutxi aztertu dute aluminioaren eragina nerbio-sistema zentrolean. Berriki, aluminioarekin tratatu ziren arratoien hipokanpoetan diferentzialki adierazitako mRNAk eta lncRNAk identifikatu dituzte, horien artean glia-zelulekin erlazionaturiko geneak (Xu et al., 2018).

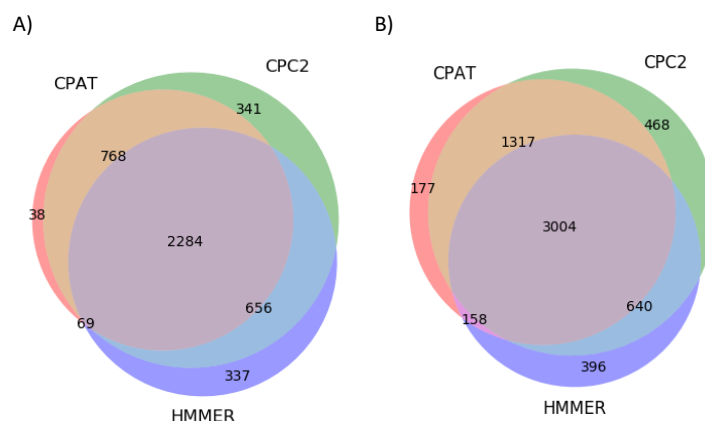
Aurrekari hauek kontuan izanda, analisi honen bitartez lortu nahi den helburua bikoitza da: Batetik, ardiaren RNA ez-kodetzaile luzeen identifikazioa egitea RNA-seq datuetatik eta azterketa deskriptiboa eta funtzionala egitea, ezaugarri orokorrak eta gene-kodetzaileekin duten erlazioa aztertuz. Bestetik, tratamenduarekin lotura duten lncRNAk identifikatzea, hau da, aluminioaren aurrean jokaera aldatzen dutenak.

3. Ikerketaren muina

3.1. lncRNA berrien identifikazioa eta sailkapena

Anotatu gabe dauden transkripto berriak aurkitu ahal izateko, adierazpenaren kuantifikaziorako ohiko prozedura bioinformatikoez gain transkriptoma mihiztatu egin zen eta Ensembl datu-baseko ardi geneekin konparatu. Hainbat iragazki ezarri ziren lncRNAk beste transkriptoetatik bereizteko. Baztertu egin ziren anotatutako geneekin bat egiten zuten transkriptoak eta hautatu egin ziren intergenikoak, intronikoak edo kontrako harizpikoak zirenak. 200 nukleotidoko luzera minimoa ezarri zen transkripto poliexonikoentzako eta 2000koa exon bakarra zutenentzako. Ondoren, proteinak kodetzeko potentziala kalkulatu zen transkripto bakoitzeko hiru programa erabilita: CPAT eta CPC2, ikasketa automatikoa oinarrituta, eta HMMER, sekuentzietan proteina domeinu ezagunak bilatzen dituen (1. irudia). Kodetzaile bezala sailkatutakoak arbuiatu egin ziren. Iragazki guztiak aplikatu eta gero odoleko zelula mononuklearretatik eta entzefalotik eratorritako 2284 eta 3004 transkriptorekin geratu ginen, hurrenez hurren.

