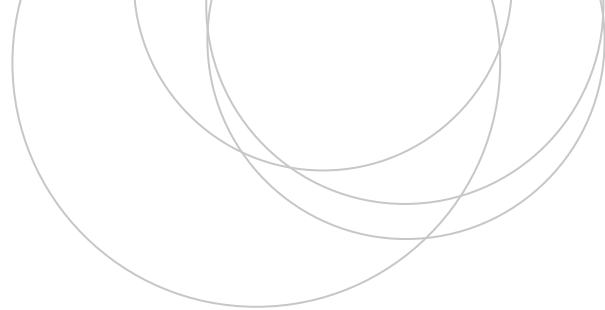




Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

ZIENTZIA
ETA TEKNOLOGIA
FAKULTATEA
FACULTAD
DE CIENCIA
Y TECNOLOGÍA



Gradu Amaierako Lana
Biologiako Gradua

Ardiaren ehun desberdinetan karakterizatutako eta kuantifikatutako miRNAen adierazpenaren banaketaren azterketa

Egilea:

Aitor Guisasola Serrano

Zuzendaria:

Begoña Marina Jugo

© 2021, Aitor Guisasola Serrano

Leioa, 2021ko ekainaren 17a

AURKIBIDEA

LABURPENA	1
1. SARRERA	2
1.1. MiRNAen biogenesis	2
1.2. MiRNAen adierazpen analisia	4
1.3. Analisi bioinformatikoa	5
1.4. MiRNAen ikerketa etxe-abereetan	6
2. HELBURUAK	7
3. MATERIALA ETA METODOAK	7
3.1. Laginen hautaketa	7
3.2. Laginen kalitate kontrola eta preprozesaketa	7
3.3. MiRNAen karakterizazioa eta kuantifikazioa	8
3.4. Laginen arteko antzekotasun analisia	9
3.5. MiRNAen banaketaren eta ehun espezifikotasunaren azterketa	10
4. EMAITZAK	11
4.1. Bildutako datuen deskribapena	11
4.2. Sekuentzien preprozesaketa eta karakterizazioa	12
4.3. Laginen arteko antzekotasun analisia	13
4.4. MiRNAen adierazpenaren banaketa eta ehun espezifikotasuna	15
5. EZTABAIDA	17
6. ONDORIOAK	20
7. ESKER ONAK	21
8. BIBLIOGRAFIA	21

LABURPENA

MicroRNak (miRNak) 22 base-pare inguruko RNA ez-kodetzaile txikiak dira eta prozesu biologiko gehien erregulazioan parte hartzen dute. Etxe-abereekin lan egitean miRNak oso erabilgarriak izan daitezke osasun egoeraren bioindikatzailerik bezala edo hazkuntza programetan intereseko fenotipoen hautaketarako markatzaile bezala. Ardian ehun desberdinetako miRNAen adierazpena ikertu den arren ez da ehun asko biltzen dituen lan integraturik egin eta lan mota horiek garrantzitsuak dira miRNA ehun espezifikoak identifikatzeko eta ehunetako miRNAen adierazpenaren ikuspegi globalago bat izateko. Hori dela eta, lan honetan datu-baseetan aurkitutako ardiaren 15 lan desberdinetako 152 miRNA-seq lagin gordin bildu eta 19 ehunetan sailkatu ditugu. MiRNA berriak aurrean eta ezagunekin batera kuantifikatu ditugu miRDeep2 programaren bidez. Prozesu eta analisi guztiak programa informatiko, metodologia eta genoma berdinarekin burutu dira, datuak era konparagarri eta uniformeaz aztertu ahal izateko. MiRNA ehun espezifikoak identifikatzeko tau izeneko ehun espezifitate indizea (EEI) erabili dugu. Aztertutako ehunen miRNAen adierazpenaren portzentajerik handiena miRNA gutxi batzuek hartzen dutela eta, batzuetan, miRNA bakar batek har dezakeela ikusi dugu. Baita ere ehunetako miRNA adierazienek ziklo zelularren erregulazioan parte hartzen dutela ikusi dugu eta, orokorrean, ehun espezifitasun baxua izan dutela. Ehun askotan miRNA adieraziena mir-143 izan da, ziklo zelularren erregulazioan parte hartzen duena eta miRNA adierazien artean EEI baxuena izan duena. Horrez gain, identifikaturiko miRNA ehun espezifiko batzuk beste espezie batzuetan ere miRNA ehun espezifiko bezala identifikatu direla ikusi da. Bestalde, lan desberdinetan erabilitako metodoak edo ardi arrazak ehun jakin bateko miRNAen adierazpenaren desberdintasun nabariak eragin ditzakeela ikusi dugu eta datu-base eguneratu eta fidagarrien eskuragarritasunak duen garrantzia agerian geratu da. Azkenik, lan honek identifikaturiko miRNAek ehun desberdinetan zehar duten adierazpen profila erakutsi du lehen aldiz ardian.

ABSTRACT

MicroRNAs (miRNAs) are a class of small noncoding RNAs of about 22 base pairs that play multiple roles in regulating most biological processes. MiRNAs can be very useful in livestock as animal health biomarkers or as markers of phenotypes of interest during the process of genomic selection in breeding programs. Although studies have been carried out on the expression of miRNAs in different tissues of the sheep, there has not been a study that integrates many tissues of the sheep and this type of work is very important to identify tissue-specific miRNAs and to have a global approach of the expression of miRNA across tissues. Therefore, in this work we have collected 152 miRNA-seq raw samples of 15 different works from databases and classified them into 19 tissues. We have predicted novel miRNAs and quantified them together with known miRNAs using miRDeep2. All processes and analyses have been done using the same computer programs, methodology and genome, to analyze data in a comparable and uniform way. To identify the tissue-specific miRNAs we have used a tissue specificity index (TSI) called tau. We have seen that the expression in the tissues we have studied is dominated by a few miRNAs, with a single miRNA dominating in some tissues. We have also seen that the more expressed miRNAs have a role in cell cycle regulation and generally present a lower TSI. MiR-143 has been the most expressed miRNA in various tissues, which has a role in cell cycle regulation and has presented the lowest TSI. Furthermore, some of the tissue-specific miRNAs identified in this work have also been identified as tissue-specific miRNA in other species. On the other hand, we have seen that the methods or sheep breeds used in different works can cause differences in miRNAs expression in a given tissue and the importance of

accessibility to updated and reliable databases has been seen. Lastly, this work has shown for the first time the expression profile of the identified miRNAs across different tissues in sheep.

1. SARRERA

MicroRNak (miRNak) 22 base-pare inguruko RNA ez-kodetzaile txikiak dira eta transkripzio osteko gene adierazpena erregulatzen dute (Bartel, 2004). Lehen miRNA, Lin-4, 1993 urtean aurkitu zen *Caenorhabditis elegans* nematodoan. Lin-4 RNak LIN-14 proteina kodetzen duen RNA mezulariaren (mRNA) 3' UTRrekin lotu eta proteinaren ekoizpena erregulatzen duela ikusi zuten (Lee *et al.*, 1993).

Hasieran, ez zen Lin-4 bezalako RNarik ezagutzen, baina let-7ren aurkikuntzak adierazi zuen ez zela salbuespen bat. Are gehiago, let-7 gizakian eta beste zenbait organismotan aurkitu da (Hammond, 2015). Gaur egun, ezagutzen da miRNak landare eta animalien erregulazioan eginkizun garrantzitsu bat dutela eta estimatu da proteinen adierazpenaren herena erregulatzen dutela (Bartel, 2004). Horrez gain, ikusi da animalietan eta landareetan aurkitutako miRNA asko eboluzioan zehar kontserbatuak daudela (Bartel, 2018)

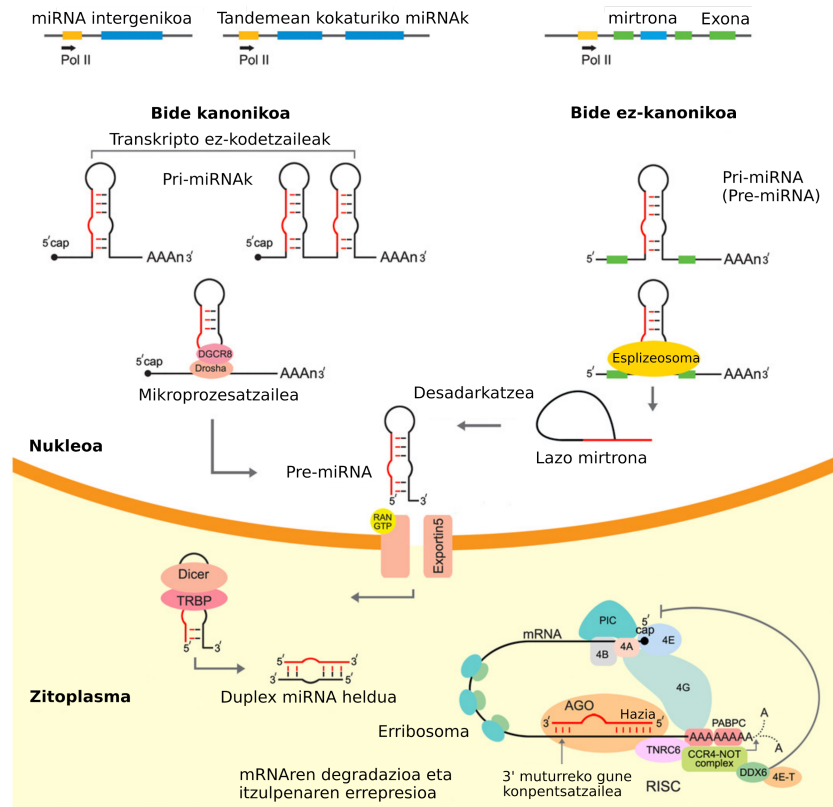
MiRNAen genomako kokapena aztertzean ikus daiteke intronen barruan (mirtronak) edo exon ez-kodetzaileetan (miRNA intergenikoa) aurki ditzakegula (Saliminejad *et al.*, 2018). MiRNA intergenikoak ez dira agertzen beste geneetatik gertu, beraz, promotore propioa duten transkripzio unitateak dira (Bartel, 2004). MiRNAren portzentaje txiki bat tandemean kokatuta daude, bata bestearen ondoan, eta transkripzio unitate bakar bat bezala transkribatzen dira. Tandemean kokaturiko miRNA horiei miRNA polizistronikoa deritze (Ozsolak *et al.*, 2008).

1.1. MiRNAen biogenesisia

II polimerasak (pol II) miRNAen portzentaje handi bat transkribatzen du eta sortzen den lehen RNA transkripto horri pri-miRNA deritzo (Saliminejad *et al.*, 2018). 1. irudian ikus daiteken bezala, Pri-miRNA horiek gutxienez urkila bat dute, 5' muturrean 5'-kapela sekuentzia daramate eta 3' muturrean zenbait adenina monofosfatoz osaturiko isatsa (poly-A isatsa) eraman dezakete (Bartel, 2018). Mikroprozesatzaile konplexu batek pri-miRNA zatitzen du, gutxi gora behera 70 nukleotido dituen eta pre-miRNA deritzon begizta formako RNA molekula bat sortuz (García-Lopez *et al.*, 2013). Prosezu hau nukleoan gertatzen da eta mikroprozesatzaile konplexua Droscha eta DGCR8 izeneko bi azpiunitatez osatuta dago. Droscha endorribonukleasa bat da eta kate bikoitzeko RNA mozten du. DGCR8, ordea, ez du erreakzio katalitikorik egiten eta kate bikoitzeko RNarekin lotzen da. Sortzen den pre-miRNA nukleotik zitoplasmara garraiatzen da exportin 5 (XPO5) proteinaren bidez (Saliminejad *et al.*, 2018). Zitoplasman prosezaketa-konplexu batek pre-miRNA mozten du begiztaren alboan eta 22 base-pare (bp) inguruko kate bikoitzeko miRNA heldua (duplex miRNA heldua) sortzen da (Bushati *et al.*, 2007). Kate bikoitzeko RNarekin lotzen den TAR proteinak (TRBP) eta Dicer endorribonukleasak prosezaketa-konplexua osatzen dute. Duplex miRNA heldua proteina argonauta (AGO) batekin lotzen da eta RNA kateetako bat (harizipi bidaiaria) galtzen du, kate bakarreko miRNA heldu bat sortuz (Bartel, 2018). AGO proteinak RISC deritzon konplexu erribonukleoproteiko baten eraketa indutzen du eta horrek, miRNA helduak gidatuta, RNA mezulariaren (mRNA) transkripzio osteko isilpena eragiten du (Saliminejad *et al.*, 2018).