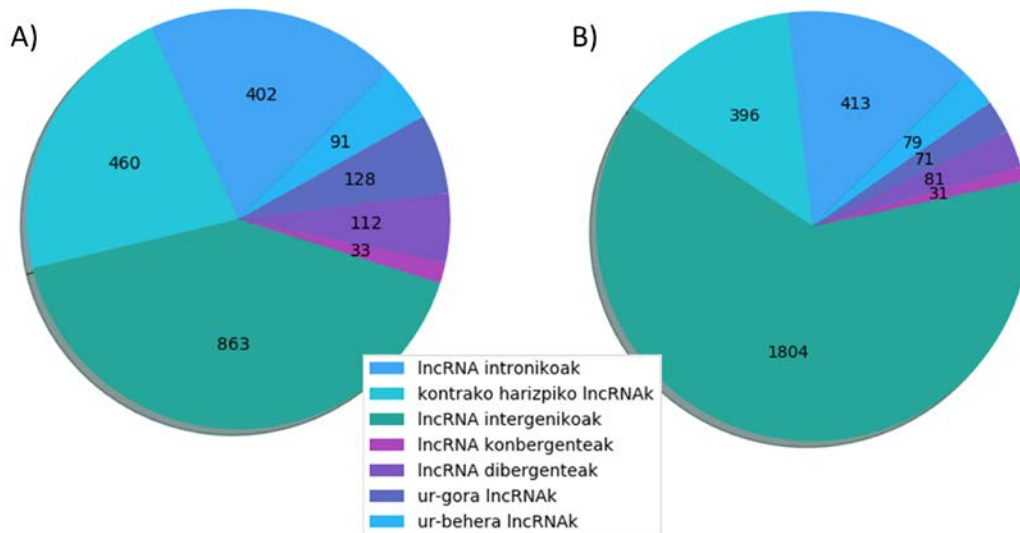
1. irudia. Transkriptoen potentzial-kodetzailearen ebaluazioa hiru tresna erabiliz. A) Odol-zelula mononuklearretan. B) Entzefaloan



lncRNA berrien ezaugarriak antzekoak izan ziren bi ehunetan eta beste ikerketa batzuetan lortutako emaitzekin bat egiten dute. Luzera askotako transkriptoak lortu ziren, tamaina ohikoena 2000 eta 5000 nukleotido artekoa bazen ere, entzefaloan 5000 nukleotido baino gehiagoko transkripto kopuru handiagoa aurkitu zen. Exon bakarreko gehienak 2000-5000 nukleotido zituzten, hau da, luzera txikikoak ziren. 2000 nukleotidotik beherako transkripto monoexonikorik ezin zen egon, analisisik kanpo utzi ziren, sekuentziazioa irakurketa motzeko RNA-seq bat zenez transkripto hauek ezin baitira fidagarritasun handiarekin *de novo* mihiztatu. lncRNA guztiak gertuen duten gene kodetzailearekiko posizio erlatiboaren arabera sailkatu

ziren (2. irudia). Espero bezala, lncRNA intergenikoen klasea gailentzen zen, entzefaloan hauen proportzioa handiagoa bada ere.

2. irudia. LncRNA transkriptoekokapenaren araberrako sailkapena klase desberdinetan. A) Odol-zelula mononuklearretan. B) Entzefaloan

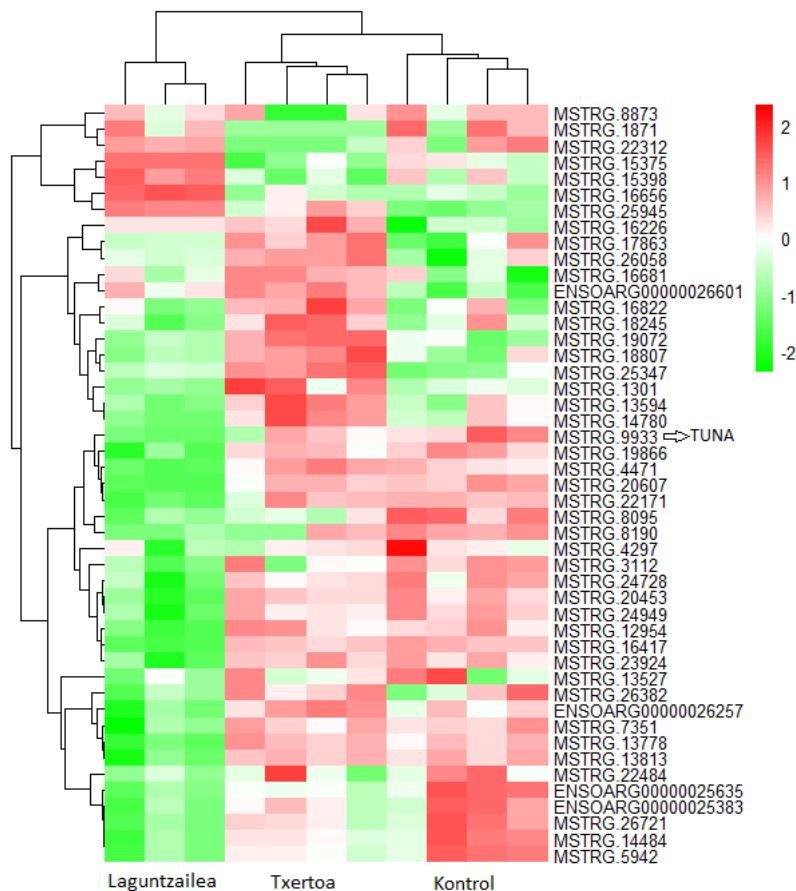


Eraikitako lncRNA-ereduak beste espezie batzuetan agertzen badira, geneok sekuentzia- edo prozesamendu-erroreengatik agertu izanaren probabilitate txikiagoa egongo litzateke eta funtzionalak izateko babes handiagoa izango lukete. Hortaz, sekuentzien kontserbazioa aztertzeko asmoz, Blast tresna erabili zen transkripto bakoitza RNACentral datu basearekin konparatuz. Identitate ona izateaz gain, sekuentziek antzeko luzera izatea bilatu zen. Analisi honek emaitza gutxi eman zituen, izan ere, aipatu bezala, sekuentzia mailako kontserbazioa txikia da lncRNAetan. Hala ere, gizakian edo behian aurkitu diren batzuk bat egin zuten, esaterako NORAD, HCG11, TUNA edo COPG2IT1. Gauzak horrela, sinteniaren analisia burutu zen: behiaren eta gizakiaren genomak gene kodetzaileen ordena ardiaren anotazioarekiko mantentzen diren eremuak identifikatu ziren eta bi espezieetan lncRNArrik zegoen aztertu zen. Ardia eta gizakiaren arteko konparaketan 279 kontserbaturiko bloke sinteniko aurkitu ziren eta ardia eta behiaren arteko konparaketan 871. Ondoren, bloke berdinean kokatzen diren lncRNAekin binakako lerrokaketak egin ziren, 20 nukleotido baino gehiagoko azpisekuentzien identitate perfektuan arreta jarritz. Behiaren konparaketan 39 bat egite lortu ziren eta gizakiarenean bakarra.

3.2. Adierazpen diferentziala

Esperimentu honetan animaliak hiru taldetan bananduta zeuden: kontrol taldea, txerto komertzialak jaso zituzten ardiaren taldea eta laguntzailea soilik jaso zuten ardiaren taldea. Binakako hiru konparaketetan guztira 304 eta 47 lncRNA identifikatu ziren diferentzialki adierazita (DA), odol zelula mononuklearretan eta entzefaloan, hurrenez hurren. 3. irudian entzefaloaren DA lncRNAk erakusten dira. Kopuru hauetan gehituta daude Ensembl-en lncRNA bezala anotatuta dauden eta DA zeuden 28 eta 4 gene. Aipatzekoa da, entzefaloan, behian eta gizakian kontserbatuta dagoen TUNA lncRNA laguntzailea jaso zuten animalietan negatiboki erregulatuta zegoela. lncRNA hau nerbio-sistema zentratean soilik adierazten da eta beharrezkoa da neuronan desberdintzapenerako eta zelula amen pluripotenzia mantentzeko (Lin et al., 2014).