Orain arte azalduko miRNAen biogenesi bidea ohikoena izan arren badira ez-kanonikoak diren bide alternatibo batzuk. Bide alternatibo horietan bide kanonikoko zenbait urrats egin gabe miRNA helduak sortzen dira. Mirtronak, adibidez, Droscha eta DGCR8 gabe sortzen dira. Horien ordez esplizeosomak pri-miRNak

prozesatzen ditu lazo mirtrona izekeko begizta formako egitura bat sortuz. Lazo mirtrona desadarkatu egiten da eta sortzen den pre-miRNAk bide kanonikoa jarraitzen du (Westholm *et al.*, 2011). Landareetan Drosha eta Dicer endorribonukleasen ordez DCL familiako endonukleasak sortzen dute duplex miRNA heldua (Rogers *et al.*, 2013).



1. irudia: tandemean kokaturiko miRNAen, miRNA intergenikoaren eta mirtronen biogenesiaren diagrama. Zelula nukleo barruan eta zitoplasman gertatzen diren prozesuak bereizten dira (Saliminejad *et al.*-etik moldatua, 2018).

RISC konplexuak gene baten adierazpena erregulatzeko mRNA itua identifikatu behar du. Horretarako, miRNAek hazia deitzen den 2-7 base pareko sekuentzia bat dute. Hazia itu mRNAren 3' UTRko nukleotidoekin lotzen da (Bartel, 2009). Haziaz gain miRNAek mRNAekin parekatzen den gune konpentsatzaile bat izan dezakete. Gune konpentsatzaile horrek parekatzearen espezifotasuna eta afinitatea emendatzen du (Friedman *et al.*, 2009). RISC konplexuak bi mekanismo erabil ditzake mRNA isiltzeko: itu mRNAren degradazioa edo itzulpenaren errepresioa. MiRNA eta mRNAren arteko osagarritasunak bi mekanismoetatik zein gertatzen den determinatzen du. Osagarritasun perfektua edo altua gertatzean Ago 2 endonukleasak mRNA moztuko du 10 eta 11 hondarren artean eta mRNAren degradazioa eraginez. Mekanismo hau oso arrunta izaten da landareetan (Pillai, 2005). Animalietan, ordea, osagarritasun partziala gertatzen da orokorrean, proteina argonauta zitoplasmako gorputz prozesatzaileen barruan (p-gorputzetan) barneratzen da eta bertan itzulpenaren errepresioa gertatzen da. P-gorputzen barnean mRNA desegonkortzeko hainbat mekanismo ezagutzen dira, horietako bat da poly-A isatsaren kentzea CCR4-NOT deadenilasa konplexuaren bitartez (Hammond, 2015).

Batzuetan miRNA helduaren 3' eta 5' muturretan nukleotido gutxi batzuk gehiago edo gutxiago ager daitezke. MiRNA heldu batetatik eratorritako isoforma horiei IsomiR deritze. IsomiR gehienak Drosha eta Dicer

endorribonukleasen mozketa zehaztugabearengatik sortzen dira. IsomiR-ek miRNA-en egonkortasuna handitu, RISC konplexuan kargatzeko ahalmena handitu eta miRNA-en eraginkortasuna aldatu dezakete. Beraz, ez dira zoriz sorturiko miRNA isoforma hutsak eta arrazoi fisiologikoengatik sor daitezke. IsomiR baten lehentasuna zelularen arabera alda daiteke eta gaixotasun batek isomiR baten lehentasuna eragin dezakete (Hammond, 2015).

1.2. MiRNA-en adierazpen analisia

MiRNA-en ikerketa gero eta garrantzi gehiago hartzen ari da, izan ere, gero eta diagnostiko molekular gehiago egiteko erabiltzen ari dira miRNA-ak eta gaixotasun askorekin harremandu dira. Hori dela eta, miRNA-en adierazpen analisia egiteko metodo sendoak ezagutzea funtsezkoa da. Ondo finkatutako hiru ikuspegi desberdin ezagutzen ditugu miRNA-en adierazpen analisia egiteko: Alderantzizko PCR kuantitatiboa (qRT-PCR), microarray-ak eta miRNA-en sekuentziak (miRNA-seq) (Pritchard *et al.*, 2012).

RT-qPCR-a metodo espezifiko eta sentikorra da eta era erraz batean txertatu daiteke laborategi baten eguneroko lanean. DNA osagarria (cDNA) lortzeko bi estrategia erabiltzen dira: poly-A isatsa gehitzea edo begizta formako hasle bat erabiliz alderantzizko haslea lotzeko gunea sortzea (Pritchard *et al.*, 2012). Microarray-a hibridazioan oinarrituriko errendimendu altuko teknika da eta milaka zunda dituen euskarri solido batean oinarritzen da. Teknika merkea da, baina harreman estuko miRNA familia asko nukleotido bakar baten desberdintzen dira eta, beraz, espezifikotasun baxuko teknika da (Saliminejad *et al.*, 2018).

Microarray-aren eta RT-qPCR-aren ahultasunak denborarekin sendotzen joan diren arren bi teknikek oinarritzko desabantaila bat azaltzen dute: ezin dira miRNA berriak (novel) identifikatu (Gao *et al.*, 2016). Hasieran uste zena baino miRNA gehiago aurkitu dira eta oraindik aurkitu gabeko miRNA asko daudela estimatzen da. Horregatik, garrantzitsua da miRNA berriak detektatu ditzakeen metodo bat erabiltzea. MiRNA ezagunen eta berrien adierazpen analisia egiteko oso erabilgarriak dira Next Generation Sequencing (NGS) teknikak (Bortolomeazzi *et al.*, 2019).

NGS teknologia 2004 urtean garatzen hasi zen eta DNA edo RNA laginak prestatzeko, amplifikatzeko, sekuentziatzeko eta ensanblatzeko edo genomarekin eratzeko estrategia multzoa da. RNA eta DNA era berean sekuntziatzen dira, baina RNA sekuentziak (RNA-seq) egitean cDNA lortu behar da (Metzker, 2009). Beste RNA-seq teknikekin alderatuta miRNA-en sekuentziak (miRNA-seq) egitean lagina tratatu behar da RNA txikiak (sRNA) isolatu eta aberastu ahal izateko. Izan ere, miRNA-ek RNA masa totalaren zati txiki bat (%1 inguru) osatzen dute (Pritchard *et al.*, 2012). Horrez gain, bere tamaina txikiagatik, sRNA liburutegia prestatzean adaptadoreak (3' edo 5' muturretan) edo poly-A isatsak (3' muturrean) gehitu behar zaizkie (McCormick). Liburutegia prestatu eta gero lagina sekuentziatu egiten da. Sekuentziatzean kontuan hartu behar den faktore bat sakontasuna da. Sakontasuna sekuentziak prosezuan zehar nukleotido bakoitzaren irakurketa kopuruaren batezbestekoa da. MiRNA askok adierazpen maila baxuak dituzte eta sekuentziak sakonak behar dira horien adierazpena aztertzeko. Are gehiago, sekuentziak sakontasuna estuki harremandu dago adierazpen maila erlatiboaren azterketarekin (Saliminejad *et al.*, 2018). Gaur egun NGS teknika ugari ezagutzen dira, baina guztietatik sakontasun onena duena eta miRNA-ak aztertzeko erabili dena Illumina teknologia da (Fromm, 2016).

1.3. Analisi bioinformatikoa

MiRNA-seq teknikari esker sRNA ugari sekuentziazten dira, baina irakurketa guzti horietatik asko ez dira benetako miRNAk (Chiang *et al.*, 2010). Gaur egun miRNAen auresatea eta horien adierazpenaren analisia ordenagailuen eta software espezializatuaren zeregina da, hau da, bioinformatikarena. Erabiliko diren programek ezagutu behar dituzte miRNAen egituraren ezaugarri espezifikoak miRNAk auresan ahal izateko. Ezaugarri horietako batzuk ondo finkatuta daude: 22 nukleotido inguruko sekuentziak dira, genomak kodetuta daude, urkila formako prekursore batetik datoz, duplex miRNAren bi kateak adierazten dira eta espezieetan zehar kontserbatuak daude (Pritchard *et al.*, 2012). Badira bestelako ezaugarriak, baina ez dira kontuan hartzen duela gutxi aurkitu berri direlako eta animali talde konkretuetan soilik behatu direlako (Fromm, 2016).

MiRNA-seq-etik lorturiko datu gordinekin burutu beharreko lehen urratsa da liburutegiak prestatzean gehituriko adaptadorea kentzea (trimming). Urrats honetan, adaptadorea kentzeaz gain, 22 nukleotido ingurukoak ez diren sekuentzia guztiak baztertzen dira (Fromm, 2016). Datu gordinak eta prozesatuak Fastq formatuan daude eta sekuentziez gain nukleotido bakoitzak kalitate puntuazio bat du. Hori dela eta, gomendagarria izaten da trimming egin aurretik eta ondoren kalitate kontrola egitea, irakurketa guztien kalitatea era grafiko batean laburtzen duen programa bat erabiliz (Andrews, 2010).

Ondoren, adaptadorerik gabeko laginak aztertu behar dira miRNAk detektatzeko eta horien adierazpena kuantifikatzeko programa espezializatuekin. MiRNAen aurkikuntza eta karakterizazioarako hainbat programa erabil daitezke, baina erabiliena miRDeep2 da (Bortolomeazzi *et al.*, 2019). MiRDeep2-k, beste programa gehienek bezala, hasteko, irakurketak genomaren aurka erkatzen ditu (mapeoa) (Friedländer *et al.*, 2011). MiRNAen ezaugarri bat genomak kodetuta egotea denez, mapeoari esker ezaugarri hori betetzen ez duten irakurketak baztertzen dira (Fromm, 2016).

Azkenik, miRDeep2-k miRNA berriak auresan eta miRNA ezagunekin batera kuantifikatzen ditu. Programek miRNA berriak auresateko prozedura desberdinak egin ditzakete (Bortolomeazzi *et al.*, 2019). MiRDeep2-k novelak auresateko hurrengoa egiten du: mapeoan agerturiko locus posibleetatik miRNA prekursore posibleak hautatzen ditu. Ondoren, RNAfold-en bidez, miRNA prekursore horien RNAREN egitura sekundarioen energia askea kalkulatu du. Azkenik, aurreko informazioa kontuan hartuta, prekursore bakoitza puntuatu egiten du eta benetako prekursore bat izateko probabilitatea kalkulatu du (Friedländer *et al.*, 2011).

Kuantifikazioaren ostean, aztertutako lagin bakoitzeko miRNA ezagunen eta auresandako miRNAberrien adierazpena lortzen da. Lagin ugari izatekotan garrantzitsua da datuak normalizatzea, sekuentziazioaren sakontasunaren eragina ezabatzeko (Pritchard *et al.*, 2012). Adierazpen normalizatuarekin datuen analisi estatistikoa egin daiteke.

Auresandako eta ondo karakterizaturiko miRNA berrien sekuentziak, nomenklatura eta anotazio egokiarekin datu baseetan gorde behar dira, etengabe miRNAk auresaten egon behar ez izateko. MiRBase miRNAen biltegi publiko erabiliena da. 2002 urtetik dago erabilgarri eta konfiantza altuko datu basea da, izan ere, bikotekako errebisia (peer-review) egiten duten aldizkarietan publikatutako artikuluetan identifikaturiko sekuentziak gordetzen dira soilik (Kozomara *et al.*, 2013). Hala ere, MiRBase-ren azken argitaratzea 2018an izan zen v22 bertsioarekin eta gizakian 2654 miRNA heldu gordeta dauden arren, bestelako ugaztunetan, ardian adibidez, 153 miRNA heldu daude soilik (www.mirbase.org). Hori arazo bat da, izan ere, NGS teknikei esker gero eta miRNA

berri gehiago aurkitzen dira (Saliminejad *et al.*, 2018) eta miRNAen erregistro estandarizaturik ezean, ikerketa desberdinetan anotazio desberdina jasotzen duten miRNA berri berdinak aurkitzea gerta daiteke, ezagutzan aurrera egitea oztopatuz.

1.4. MiRNAen ikerketa etxe-abereetan

MiRNAk gizakietan oso aztertuak izan diren arren, gero eta ikerketa gehiago egiten ari dira miRNAek animalien biologian duten eraginaren inguruan (McDanel, 2009). MiRNA ugari identifikatu dira behian, oiloan, txerrian eta ardian, eta miRNA horietako askok adierazpen espazio-temporal oso erregulatu bat dute. Horrek miRNAk ehun eta garapen estadio espezifikoetan eginkizunak dituztela adierazten digu (Wang *et al.*, 2013). Horretaz gain, etxe-abereen miRNAen adierazpen profila, eta bereziki miRNA zirkulatzailen (C-miRNA) adierazpen profila, zenbait gaixotasunekin edo egoera fisiologikoenkin alda daiteke. C-miRNAk zeluletatik kanpo, gorputzeko fluidoetan besikuletan bilduta agertzen diren miRNAk dira. Zelula arteko komunikazioan parte hartzen dutela uste da eta C-miRNAen adierazpen aldaketak minbizi, aurdunaldi, inguru-estres eta bestelako gaixotasun eta egoera fisiologikoenkin harremandu dira (Miretti *et al.*, 2020).

Aurreko guztia kontuan hartuta, garrantzitsua da etxe-abereetan miRNAen adierazpen profila ikertzea eta horri buruz dugun ezagutza handitzea. Izan ere, etxe-abereen osasuna eta produktuen kalitatea kontrolatzeko bioindikatzailen oso erabilgarriak dira. MiRNAek bioindikatzailen konbentzionalek baino informazio gehiago ematen dute animalien egoera fisiologiko desberdinen eta estresen aurrean aktibatzen eta desaktibatzen diren mekanismo zelularrei buruz. Gainera, C-miRNAk animalien gorputzeko fluidoetan agertzen direnez (esnean, adibidez) teknika ez-inbaditzaileen bidez eta ohiko odol analisiekin batera azter daitezke (Miretti *et al.*, 2020).

Bioindikatzailen erabilgarriak aurkitzeaz gain, miRNAekin eginiko ikerketek zenbait ekarpen egin ditzakete etxe-abereen hazkuntza programetan, intereseko fenotipoak hautatzeko eta hobetzeko. Izan ere, etxe-abereen produktuen ekoizpen altuarekin edo etxe-aberearen osasun egoera hobearrekin harreman dutako nukleotido bakarrek polimorfismoak (SNP) aurkitu dira intereseko geneen miRNA lotze guneetan (Liu *et al.*, 2010).

Etxe-abereetan miRNAei buruz eginiko ikerketa gehienak garrantzi ekonomikoa duten ezaugarrien inguruan izan dira. Hainbat lanetan aztertu da miRNAen ikuspuntutik esne, okel eta arrautzen ekoizpena, bai eta etxe-abereen ugalkortasuna eta gaixotasunekiko erresistentzia (Fatima *et al.*, 2013). Hala ere, garrantzitsua da hainbat ehun era integratuan aztertzea nahiz eta garrantzi ekonomikorik ez izan. Modu horretan, miRNA ehun espezifikoak identifikatu daitezke eta ehunak miRNA adierazpen profilararen arabera taldekatu daitezke. Etxe-abereetan hainbat ehun era integratuan aztertzen dituzten ikerketak egin dira, zaldian adibidez (Pacholewska *et al.*, 2016). Ardian, ordea, ez da klase honetako ikerketarik egin.

Gure lan taldeak ardiaren miRNAei buruzko zenbait ikerketa egin ditu. Horietako bi lanetan garun-azaleko eta odoleko zelula mononukleatu periferikoetako (PBMC) miRNAk aztertu ziren, ardiak aluminio hidroxido txerto-laguntzailea soilik zuten txertoekin *in vivo* tratatu ostean. Bi lanetan ondorioztatu zen aluminio hidroxidoak, txertoaren garraioan parte hartzeaz gain, garun-azaleko eta PBMCetako transkriptoman aldaketak sor zitzakeela eta ondorio negatiboak sor zitzakeela (Varela-Martínez *et al.*, 2018; Varela-Martínez *et al.*, 2020). Beste lan batean, hausnarkari txikien lentivirus (SRLV) infekzioaren aurreko miRNA adierazpenaren erantzuna aztertu zen lehen aldiz ardian. Ikerketa birikietako zelulekin egin zen eta infekzio osteko birikietako lesioen eragilea oar-miR-21 miRNA izan zitekeela ondorioztatu zen (Bilbao-Arribas *et al.*, 2019).

2. HELBURUAK

Lan honetan ardiaren ehun ugaritan eginiko miRNA-seq datuak bilduko dira eta, ardiaren erreferentziako genoma eguneratuena erabiliz, tresna bioinformatikoekin landu eta estatistikoki aztertuko dira, era konparagarri eta uniformean. Horrez gain, lan honetan lortu nahi diren helburu espezifikoak ondokoak dira:

1. Ardiaren ehun bakoitzean miRNAen banaketa aztertzea
2. Ardiaren ehunetako miRNA adierazienak zeintzuk diren eta zein funtzio betetzen dituzten ikustea.
3. Ardiaren ehun bakoitzean miRNA ehun espezifikorik dagoen determinatzea, identifikatzea eta beste espezie batean ere identifikatu den ikustea.

3. MATERIALA ETA METODOAK

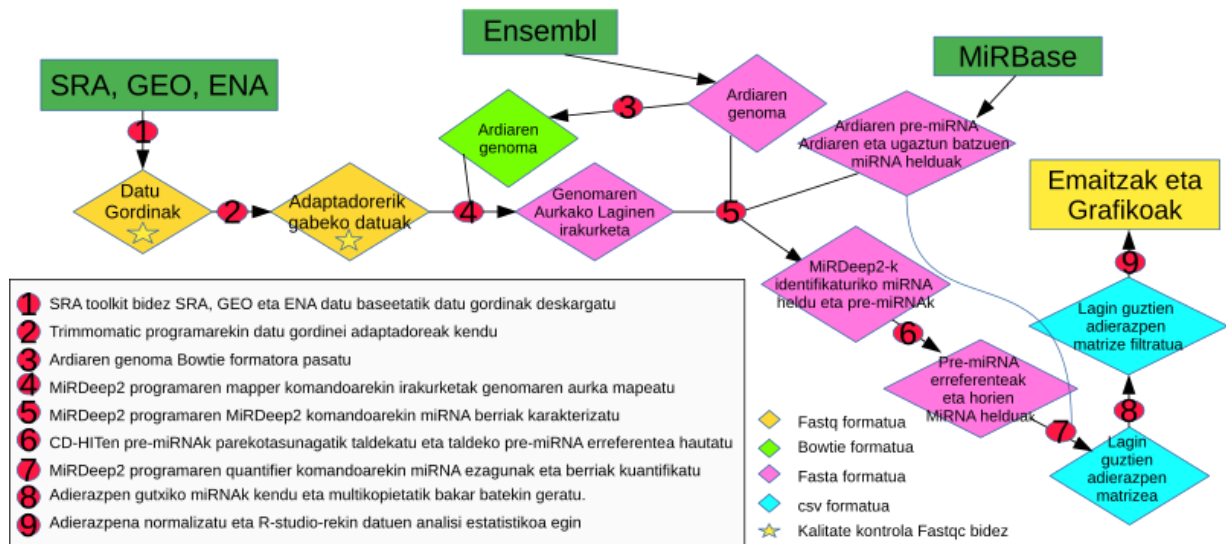
3.1. Laginen hautaketa

Ardiaren ahalik eta ehun gehienekin lan egiteko, PubMed-en “mirna” eta “sheep” hitzekin agertutako artikuluguztiak errebisatu ziren, denetara 210 artikuluguztiak. Soilik aukeratu ziren GEO, SRA edo ENA datu baseetako kodeak zituzten artikuluguztiak. Horrez gain, SRA datu baseko artxiboetan “sheep mirna-seq” hitzekin agertzen ziren kode guztiak errebisatu ziren eta lagin horiek igotako taldeak idatzitako artikuluguztiak bilatu ziren. Laginak igota egon arren arikulurik ez egotekotan (artikulua idazteke balego, adibidez) lagin multzoaren bioproject-eko kodea gorde zen.

Artikuluguztiak banan-banan errebisatu ziren eta azterketa *Ovis aries*-ekin egin zela eta ehun bateko laginak zirela egiaztatu zen (ez zelula espezifiko batenak, ezta ardiaren jariakinenak). Ikerketan erabilitako arraza agertzen bazen apuntatu egin zen. Bildutako laginek tratamendu berezirik jaso ez zutela egiaztatu zen eta, tratamendu berezirik jaso balute, kontrol laginak hautatu eta beste guztiak deskartatu ziren. Laginetan egon litezkeen erroreek emaitzetan eragin handirik ez izateko, bi lagin baino gutxiago zituzten ehunak deskartatu ziren. Espezifikoak ziren ehunetako laginak talde orokorragoetan bildu ziren. Laginak euren artean konparagarriak izateko metodo eta programa berdinekin landu ziren, 2. irudian ikus daitekeen lan fluxua jarraituz. Hori dela eta, lehendabizi, lagin guztiak deskargatu ziren SRA toolkit programaren 2.10.8 bertsioarekin (SRA Toolkit Development Team).

3.2. Laginen kalitate kontrola eta preprozesaketa

Lagin guztien egoera orokorra ezagutzeko Fastqc programaren 0.11.5 bertsioa erabiliz kalitate kontrola egin zen (Andrews, 2010). Fastqc-k lagin guztietan eginiko analisiak txosten bakar baten biltzeko Multiqc programaren 1.8 bertsioa erabili zen (Ewels *et al.*, 2016). Multiqc-ren txostenan zenbait analisi aztertu ziren: irakurketen luzeera 50 bp ingurukoa zela, laginen irakurketen kopuru handi bat kalitate onekoak zela, 20 bp ingurutik aurrera adaptadoreak detektatzen zirela eta ez zeudela gainadierazitako sekuentzia ezezagun asko. Fastqc programak, kalitate kontrola egiteaz gain, laginek dituzte adaptadoreak detektatzen ditu. Horien izenak eta sekuentziak fasta artxibo batean bildu ziren.



2. irudia: Lan honetan analisi bioinformatikoa egiteko lagin guztiekin jarraitutako lan-fluxua. Pausu bakoitzean erabili diren programak ageri dira, hau da: SRA toolkit, Fastqc, Trimmomatic, MirDeep2, CD-HIT eta R-studio.

Adaptadoreak kentzeko Trimmomatic programaren 0.39 bertsioa erabili zen (Bolger *et al.*, 2014). Trimmomatic programaren ILLUMINACLIP parametroan aurretik sorturiko adaptadoreen fasta artxiboa sartu zen. Adaptadorea detektatzeko, programak laginetako irakurketak eta adaptadorearen sekuentziak lerrotzen ditu eta parekotasuna aztertzen du. Batzuetan sekuentziak ez ziren guztiz berdinak, hortaz, programari bi base-parekatze oker onartzeko eskatu zitzaion.

Adaptadoreak kentzeaz gain, Trimmomatic-ek bestelako prozedurak egin zituen. Alde batetik, lehen eta azken base-parea aurretik ezarritako balio minimo batetik beherako kalitate bat bazuten, programak kendu zituen. Kalitate minimo hori 20 izatea eskatu zitzaion. Bestetik, programari eskatu zitzaion irakurketak eliminatzea horien luzeera 16 bp-tik beherakoa bazen. Azkenik, irakurketen bp-ak bosnaka aztertzea eskatu zitzaion eta bost bp-en kalitatearen batezbestekoa 20 baino txikiagoa bazen, irakurketa eliminatzea. Programak preprozesatutako laginak itzuli zituen.

Preprozesatutako laginekin beste kalitate kontrol bat egin zen, Fastqc programa erabiliz. Adaptorerik gabeko laginen eta lagin gordinen Multiqc txostenak konparatu ziren, adaptadoreak ondo kendu zirela, laginetako irakurketen luzeera egokia zela eta orokorrean irakurketek kalitate ona zutela egiaztatzen.

3.3. MiRNAen karakterizazioa eta kuantifikazioa

Lagin preprozesatuen miRNAk karakterizatzeko MiRDeep2 programaren 0.1.2 bertsioa erabili zen (Friedländer *et al.*, 2011). MiRDeep2-k irakurketak genomaren aurka mapeatu behar zituen. Hori dela eta, Ensembl-etik deskargatu zen ardiaren erreferentziako genoma eguneratuena, hau da, Oar_rambouillet_v1.0 (International Sheep Genome Consortium, 2017) eta bowtie formatora pasatu zen.