3. irudia. Diferentzialki adierazita dauden lncRNA guztien eta laginen taldekatzea entzefalo ehunean. Berdeak azpiadierazpena adierazten du eta gorriak gainadierazpena. TUNA lncRNA kontserbatua MSTRG.9933 kodeari dagokio.



4. Ondorioak

Lan honetan RNA-seq datu-multzo bat erabilia ardien lncRNAk identifikatu, kuantifikatu eta aluminioaren aurrean duten jokoera aztertu da. Ugaztunen transkriptomak oso konplexuak direla argi geratu da azkenaldian eta mRNA klaseaz haraindi bestelako funtzioak dituzten RNA klaseak garrantzia hartzen ari dira. LncRNA geneen eboluzioa aztertzeko sekuentzietatik haratago doazen hurbiltze berriak beharrezkoak agertu dira, filogenetikoki gertu dauden espezieen arteko konparaketen kasurako ere, behiaren eta ardiaren artekoak adibidez. Azkenik, txertoen laguntzaile bezala erabiltzen den aluminioak dirudenez lncRNA batzuen adierazpena aldatzen du odol zelula mononuklearretan, baina harrigarriagoa dena, entzefaloan ere eragina du.

5. Etorkizunerako planteatzen den norabidea

Eskura ditugun sekuentziazio datuak probestuz lncRNA, miRNA eta mRNA datuak integratzen ari gara. Geneen ko-adierazpen analisien bitartez aztertuko dugu lncRNA hauek zein gene kodetzaileekin batera adierazten diren, lncRNAk miRNA-ituak diren eta miRNA hauek mRNA molekulengan egiten duten errepresioa lncRNA transkriptoen eraginez ahuldu daitekeen. Bestalde, RNA ez-kodetzaile hauen elkarrekintzak auresaten ari gara. Horrela, ardien lncRNA geneen ezagutza zabaldu eta aluminioarekin lotutako hautagaiak identifikatuko dira eurekin analisi funtzionalak egiteko.