MiRDeep2 programaren Mapper.pl komandoak egin zuen mapeoa: hasteko, lagin batean guztiz identikoak diren sekuentziak kolapsatzen dira, erreduantziak ekiditeko. Ondoren, bowtie programa erabiliz, irakurketak genomaren aurka mapeatzen dira. Irakurketaren hazi sekuentzian (lehen 18 bp-etan) parekatze oker bat

onartzeko eskatu zitzaion programari. Hazi sekuentziaren ostean programak bi parekatze oker onartzen ditu. Bukatzeko, programak genomaren aurkako mapeoaren artxibo bat itzuli zuen.

Ondoren, lagin bakoitzeko miRNAk karakterizatu ziren. Horretarako, MirBasetik deskargatu ziren ardiaren miRNA helduak, miRNA prekursorak eta zenbait ugaztunen (*Homo sapiens*, *Bos taurus*, *Equus caballus*, *Sus scrofa* eta *Capra hircus*) miRNA helduak. Sekuentzia horiek guztiak, ardiaren genomarekin, preprozesaturiko laginekin eta genomaren aurkako irakurketen artxiboarekin batera, miRDeep2 programaren miRDeep2.pl komandoan sartu ziren. miRDeep2-k ardian ez baina beste ugaztunetan aurkitu diren hainbat miRNA eta guztiz berriak ziren zenbait miRNA identifikatu zituen. MiRDeep2-k puntuazio bat ematen die sekuentzia guztiei eta 4 puntu edo gehiago zituzten sekuentziak hautatu ziren, benetako miRNAk izateko $0,79 \pm 0,01$ eko probabilitatea baitzeukaten. Puntuazio hori miRDeep2k defektuz ematen diona ($0,5 \pm 0,01$) baino zorrotzagoa da.

MiRDeep2-k identifikaturiko miRNA asko euren artean oso antzekoak ziren, hau da, base-pare bakarra edo bi base-pare desberdin zituzten euren artean. Hori gertatzen da SNP baten alelo desberdinak direlako edo transkripzio osteko prozesaketa gertatzen delako. Hortaz, erredundantziak ekiditeko, CD-HIT programaren bidez miRNA berri guztiak taldekatu ziren (Huang *et al.*, 2010). Taldekapenak sortzeko pre-miRNAen sekuentziak erabili ziren eta 0,9ko identitate minimoa ezarri zen. Talde bakoitzeko erreferentziako pre-miRNA bat hautatu zen. MiRDeep2-k miRNA heldu bezala identifikatu zituen sekuentziak bere Pre-miRNAren lehen 20 bp-etan agertzen bazen, sekuentzia hori pre-miRNAren 5' muturrean zegoen. Kasu horietan sekuentziaren izenari "-5p" identifikadorea ezarri zitzaion eta harizpi bidaiariaren izenari "-3p" identifikadorea. Ordea, miRNA heldua azken 20 bp-etan agertzen bazen, identifikadoreak alderantziz ipintzen ziren.

MiRDeep2 programaren quantifier.pl komandoan sartu ziren MirBaseko pre-miRNA eta miRNA heldu ezagunak eta CD-HIT-etik ateratako erreferentziako pre-miRNA eta horietan kodeturiko miRNA helduak ("-5p" eta "-3p" identifikadoreak zeramatenak). Kuantifikazioa egitean MiRDeep2-k lagin guztien adierazpen matrize bat itzultzen du. Genomako hainbat lekuk miRNA heldu bera kodetu dezakete (multikopiak), beraz, programari multikopia bakoitzari miRNA heldu horren adierazpen totala esleitzeko eskatu zitzaion eta multikopietako bakar bat kontserbatu zen, besteak deskartatuz. Gainera, gutxienez ehun batean batezbesteko hamar irakurketa ez zituzten miRNAk deskartatu ziren. Lorturiko adierazpen matrize filtratuarekin laginen analisi estatistikoa, miRNAen adierazpenaren banaketaren azterketa eta miRNA ehun espezifikoan identifikazioa egin zen.

3.4. Laginen arteko antzekotasun analisia

Hasteko, datuak normalizatu ziren, laginak euren artean konparagarriak izateko eta sekuentziazioaren sakontasunaren eragina ezabatzeko. Adierazpen balioak milioikoetara normalizatu ziren (CPM). Horretarako, EdgeR R paketeko 3.26.8 bertsioako cpm funtzioa erabili zen. Funtzio horrek lagin bakoitzeko irakurketa kopuru totala kalkulatu du eta lagin horretako miRNA bakoitzaren adierazpena zatitzen du kopuru horrengatik. Ondoren, emaitza milioi batengatik biderkatzen du.

Ehun berdineko lagin guztiak elkarrekin taldekatzen ote ziren aztertu zen. Horretarako, normalizatutako adierazpen matrizeko balioak \log_{10} -gatik transformatu ziren eta matrize berri horrekin t-SNE teknika erabili zen. t-SNEa, osagai nagusien analisia (PCA) bezala, dimentsio ugari dituzten datuak bi dimentsiotan murrizteko teknika bat da. PCA metodo lineala da eta datu guztien forma globala mantentzen du. Ordea, t-SNE metodoak

taldekapan lokalak mantentzen ditu, ehun berdineko laginak elkarrekin taldekatzen diren ikustea erraztuz. t-SNEa egiteko Rtsne R paketeko 0.15 bertsioiko Rtsne funtzioa erabili zen.

Horren ostean, laginen arteko korrelazioa aztertu zen, ehun berdineko laginen arteko eta ehun desberdinen arteko korrelazioa ikusteko. Horretarako, binaka laginen arteko korrelazio balioak kalkulatu ziren pearsonen metodoa erabiliz. Korrelazio balioak kalkulatzeko laginen adierazpen matrize normalizatu eta transformatutik (\log_{10} -gatik) abiatu zen. Lorturiko korrelazio matrizearekin Heat Map bat osatu zen, Pheatmap R paketeko 1.0.12 bertsioiko pheatmap funtzioa erabiliz. Heat Map-a bi dimentsioetako datuak bistaratzeko teknika bat da. Korrelazio matrizean agertzen den balio txikienetik handienara kolore gradiente bat sortzen du eta balio bakoitzari kolore bat egokitzen dio. Ondoren, balioak koloreengatik ordezkatzeko dituzte. Gainera, Heat map-ari ehunak korrelazio bidez taldekatzeko eskatu zitzaion eta, pearsonen metodoa erabiliz, laginak dedrograma batean banatu zituen.

3.5. MiRNAen banaketaren eta ehun espezifikotasunaren azterketa

Ehun bakoitzeko miRNA adierazienak adierazpen totalaren zer portzentaje osatzen zuten aztertu zen. Horretarako, ehun bakoitzeko lagin guztien miRNAen batezbestekoa kalkulatu zen. Ondoren, ehun bakoitzeko miRNAen irakurketa totaletik miRNA bakoitzaren portzentajea kalkulatu zen ehunekotan. Portzentajerik altuenak zituzten lehen hamar miRNAk hautatu ziren. Lanen arteko desberdintasunak ikusteko, prozedura berdina egin zen bareako lan desberdinetako laginekin.

Ondoren, miRNAen ehun espezifikotasuna aztertu zen. Horretarako, lehenik eta behin, ehun bakoitzean miRNA bat adierazten zen edo ez ikusi zen. MiRNA bat ehun batean adierazita dagoela kontsideratzeko ehun horretako laginen erdian baino gehiagoetan miRNA baten adierazpena aurretik erabakitako balio minimoa baino handiagoa edo berdina izan behar zuen. Ehun espezifikotasuna aztertzeko erabakitako balio minimoa 5 CPM izan zen. Ehun batean miRNA bat adierazten zela kontsideratzen bazen 1 zenbakiarekin adierazten zen, bestela 0 zenbakiarekin.

Ehun guztietako miRNA presentzia/absentzia matrizea sortu zen (1 eta 0 balioak erabiliz) eta matrizeetik abiatuz Upset grafikak sortu ziren, UpSetR R paketeko 1.4.0 bertsioiko upset funtzioa erabiliz. Upset multzoak eta euren arteko intersekzioak era ulergarri bistaratzeko teknika bat da. Upsetek miRNA bakoitza adierazten den ehuna edo ehun multzoa determinatzen du, presentzia/absentzia matrizea erabiliz. Ondoren, ehunak edo ehun multzoak ordenatu egiten ditu, eskusiboki adierazten diren miRNA gehien dituztenetatik gutxien dituztenetara. Ehun edo ehun multzoak X ardatzean adierazten dira eta horietan eskusiboki adierazten diren miRNA kopurua Y ardatzean adierazten da. Modu honetan, ehun batean edo zenbaitetan bakarrik adierazten diren miRNA kopurua erraz ikus daiteke.

Azkenik, miRNAen ehun espezifikotasuna aztertzeko Yanai eta bere lan taldeak garatutako tau (τ) ehun espezifikotasun indizea (EEI) erabili zen. Tau neurri eskalar graduatu bat da, zeinak miRNA baten ehun bateko adierazpenaren espezifikotasuna beste ehunekiko adierazten duen. Tau balioak 0 eta 1 tartean ager daitezke, 1 balioa duten miRNAk ehun bakar batean adierazten dira (oso ehun espezifikokoak dira) eta 0 balioa dutenak, ehun guztietan portzentaje berean adierazten dira.

Tau modu honetan definitzen da:
$$\tau = \frac{\sum_{i=1}^n (1 - x_i)}{n - 1},$$

non n aztertutako ehun kopurua den eta x_i miRNA baten ehun desberdinetako adierazpen normalizatua den, miRNA horren adierazpen balio handienagatik zatitua. (Yanai *et al.*, 2004). Ehun bakoitzeko lagin guztien miRNAen adierazpen normalizatuaren eta \log_{10} -gatik transformatuaren batezbestekoa kalkulatu zen. MiRNA bakoitzaren EElak kalkulatu ziren eta 0,9-tik gorako balioa zuten miRNAen adierazpen normalizatua Heat Map batean adierazi zen, kolore gradiente batekin zein ehunetan gehiago adierazten diren ikusteko. Programari eskatu zitzaion ehunak eta miRNAk korrelazio dendrograma batekin taldekatzea.

4. EMAITZAK

4.1. Bildutako datuen deskribapena

PubMed-eko 210 artikulua errebisatu ostean, ardiaren ehunetan miRNA-seq egiten duten eta sekuentziatutako laginak SRA, ENA edo GEO datu baseetara igo dituzten 15 lan aurkitu ziren (1. taula). Hautatutako 15 lanetatik 5 ardiaren ugalek aparatuari buruzkoak ziren (endometrioia, obulutegia eta gorputz luteoa) eta beste lan batean (International Sheep Genome Consortium, 2017) aztertutako ehunen artean ugalek aparatuaren beste bi lagin agertzen ziren (obiduktua). Bareako lau lagin gehiago lortu ziren, gure lan taldeak aztertu zituenak eta datu baseetan igota ez zeudenak. Guztira 152 lagin lortu ziren.

1. taula: Lan honetarako hautatutako ikerketak, horiek aztertutako ehunak, ehun bakoitzeko lagin kopurua parentesien artean eta NCBI-n laginak aurkitzeko kodea biltzen dituen taula.

Lana	Ehunak eta lagin kopurua	Kodea
Yang <i>et al.</i> , 2019	Obulutegia (6), Endometrioia (2)	SRP142554
Zhou <i>et al.</i> , 2017	Ehun adipotsua (2)	SRP093866
Pokharel <i>et al.</i> , 2018	Obulutegia (10)	PRJEB22101
Pokharel <i>et al.</i> , 2020 (a)	Gorputz luteoa (10)	PRJEB32852
Pokharel <i>et al.</i> , 2020 (b)	Endometrioia (10)	PRJEB32852
Hou <i>et al.</i> , 2018	Hesteak (3)	GSE100649
Gao <i>et al.</i> , 2017	Ile sustraiak (3)	PRJNA308631
Gossner <i>et al.</i> , 2017	Nodulo linfatikoa (12)	PRJEB20781
Gu <i>et al.</i> , 2019	Obulutegia (9)	GSE122607
Bilbao-Arribas <i>et al.</i> , 2019	Birikiak (5)	GSE115415
Zhang <i>et al.</i> , 2020	Hipotalamoa (12)	PRJNA532808
Long <i>et al.</i> , 2019	Bihotza (6), Muskulua (6), Birikiak (6), Giltzurrunak (4), Gibela (5), Barea (6)	PRJNA511987
International Sheep Genome Consortium, 2017	Behazun xixkua (2), Bihotza (2), Azala (2), Muskulua (2), Nodulo linfatikoa (2), Kolona (2), Omasoa (2), Errumena (2), Obiduktua (2), Ureterra (2)	PRJNA414087
Varela-Martínez <i>et al.</i> , 2018	PBMC (6)	GSE113899
Varela-Martínez <i>et al.</i> , 2020	Garun-azala (5)	GSE128597
Varela-Martínez <i>et al.</i> , prestaketan	Barea (4)	prestaketan

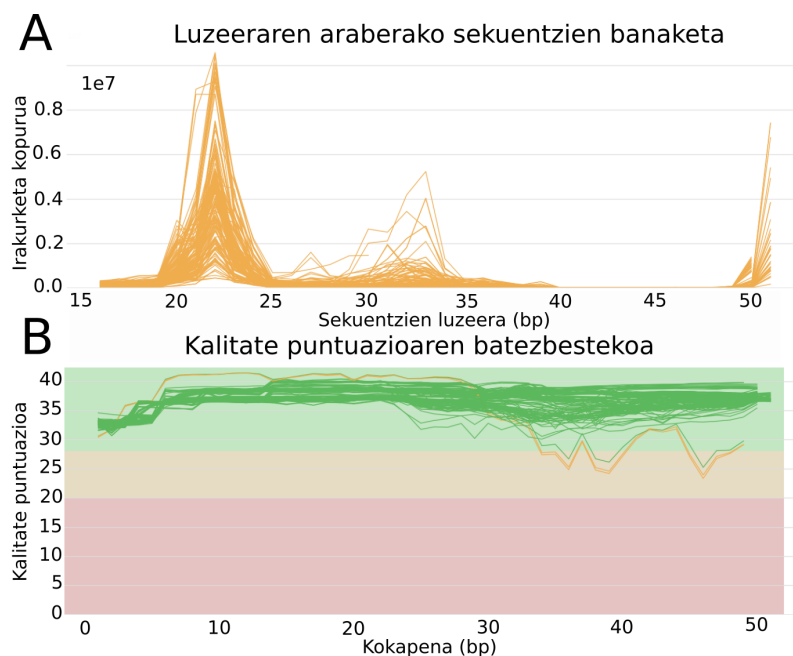
Ehun batzuk nahiko espezifikoak zirela eta, talde orokorragoetan bildu ziren: obulutegiko eta gorputz luteoko laginak “Obulutegia” taldean bildu ziren; garun-azaleko eta hipotalamoko laginak “Garuna” taldean; kolona eta hesteak “Hesteak” taldean eta omasoa eta errumena “Urdaila” taldean (2. taula). Guztira 19 ehun aztertu ziren eta obulutegia izan zen lagin gehien zituen ehuna.

2. taula: Lan hau egiteko azterturiko ehunak eta ehun bakoitzeko lagin kopurua biltzen dituen taula. Ehunak lagin kopuru handienetik txikienera ordenatuta daude.

Ehuna	Lagin kopurua	Ehuna	Lagin kopurua
Obulutegia	35	Hesteak	5
Garuna	17	Urdaila	4
Nodulo linfatikoa	14	Giltzurrunak	4
Endometrioa	12	Ile sustriak	3
Birikak	11	Behazun xixkua	2
Barea	10	Azala	2
Bihotza	8	Obiduktua	2
Muskulua	8	Ureterra	2
PBMC	6	Ehun adipotsua	2
Gibela	5		

4.2. Sekuentzien preprozesaketa eta karakterizazioa

Datuen preprozesaketa egin baino lehen, laginek batez beste $13,43 \pm 0,49$ milioi irakurketa zituzten eta ostein $12,51 \pm 0,5$ milioi. Falta ziren irakurketa guztiak, beraz, ez zituzten pasatu Trimmomatic programaren filtroak. Preprozesaketa osteko laginetan ez zen adaptadorerik detektatu, hortaz, preprozesaketa modu egokian burutu zen. Lagin preprozesatuaren luzeeraren arabera sekuentzien banaketa aztertzean (3. irudia, A atala) lagin gehienetako irakurketen portzentaje handi bat 20-25 bp-ko luzeera zutela eta puntu altuena 22 bp-tan zegoela ikus daiteke. MiRNAk 22 bp inguru zituztenez, sekuentzia multzo hori, ia bere osotasunean, miRNAez osatuta egotea espero daiteke.



3. irudia: Sekuentziei adaptadoreak kendu eta gero Fastqc programak eginiko 152 laginen kalitate kontrolaren txostenetik ateratako bi grafika. A atalean laginetako sekuentzien luzeeraren arabera banaketa ikusten da. B atalean sekuentzien kalitatearen batezbestekoa ikusten da base-pareko.

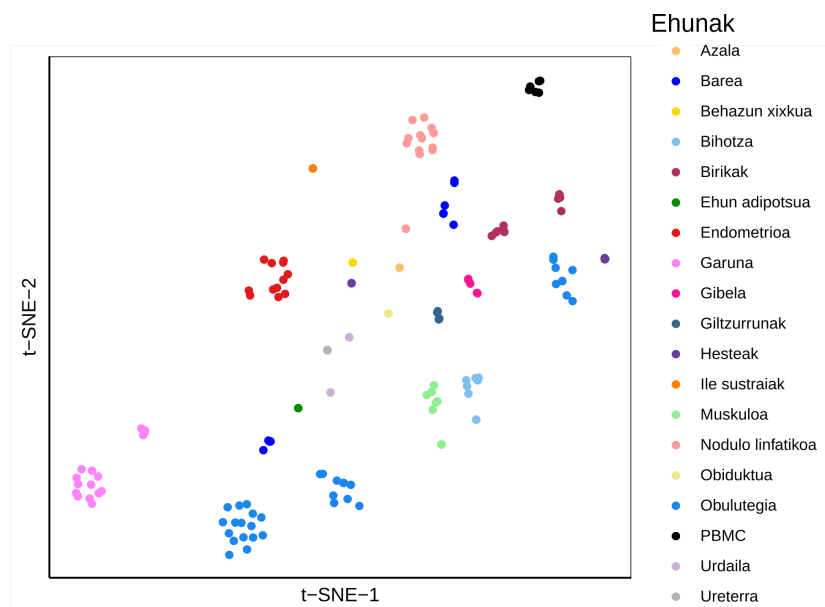
Hala ere, sekuentzien luzeeraren banaketan (3. irudia, A atala) bestelako luzeera zuten sekuentziak ikus daitezke. Alde batetik, 33 bp inguruan irakurketen portzentaje aipagarri bat ikus daiteke zenbait laginetan. Luzeera hori izanik sekuentzia multzo hori PIWI konplexuarekin elkarrekiten duten RNAz (piRNA) osatuta egotea espero daiteke. Azkenik, 50 bp edo handiagoak ziren sekuentzia multzoa RNA kate luzeagoen fragmentoak izatea espero da, izan ere, zenbait laginetan ikusi zen 50 bp-ko sekuentziek kalitate baxuagoa erakusten zuten, zati degradatuak zirela adieraziz (3. irudia, B atala).

MiRDeep2k laginetako sekuentzien batezbesteko % $51,93 \pm 1,42$ mapeatu zuten ardiaren genomaren aurka. 1336 sekuentzia hautatu ziren benetako miRNA izateko $0,79 \pm 0,01$ eko probabilitatea zutenak. Badira zenbait miRNA ezagun ez zirela iristen benetako miRNA kontsideratzeko ezarritako miRDeep2ren puntuazio minimora: oar-miR-134-5p, oar-miR-107, oar-miR-323c, oar-miR-221, oar-let-7f, oar-miR-487b-3p, oar-miR-544-3p, oar-miR-376a-5p, oar-miR-134-5p eta oar-miR-3956-5p.

CD-HIT-en miRDeep2ko puntuazio minimoa lortu zuten miRNA berrien pre-miRNAen sekuentziak sartu eta, 0,9ko identitatearekin, 1282 cluster sortu zituen. Multikopia bakarra hautatu eta gutxienez ehun batean 10 irakurketa ez zituzten miRNAk kendu ostean 1182 miRNA lortu ziren. Horietatik 1038 (% 87,8) miRNA berriak ziren ardian eta 144 (% 12,2) ardian ezagunak. MiRNA berrien artean 768 miRNA kontserbatuta agertu ziren gizakian, txerrian, ahuntzan, behian edo zaldian eta beste 270 ez ziren espezie horietan aurkitu.

4.3. Laginen arteko antzekotasun analisia

Lagin guztietako miRNAen adierazpen matrize filtratuarekin eginiko lehen analisia tSNEa izan zen. tSNE grafikoa (4. irudia) orokorrean ehunak ondo taldekatzen zirela ikus daiteke, baina baziren taldekatze gune desberdinetan multzokatu ziren ehun bereko laginak, hain zuzen ere, garuna, birrikak, obulutegia, barea, hesteak eta nodulu linfatikoa.

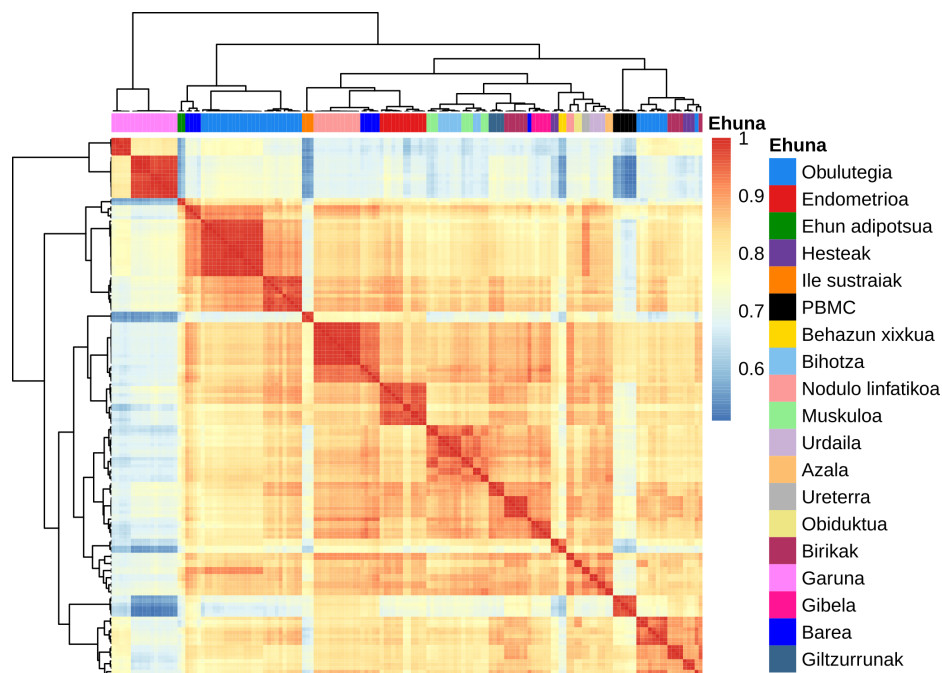


4. irudia: Lagin guztien adierazpen normalizatu eta \log_{10} -gatik transformatutik sorturiko tSNE grafika. Puntu bakoitza lagin bati dagokio eta laginak kokatzen dira beste laginekiko duten adierazpenaren antzekotasunengatik: adierazpen antzekoena duten laginek euren artean hurbilago kokatzen dira. Puntuen kolore bakoitza eskuinean zerrendaturiko ehun bakoitzari dagokio.

Ehun berekoak ziren eta banatuta agertzen ziren lagin-multzoak maiz bat zetozen ehun berdina aztertu zuten lan desberdinen lagin multzoekin. Barea, adibidez, bi lan desberdinetan aztertu zen: gure taldeak Rasa aragonesa-ren lau lagin aztertu zituen eta Keren Long eta bere taldeak Tibetan sheep-ren sei lagin aztertu zituzten. tSNEan argi ikusten da barea bi taldetan banatu zela: sei laginez osatutako bat eta lauko beste bat. Birikiekin berdina ikusten da eta obulutegiarekin ez da oso ondo antzematen. Beraz, lanen arteko desberdintasunek miRNAen adierazpenean eragin dezakete. Garuna ere bi taldetan banatzen zen, baina garuna aztertu zuten bi taldeek garunaren bi atal desberdin aztertu zituzten: gure taldeak garun-azala aztertu zuen eta Zhang *et al.*-ek hipotalamoia.