6. Erreferentziak

- Asín, J., Pérez, M., Pinczowski, P., Gimeno, M., & Luján, L. (2019). From the bluetongue vaccination campaigns in sheep to overimmunization and ovine ASIA syndrome. *Immunologic Research*. <https://doi.org/10.1007/s12026-018-9059-7>
- Bush, S. J., Muriuki, C., McCulloch, M. E. B., Farquhar, I. L., Clark, E. L., & Hume, D. A. (2018). Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genetics Selection Evolution*, *50*(1), 20. <https://doi.org/10.1186/s12711-018-0391-0>
- Cao, H., Wahlestedt, C., & Kapranov, P. (2018). Strategies to Annotate and Characterize Long Noncoding RNAs: Advantages and Pitfalls. *Trends in Genetics*, *34*(9), 704–721. <https://doi.org/10.1016/j.tig.2018.06.002>
- Deveson, I. W., Hardwick, S. A., Mercer, T. R., & Mattick, J. S. (2017). The Dimensions, Dynamics, and Relevance of the Mammalian Noncoding Transcriptome. *Trends in Genetics*, *33*(7), 464–478. <https://doi.org/10.1016/j.tig.2017.04.004>
- Dykes, I. M., & Emanuelli, C. (2017). Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. *Genomics, Proteomics and Bioinformatics*, *15*(3), 177–186. <https://doi.org/10.1016/j.gpb.2016.12.005>
- Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*, *11*(7), 1110–1122. <https://doi.org/10.1016/j.celrep.2015.04.023>
- Koufariotis, L. T., Chen, Y. P. P., Chamberlain, A., Jagt, C. Vander, & Hayes, B. J. (2015). A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS ONE*, *10*(10), 1–24. <https://doi.org/10.1371/journal.pone.0141225>
- Kuo, R. I., Tseng, E., Eory, L., Paton, I. R., Archibald, A. L., & Burt, D. W. (2017). Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. *BMC Genomics*, *18*(1), 323. <https://doi.org/10.1186/s12864-017-3691-9>
- Laurent, G. S., Wahlestedt, C., & Kapranov, P. (2016). The Landscape of long non-coding RNA classification Georges. *Trends in Genetics*, *31*(5), 239–251. <https://doi.org/10.1016/j.tig.2015.03.007>
- Lin, N., Chang, K. Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., Zhang, D., Han, T., Yang, C. S., Cunningham, T. J., Head, S. R., Duester, G., Dong, P. D. S., & Rana, T. M. (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Molecular Cell*, *53*(6), 1005–1019. <https://doi.org/10.1016/j.molcel.2014.01.021>
- Luo, S., Lu, J. Y., Liu, L., Yin, Y., Chen, C., Han, X., Wu, B., Xu, R., Liu, W., Yan, P., Shao, W., Lu, Z., Li, H., Na, J., Tang, F., Wang, J., Zhang, Y. E., & Shen, X. (2016). Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell*, *18*(5), 637–652. <https://doi.org/10.1016/j.stem.2016.01.024>
- Muret, K., Klopp, C., Wucher, V., Esquerré, D., Legeai, F., Lecerf, F., Désert, C., Boutin, M., Jehl, F., Acloque, H., Giuffra, E., Djebali, S., Foissac, S., Derrien, T., & Lagarrigue, S. (2017). Long noncoding RNA repertoire in chicken liver and adipose tissue. *Genetics Selection Evolution*, *49*(1), 1–17. <https://doi.org/10.1186/s12711-016-0275-0>
- Scott, E. Y., Mansour, T., Bellone, R. R., Brown, C. T., Mienaltowski, M. J., Penedo, M. C., Ross, P. J., Valberg, S. J., Murray, J. D., & Finno, C. J. (2017). Identification of long non-coding RNA in the horse transcriptome. *BMC Genomics*, *18*(1), 1–11. <https://doi.org/10.1186/s12864-017-3884-2>

- Signal, B., Gloss, B. S., & Dinger, M. E. (2016). Computational Approaches for Functional Prediction and Characterisation of Long Noncoding RNAs. *Trends in Genetics*, 32(10), 620–637. <https://doi.org/10.1016/j.tig.2016.08.004>
- Ulitsky, I. (2016). Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics*, 17(10), 601–614. <https://doi.org/10.1038/nrg.2016.85>
- Varela-Martínez, E., Abendaño, N., Asín, J., Sistiaga-Poveda, M., Pérez, M. M., Reina, R., De Andrés, D., Luján, L., & Jugo, B. M. (2018). Molecular signature of aluminum hydroxide adjuvant in ovine PBMCs by integrated mRNA and microRNA transcriptome sequencing. *Frontiers in Immunology*, 9(OCT), 2406. <https://doi.org/10.3389/fimmu.2018.02406>
- Xu, Y., Zhang, H., Pan, B., Zhang, S., Wang, S., & Niu, Q. (2018). Transcriptome-Wide Identification of Differentially Expressed Genes and Long Non-coding RNAs in Aluminum-Treated Rat Hippocampus. *Neurotoxicity Research*, 34(2), 220–232. <https://doi.org/10.1007/s12640-018-9879-1>

7. Eskerrak eta oharrak

Lan hau Nafarroako Agrobioteknologia Institutuarekin eta Zaragozako Unibertsitateko Albaitaritza Fakultatearekin batera dugun proiektuari (MINECO: AGL2013-49137-C3-3-R, AGL2013-49137-C3-2-R eta AGL2013-49137-C3-1) esker burutu da. UPV/EHUko doktorego aurreko kontratua dute M. Bilbao-Arribas-ek (PIF17/306) eta E. Varela-Martínez-ek (PIF15/361), eta UPV/EHUko doktorego ondoko kontratua N. Abendaño-k (ESPDOC16/43).