Jarraian, laginen arteko korrelazio matrizea kalkulatu eta Heat Map grafikoa sortu zen (5. irudia). Orokorrean ehun berdineko laginak taldekatuak agertu ziren, baina baziren oso banatuak agertzen ziren ehunak, tSNEan banatuta ikusitakoak, hain zuzen ere. Analisi honetan lagin bakoitzaren izena ikusteko aukera izan genuen eta, espero bezala, ehun berdinekoak diren eta multzo desberdinetan banatzen ziren laginak jatorrizko lanen arabera taldekatzen ziren. Bihotzeko eta muskuluko laginak tartekatuta agertu ziren, baina bihotza muskulu bat denez antzeko adierazpena espero daiteke.

Garunak nahiko adierazpen berezia zuen eta hasieratik banatu zen beste ehun guztietatik. Hala ere, tSNE grafikan bezala, garunean ikus daitezke elkarrengatik desberdin ziren bi lagin talde (garun-azala eta hipotalamoia). Irudian ikus daiteke, garunaz gain, badirela zenbait ehun besteekiko korrelazio oso baxua erakusten zutenak, hau da, desberdinagoak zirenak: ile sustraiak, behazun xixkua, PBMCak eta hestearen bi lagin.

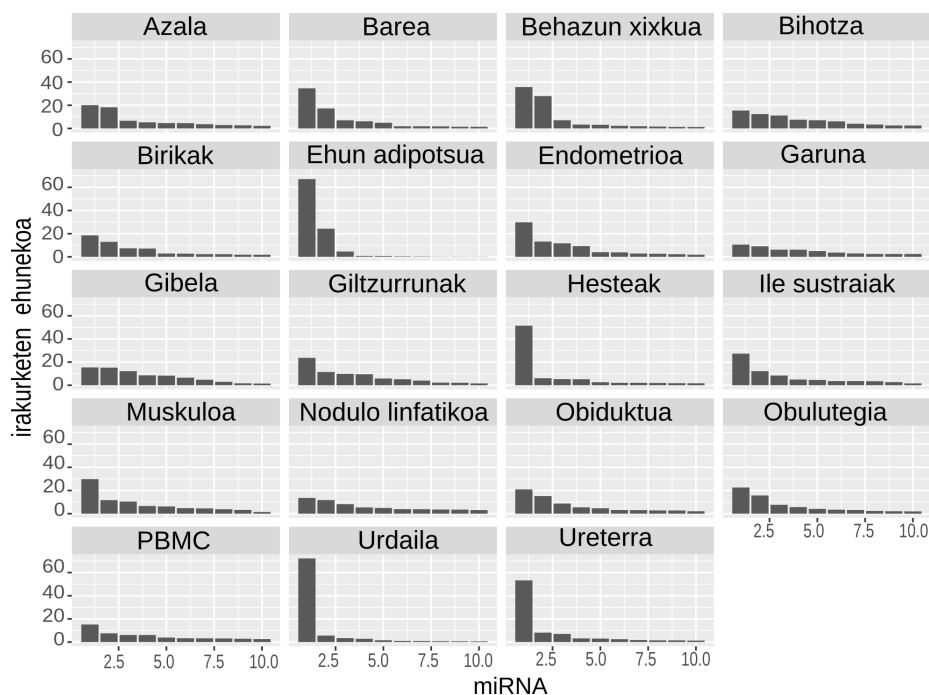


5. irudia: Adierazpenaren araberako laginen arteko korrelazio Heat Map-a, \log_{10} -gatik transformaturiko adierazpen normalizaturik abiatuta. Korrelazio balioak pearsonen metodoarekin kalkulatu dira eta korrelazio maximoa duten laginak kolore gorriarekin agertzen dira eta korrelazio txikiena dutenak urdinarekin.

Ehunen arteko korrelazioak ez dirudi korrelazio anatomikorik edo funtzionalik jarraitzen zuenik. Izan ere, ugal aparatuko zenbait lagin beste sistema edo aparatu batzuetako laginekiko korrelazio handiagoa erakusten zuten (PBMC, adibidez) ugal aparatuko beste laginekiko baino.

4.4. MiRNAen adierazpenaren banaketa eta ehun espezifikotasuna

MiRNAen banaketa aztertzeko, ehun bakoitzeko hamar miRNA adierazienak identifikatu eta adierazpen totalaren batezbesteko zein portzentaje betetzen zuten aztertu zen (6. irudia).



6. irudia: Ehun guztietako hamar miRNA adierazien batezbesteko portzentajea ehunekotan. Grafikoetako barra bakoitza miRNA bati dagokio eta ezkerretik eskuinera hamar miRNAk adierazpen altuenetik baxuenera kokatuak daude. Ardatz bertikalean miRNA bakoitzaren portzentajea adierazten da.

Grafikoan ikus daitekeen bezala, kasu batzuetan miRNA bakar batek hartzen zuen adierazpenaren portzentajerik altuena (hestean, urdailean eta ureterean, bereziki) eta miRNA gutxi batzuk antzeko mailan adierazten ziren. Ehun guztietan zortzigarren miRNAtik aurrera adierazpenaren portzentajea oso txikia bihurtzen zen. Orokorrean ehun bakoitzeko miRNA adieraziena beste ehunetako lehen hamar miRNA adierazien artean aurkitzen zen. Ehun ugaritan miRNA adieraziena oar-miR-143 izan zen (hestean, behazun xixkuan, nodulo linfatikoan, urdailean, obiduktuan, ureterean, biriketean eta barean).

Ehun berdina aztertu duten lan-talde desberdinen arteko heterogenitatea ikusteko, bareako laginak lan-taldetan banatuta aztertu ziren (7. irudia). Lehenik eta behin, grafikoan desberdintasun argi bat ikusten da lehen hamar miRNAen portzentajeetan. Izan ere, Keren Long *et al.*, -en laginetan miRNA adierazienetik hurrengoetara salto handiago bat zegoen. Bi lan-taldeen miRNA adierazienak bareako miRNA adierazpen totalaren % 40 baino gehiago hartzen zuen, baina miRNA hau zein den aztertzean bi lan taldeetan miRNA desberdinak agertu ziren. Keren Long *et al.*, -en laginetan oar-miR-143 agertu zen eta Varela-Martinez *et al.*, -en laginetan hsa-miR-486-5p. Ondorengo beste miRNA guztiek posizioa desberdinak hartu zituzten eta hsa-miR-486-5p ez zen Long *et al.*, -en laginetako lehen hamar miRNA adierazien artean aurkitu.

Ehun bateko laginen erdian gutxienez 5 cpm-ko adierazpena zuten 829 miRNA lortu ziren. Upset grafikoan ikus daitekeen bezala, miRNA multzo handiena ehun guztietan adierazten da eta 119 miRNA biltzen ditu. Hala eta guztiz ere, badira zenbait miRNA ehun bakar batean eskusiboki adierazten zirenak. Garunean 66 miRNA ehun espezifiko detektatu ziren analisi honetan. Hurrengo miRNA multzo handiena PBM Cetan agertu zen, 64 miRNA ehun espezifikoekin. Horien atzetik ile sustraiak eta behazun xixkua zeuden. Aipaturiko ehun horiek bat zetozen 5. irudian beste ehunekiko korrelazio baxuagoa azaltzen zuten ehunekin. Beraz, Garunean, PBM Cetan, ile sustraietan eta behazun xixkuan zenbait miRNA ehun espezifiko aurkitzen direla antzeman daiteke.

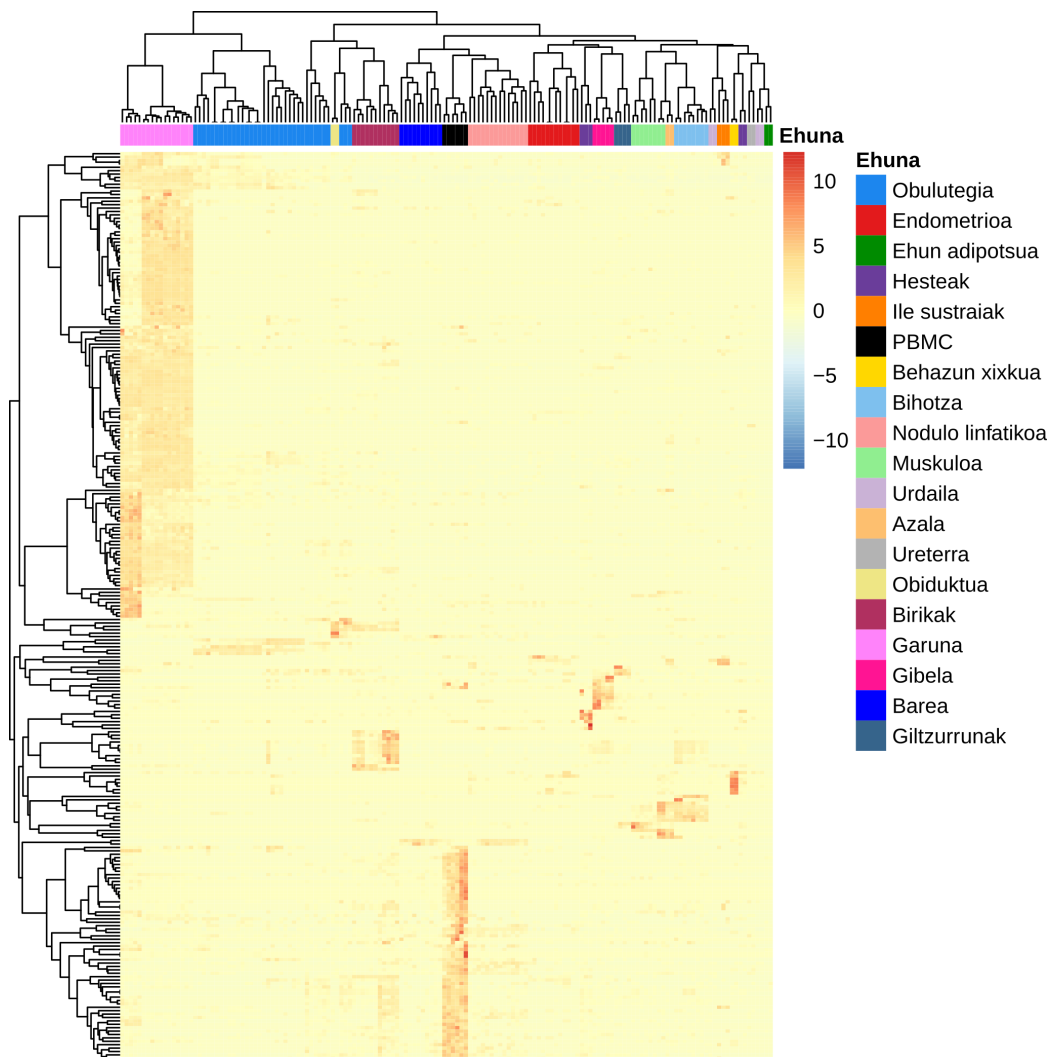
MiRNA ehun espezifiko horiek identifikatzeko tau izeneko EEI-a kalkulatu zen eta 0,9 baino balio altuagoa zuten miRNA en adierazpenarekin Heat Map bat sortu zen (9. irudia). 267 miRNA k izan zuten 0,9 edo balio altuagoa. Horietatik 24 ardiaren miRNA ezagunak ziren. MiRNA berrien artean 156 miRNA gizakian, txerrian, ahuntzan, behian edo zaldian kontserbatuta zeuden eta beste 87 ez ziren espezie horietan aurkitu. 12 miRNA k leko EEI balioa izan zuten, hau da, ehun bakar batean adierazten ziren. Horietatik bat ardiaren ezaguna zen, 7 aurreko espezieetan kontserbatuta zeuden eta 4 ez ziren aurreko espezieetan aurkitu. Heat Map-ean argi eta garbi ikusten da beste analisisetan ikusitako ehun bereziek (garunak, PBMek eta behazun xixkuak) miRNA ehun espezifiko asko edo gutxi batzuk zituztela.

Garuna da miRNA ehun espezifiko gehien zituen ehuna eta jarraian PBM Cak. Ile sustraiak eta behazun xixkuak miRNA ehun espezifiko gutxi batzuk ere zituzten. Horietaz gain, badira beste analisisetan detektatu ez ziren miRNA ehun espezifikoak zituzten ehunak, horien artean birrikak, bihotza, muskulua edo hesteak. tSNEan eta ehunen korrelazio Heat Map-ean ondo taldekatzen ez ziren ehunak hobeto taldekatu ziren miRNA ehun espezifikoekin eginiko Heat Map-ean, birrikak edo barea, adibidez. Gainera, immune sistemarekin harremanduriko ehunak (PBM Cak, barea eta nodulu linfatikoa) dendrogramaren adar batean taldekatu ziren, korrelazio funtzional bat adieraziz. MiRNA adierazien EEI, orokorrean, nahiko baxua izan zen. MiR-146a-ren EEI 0,57 izan zen eta miR-122rena 0,84. Beste miRNA guztien EEI 0,5 edo txikiagoa izan zen eta balio baxuena miR-143ek izan zuen (0,16).

5. EZTABAIDA

Lan honek izan duen zailtasun handiena miRNA berrien karakterizazioa izan da. Gaur egun erabilgarri dauden programek auresandako miRNA k eskuz zuzendu behar dira gerta daitezkeen bi arazo hauengatik: benetako miRNA berri kopurua gehiegi estimatzea edo gutxiestea (Fromm, 2016).

Alde batetik, miRNA kopuruaren gehiegi estimatzea gerta daiteke identifikaturiko positibo faltsuengandik (benetakoak ez diren miRNA k) edo programak eragindako erredundantziengatik (miRNA beretik datozen baina base-pare batean edo gutxitan desberdintzen diren sekuentziak, bi miRNA desberdin bezala identifikatuak). Positibo faltsuen identifikazioa minimizatzeke lan honetan miRDeep2 programa erabili zen, beste algoritmo batzuek baino positibo faltsu gutxiago identifikatzen zituela frogatu baitzen (Friedländer *et al.*, 2008). Gainera, lan honetan soilik hautatu ziren benetako miRNA izateke 0,79 ± 0,01eko probabilitatea zuten sekuentziak. Erredundantziak ekiditeke CD-HIT programa erabiliz miRNA berri identikoenak taldekatu behar izan ziren.



9. irudia: MiRNA ehun espezifikoeekin, hau da, EEI 0,9 edo handiago zuten miRNAen adierazpen normalizatu eta transformatuarekin sortutako Heat Map grafikoa. Ardatz horizontalean kokatzen dira eskuinean zerrendaturiko ehunak, dagokien koloreekin, eta korrelazio dendrograma batekin taldekatzen dira. Ardatz bertikalean miRNAk daude, korrelazio dendrograma batekin taldekatuak. MiRNA baten adierazpena altua bada kolore gorri adierazten da eta adierazpena baxua bada, kolore urdinez.

Bestetik, benetako miRNA kopurua gutxien bada, miRNAk kuantifikatzerakoan erroreak gerta daitezke. Hori gerta daiteke kuantifikazioa egiten dute programek base-parekatze oker bat edo gutxi batzuk onartzen dituztelako. Irakurketa bat identitate altuena erakusten duen miRNARI esleitzen zaio. Beraz, base-pare batean edo gutxitan desberdintzen diren bi miRNAetako bat ez bada identifikatzen, identifikatu ez den miRNAren irakurketak identifikatu denari esleitzea gerta daiteke (Barturen *et al.*, 2014).

Ikerketa-lan bakoitzean miRNA berri berdinak identifikatzen egon behar izateak asko oztzatzen du miRNAen ikerketa, egin beharreko prozedurak konplikatzeaz gain, lanen arteko konparagarritasuna ere zailtzen baitu. Izan ere, bi lan desberdinetan guztiz berria den miRNA bat identifikatuko balitz, autoreek izen desberdina emango liokete eta bi lanak konparatu nahi izatekotan, azterketa sekuentzia mailan egin beharko litzateke.

Zailtasun horri eman dakioken konponbiderik errazena miRNAen biltegi fidagarri eta eguneratu baten eraketa da, nomenklatura eta anotazio sistema estandarizatu bat erabiltzen duena. Aurretik aipatu den bezala, miRBase,

gaur eguneko miRNAen datu-base erabiliena eta estandarizatuena dena, www.mirbase.org, ez da 2018-tik eguneratu eta gizakia ez den beste izakietan miRNA asko falta dira anotatzeko. Hori gutxi balitz, miRBasen anotatuta dauden zenbait miRNA positibo faltsuak izan daitezke. MiRDeep2k miRNA berriak identifikatzeko prozedura berdina egiten du miRNA ezagunekin. Lan honetan identifikaturiko zazpi miRNAek benetako miRNA izateko $0,18 \pm 0,01$ eko edo gutxiagoko probabilitatea erakutsi zuten eta beste bi miRNA ez ziren iritsi lan honetan ezarritako probabilitate minimora. Ludwig eta taldekideek frogatu zuten miRBasen v21 bertsioan anotatutako miRNA berrienak beste RNA sekuentzia batzuen fragmentuak izan zitezkeela, endekatutako laginetan gehiago adierazteko joera erakusten baitzuten (Ludwig *et al.*, 2017).

Lanean topaturiko beste limitazio bat ehunkako laginen banaketa desorekatua izan da: ehun batzuek bi lagin soilik zituzten, beste batzuek ordea, hamarretik gora. Lagin askoko ehunekin lorturiko emaitzak estatistikoki sendoagoak eta fidagarriagoak dira, lagin batean gertaturiko akatsek emaitzetan eragin txikiagoa izango baitute. Literatura berrikustean, ardian miRNA-seq egin duten lan askok ugaltzat ikertzen zutelaz ohartu ginen. Lan honetan obulutegiko, endometriko, obiduktuko eta gorputz luteoetako laginek, gutxi gorabehera, lagin guztien herena osatzen zuten. Hori arrazoi komertzialengatik gertatzen da, miRNAek ugalketan izan ditzakeen eginkizunek etxe-abereen produktuen ekoizpenerako interes handia baitaukate (McDaneld, 2009).

Lan honetan aurkitutako hirugarren arazo bat izan da ehun berdina aztertu duten lan desberdinetako laginen artean desberdintasunak ikus daitezkeela. Garunean, aurretik aipatu den bezala, gune anatomiko desberdinetatik egin ziren erazketak (garun-azaletik eta hipotalamotik), baina beste ehunetan ez zegoen hain argi zerk eragiten zuten heterogenitatea. Heterogenitatea bi arrazoiengatik izan daiteke: ikerketarako erabilitako ardi arraza edo miRNAk aztertzeke erabilitako metodologia.

Arrazari dagokionez, ehun berdina aztertu zuten lan desberdinek ardi arraza desberdinak erabili zituzten kasu guztietan. Barea eta birikien kasuan, adibidez, gure taldeak Rasa aragonesa arraza erabili zuen eta Long eta taldekideek ardi Tibetarra eta Buztan laburreko Han ardia (Bilbao-Arribas *et al.*, 2019; Varela-Martínez *et al.*, prestaketan; Long *et al.*, 2019).

Metodologiari dagokionez, laginen preprozesaketak, miRNA erazteko teknikak eta sekuentziarioaren sakontasunak eragin garrantzitsu bat dute miRNAen adierazpenaren analisisian (Pritchard *et al.*, 2012). Orokorrean ikerketa guztietan RNA erazketa trizolaren protokolo estandarizatuarekin egin zen, sRNA liburutegiak Illuminaren TruSeq sRNA liburutegia prestatzeko kit-arekin sortu ziren eta HiSeq 2500 sekuentziagailuarekin sekuentziatu ziren (Bilbao-Arribas *et al.*, 2019; Long *et al.*, 2019). Hala ere, kontuan hartu ez diren beste zenbait aldagai egon daitezke. Beraz, ezin dugu zehatz jakin lanen arteko desberdintasunak zerk eragiten dituen.

Laneko emaitza adierazgarrienen artean ehun bakoitzeko miRNA adierazienek osatzen zuten adierazpenaren portzentajea da. Gehienez jota zortzi miRNAek hartzen zuten adierazpenaren portzentajerik handiena eta kasu batzuetan miRNA bakarrak hartzen zuten. MiRNA adieraziena mir-143 izan zen ehun ugartan: hestean, behazun xixkuan, nodulo linfatikoan, urdailean, obiduktuan, ureterean, biriketean eta barean, hain zuzen ere. Hestean, ureterean eta urdailan miRNA horrek adierazpenaren portzentajerik handiena bete zuen. Beste ehunetan miRNA adieraziena ez zen arren, lehen hamar miRNAen artean agertzen zela ikusi genuen eta bere ehun espezifikotasun indizearen balioa 0.16 izan zen. Yang *et al.*-ek eginiko azterketa batean iradoki zuten mir-143-3p zelulen ugartzearen eta apoptosiaren erregulazioan parte har zezakeela (Yang *et al.*, 2018).

Oro har, miR-122 eta miR-146a kenduta, ehun guztietako miRNA adierazienek (miR-143, miR-10b, miR-10a, miR-486-5p, miR-133, miR-21 eta miR-26a) EEI baxua zuten (0,5 edo txikiagoa) eta guztiek zuten harremana zelulen ugaritzearekin, apoptosiarekin eta minbiziaren garapenarekin (Peng & Croce, 2016). Horrek aditzera ematen du azterturiko ehun guztietako miRNA adierazienek zelularentzat oinarrizkoak diren funtzioak betetzen dituztela, zehazki, ziklo zelularren erregulazioan parte hartzen dutela.

Laneko emaitza adierazgarriena ehun espezifikokoak ziren miRNAen identifikazioa izan zen. MiRNA ehun espezifikokoak identifikatzeko hainbat ehunetako miRNAen adierazpena aztertu behar da. Ardian eginiko miRNAen ikerketa gehienetan, edo gutxienez lan honetarako aurkitu ziren lanetan, ehun bakarra aztertzen zen eta ehun bi baino gehiago aztertu zituzten ikerketa-lanetan ez zituzten miRNA espezikoak identifikatzeko analisirik egin (Long *et al.*, 2019). Ardia ez diren beste animalietan ehun ugarietako miRNAk aztertu zituzte eta horietako zenbait lanetan miRNA ehun espezifikokoak identifikatu zituzte: gizakian (Ludwig *et al.*, 2016), behian (Sun *et al.*, 2019), Panda handian (Wang *et al.*, 2020) eta zaldian (Pacholewska *et al.*, 2016) adibidez. Gizakiaren lanean 24 ehun desberdin aztertu ziren, ehun gehien aztertu dituen lana izanik. Lan honetan ardiaren 19 ehun aztertu diren eta behian, zaldian eta pandan, hurrenez hurren, 13, 9 eta 5 ehun.

Lan honetan identifikaturiko zenbait miRNA ehun espezifikoko beste lau lan horietan ere ehun berdinetan identifikatu ziren. Gizakian, adibidez, hurrengo miRNAk identifikatu ziren: miR-9, miR-7 eta miR-219a garunean; miR-133a muskuloan eta miR-215 hestean (Ludwig *et al.*, 2016). MiR-133 miRNAk muskuloan duen espezifikotasuna behin baino gehiagotan frogatu da, izan ere, mioblastoen desberdintzapenean eta ziklo zelularren erregulazioan parte hartzen duela ikusi da. Panda handiaren, zaldiaren, gizakiaren eta ardiaren gibelean miR-122-5p miRNA agertu zen miRNA ehun espezifikoko bezala. MiRNA hau glukosa eta lipidoen metabolismoan parte hartzen duela ikusi zen, beraz, gibeledako miRNA ehun espezifikoko bat dela suposa daiteke (Ludwig *et al.*, 2016; Sun *et al.*, 2019; Wang *et al.*, 2020; Pacholewska *et al.*, 2016). Lan honetan identifikaturiko ardiaren beste miRNA ehun espezifikoko bat miR-138 izan zen eta ardiaren garunean adierazten zela ikusi da. MiR-138 arratoiaren garunean ere adierazten da eta loaren erregulazioan parte hartzen duela egiaztatu da (Davis *et al.*, 2012).

Etorkizunera begira, lagin gutxi dituzten ehunak gehiago lagindu beharko liratezke, horiei buruz lorturiko emaitzak estatistikoki sendoagoak izan daitezkeen. Horrez gain, lanen arteko desberdintasunak zerk eragiten duen jakiteko, metodologia berdina erabiliz ardi arraza desberdinen miRNAen adierazpena konparatzea interesgarria izango litzateke. Gainera, lanen arteko konparagarritasuna handitzeko, RNA erauzketa, liburutegi prestaketa eta RNA-seq egiteko metodo estandarizatu bat finkatu beharko litzateke. Azkenik, berriz ere azpimarratu nahiko genuke miRNAk anotatzen dituen datu base eguneratu eta fidagarri baten beharra, aurrerantzean egingo diren lanak errazteko eta miRNAen ikerketan era sendoago batean aurrera egin ahal izateko.

6. ONDORIOAK

Lan honetan eginiko analisisietatik ondokoa ondoriozta daiteke:

1. Ardian aztertutako ehunetan miRNA gutxi batzuek adierazpenaren portzentajerik handiena hartzen dute. Are gehiago, ehun batzuetan, urdailean adibidez, miRNA bakar batek hartzen du adierazpenaren portzentajerik handiena.

2. Aztertutako ehunetako miRNA adierazienek, orokorrean, EEI baxua dute (0,5 edo txikiagoa) eta ziklo zelularren erregulazioan parte hartzen dute. Ehun ugartan miRNA adieraziena miR-143 izan da, ziklo zelularren erregulazioan parte hartzen duena eta EEI baxuena izan duena (0,16).
3. Ardian zenbait miRNA ehun espezifiko identifikatu dira eta horietako asko garunean eta PBM Cetan dira espezifikoak. Ardian identifikaturiko miRNA ehun espezifiko batzuk beste espezie batzuetan ere dira ehun espezifikoak.

7. ESKER ONAK

Eskerrik asko Begoña Jugori ikerketa lan hau eskeintzeagatik, lana berrikusteagatik eta egindako iradokizunengatik. Bestalde, eskerrak Endika Vareleri, berak aztertutako bareako laginen emaitzak uzteagatik, bere ikerketa publikatu aurretik. Azkenik, esker bereziak Martin Bilbaori, lanaren aspektu teknikoetan laguntzeagatik eta irakatsitako guztiagatik.

8. BIBLIOGRAFIA

Andrews, S (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 116(2), 281-297. doi:10.1016/S0092-8674(04)00045-5

Bartel, D. P. (2009). MicroRNAs: Target recognition and regulatory functions. *Cell*, 136(2), 215-233. doi:10.1016/j.cell.2009.01.002

Bartel, D. (2018). Metazoan MicroRNAs. *Cell*, 173(1), pp.20-51. doi:10.1016/j.cell.2018.03.006

Barturen, G., Rueda, A., Hamberg, M., Alganza, A., Lebron, R., Kotsyfakis, ... Hackenberg, M. (2014). sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments. *Methods in Next Generation Sequencing*, 1(1), 21-31. doi:10.2478/mngs-2014-0001

Bilbao-Arribas, M., Abendaño, N., Varela-Martínez, E., Reina, R., De Andrés, D., & Jugo, B. M. (2019). Expression analysis of lung miRNAs responding to ovine VM virus infection By ma-seq. *BMC Genomics*, 20(1). doi:10.1186/s12864-018-5416-0

Bolger, A., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bortolomeazzi, M., Gaffo, E. & Bortoluzzi, S. (2019). A survey of software tools for microRNA discovery and characterization using RNA-seq. *Briefings in Bioinformatics*, 20(3), 918-930. doi:10.1093/bib/bbx148

Bushati, N., & Cohen, S. M. (2007). MicroRNA functions. *Annual Review of Cell and Developmental Biology*, 23(1), 175-205. doi:10.1146/annurev.cellbio.23.090506.123406.

Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. V., Spies, N., Baek, D., ... Bartel, D. P. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Development*, 24, 992-1009. doi:10.1101/gad.1884710

- Davis, C. J., Clinton, J. M., & Krueger, J. M. (2012). MicroRNA 138, let-7b, and 125a inhibitors differentially alter sleep and EEG delta-wave activity in rats. *Journal of Applied Physiology*, 113(11), 1756–1762. doi:10.1152/jappphysiol.00940.2012
- Ewels, P., Magnusson, M., Lundin, S. & Källér, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048. doi:10.1093/bioinformatics/btw354
- Fatima, A., & Morris, D. G. (2013). Micrnas in domestic livestock. *Physiological Genomics*, 45(16), 685-696. doi:10.1152/physiolgenomics.00009.2013
- Friedländer, M., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S. & Rajewsky, N. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology*, 26(4), 407-415. doi:10.1038/nbt1394.
- Friedländer, M., Mackowiak, S., Li, N., Chen, W. & Rajewsky, N. (2011). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, 40(1), 37-52. doi:10.1093/nar/gkr688
- Friedman, R. C., Farh, K. K., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92-105. doi:10.1101/gr.082701.108
- Fromm, B. (2016). MicroRNA discovery and expression analysis in animals. *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing*, 121-142. doi:10.1007/978-3-319-31350-4_6
- Gao, L. & Jiang, F. (2016). MicroRNA (mirna) profiling. *Methods in Molecular Biology*, 151-161. doi:10.1007/978-1-4939-3204-7_8
- Gao, W., Sun, W., Yin, J., Lv, X., Bao, J., Yu, J., ... Hu, L. (2017). Screening candidate micrnas (mirnas) in different lambskin hair follicles in hu sheep. *PLOS ONE*, 12(5). doi:10.1371/journal.pone.0176532
- García-López, J., Brieño-Enríquez, M. A., & Del Mazo, J. (2013). MicroRNA biogenesis and variability. *Biomolecular Concepts*, 4(4), 367-380. doi:10.1515/bmc-2013-001
- Gossner, A., Watkins, C., Chianini, F., & Hopkins, J. (2017). Pathways and genes associated with immune dysfunction in sheep paratuberculosis. *Scientific Reports*, 7(1). doi:10.1038/srep46695
- Gu, B., Liu, H., Han, Y., Chen, Y., & Jiang, H. (2019). Integrated analysis of miRNA and mRNA expression profiles in 2-, 6-, And 12-month-old small Tail Han Sheep ovaries reveals that Oar-mir-432 downregulates RPS6KA1 Expression. *Gene*, 710, 76-90. doi:10.1016/j.gene.2019.02.095
- Hammond, S. (2015). An overview of microRNAs. *Advanced Drug Delivery Reviews*, 87, 3-14. doi:10.1016/j.addr.2015.05.001
- Hou, L., Ji, Z., Wang, G., Wang, J., Chao, T., & Wang, J. (2018). Identification and characterization of micrnas in the intestinal tissues of sheep (ovis aries). *PLOS ONE*, 13(2). doi:10.1371/journal.pone.0193371
- Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680-682. doi:10.1093/bioinformatics/btq003
- International Sheep Genome Consortium (2017). Oar_rambouillet_v1.0. <https://www.sheephapmap.org/>

- Kozomara, A. & Griffiths-Jones, S. (2013). Mirbase: Annotating high confidence micrnas using deep sequencing data. *Nucleic Acids Research*, 42(D1). doi:10.1093/nar/gkt1181
- Lee, R. C., Feinbaum, R. L. & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75 843–854. doi: 10.1016/0092-8674(93)90529-y
- Liu, H.-C., Hicks, J. A., Trakooljul, N., & Zhao, S.-H. (2010). Current knowledge of microRNA characterization in agricultural animals. *Animal Genetics*, 41(3), 225–231. doi:10.1111/j.1365-2052.2009.01995.x
- Long, K., Feng, S., Ma, J., Zhang, J., Jin, L., Tang, Q., ... Li, M. (2019). Small non-coding RNA transcriptome of four high-altitude vertebrates and their low-altitude relatives. *Scientific Data*, 6(1). doi:10.1038/s41597-019-0204-5
- Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., ... Keller, A. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Research*, 44(8), 3865-3877. doi:10.1093/nar/gkw116
- Ludwig, N., Becker, M., Schumann, T., Speer, T., Fehlmann, T., Keller, A. & Meese, E. (2017). Bias in recent miRBase annotations potentially associated with RNA quality issues. *Scientific Reports*, 7(1). doi:10.1038/s41598-017-05070-0
- McCormick, K. P., Willmann, M. R. & Meyers, B. C. (2011). Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence*, 2, 2-2. doi:10.1186/1758-907X-2-2
- McDaneld, T. G. (2008). Microrna: Mechanism of gene regulation and application to livestock. *Journal of Animal Science*, 87(Suppl_14). doi:10.2527/jas.2008-1303
- Metzker, M. L. (2009). Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11(1), 31-46. doi:10.1038/nrg2626
- Miretti, S., Lecchi, C., Ceciliani, F. & Baratta, M. (2020). MicroRNAs as biomarkers for animal health and welfare in livestock. *Frontiers in Veterinary Science*, 7. doi:10.3389/fvets.2020.578193
- Ozsolak, F., Poling, L. L., Wang, Z., Liu, H., Liu, X. S., Roeder, R. G., ... Fisher, D. E. (2008). Chromatin structure analyses identify miRNA promoters. *Genes & Development*, 22, 3172–3183. doi:10.1101/gad.1706508
- Pacholewska, A., Mach, N., Mata, X., Vaiman, A., Schibler, L., Barrey, E., & Gerber, V. (2016). Novel equine tissue miRNAs and breed-related Mirna expressed in serum. *BMC Genomics*, 17(1). doi:10.1186/s12864-016-3168-2
- Peng, Y. and Croce, C., 2016. The role of microRNAs in human cancer. *Signal transduction and targeted therapy*, 1(1). doi:10.1038/sigtrans.2015.4
- Pillai, R. S. (2005). Microrna function: Multiple mechanisms for a tiny rna? *RNA*, 11(12), 1753-1761. doi:10.1261/rna.2248605

- Pokharel, K., Peippo, J., Honkatukia, M., Seppälä, A., Rautiainen, J., Ghanem, N., ... Kantanen, J. (2018). Integrated ovarian mRNA and Mirna transcriptome profiling characterizes the genetic basis of prolificacy traits in sheep (*Ovis aries*). *BMC Genomics*, 19(1). doi:10.1186/s12864-017-4400-4
- Pokharel, K., Peippo, J., Weldenegodguad, M., Honkatukia, M., Li, M., & Kantanen, J. (2020). Gene expression profiling of corpus luteum reveals important insights about early pregnancy in domestic sheep. *Genes*, 11(4), 415. doi:10.3390/genes11040415
- Pokharel, K., Peippo, J., Li, M., & Kantanen, J. (2020). Identification and characterization of mirnas during early pregnancy in domestic sheep. *Animal Genetics*, 51(5), 833-836. doi:10.1111/age.12992
- Pritchard, C. C., Cheng, H. H., & Tewari, M. (2012). MicroRNA profiling: Approaches and considerations. *Nature Reviews Genetics*, 13(5), 358–369. doi: 10.1038/nrg3198
- Rogers, K., & Chen, X. (2013). Biogenesis, turnover, and mode of action of plant micromas. *The Plant Cell*, 25(7), 2383-2399. doi:10.1105/tpc.113.113159
- Saliminejad, K., Khorram Khorshid, H., Soleymani Fard, S. & Ghaffari, S., (2018). An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *Journal of Cellular Physiology*, 234(5), 5451-5465. doi:10.1002/jcp.27486
- SRA Toolkit Development Team. SRA-Tools. <http://ncbi.github.io/sra-tools/>
- Varela-Martínez, E., Abendaño, N., Asín, J., Sistiaga-Poveda, M., Pérez, M. M., Reina, R., ... Jugo, B. M. (2018). Molecular signature of aluminum hydroxide adjuvant in ovine PBMCs by integrated mrna and micromna transcriptome sequencing. *Frontiers in Immunology*, 9. doi:10.3389/fimmu.2018.02406
- Varela-Martínez, E., Bilbao-Arribas, M., Abendaño, N., Asín, J., Pérez, M., De Andrés, ... Jugo, B. M. (2020). Whole transcriptome approach to evaluate the effect of aluminium hydroxide in ovine encephalon. *Scientific Reports*, 10(1). doi:10.1038/s41598-020-71905-y
- Wang, X., Gu, Z., & Jiang, H. (2013). Micromas in farm animals. *Animal*, 7(10), 1567-1575. doi:10.1017/S1751731113001183
- Wang, C., Li, F., Deng, L., Li, M., Wei, M., Zeng, B., ... Zhang, H., 2021. Identification and characterization of miRNA expression profiles across five tissues in giant panda. *Gene*, 769, doi: 10.1016/j.gene.2020.145206
- Westholm, J. O., & Lai, E. C. (2011). Mirtrons: Micromna biogenesis via splicing. *Biochimie*, 93(11), 1897-1904. doi:10.1016/j.biochi.2011.06.017
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., ... Shmueli, O. (2004). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21(5), 650-659. doi:10.1093/bioinformatics/bti042
- Yang, J., Li, X., Cao, Y., Pokharel, K., Hu, X., Chen, Z., ... Li, M. (2018). Comparative mRNA and miRNA expression in European mouflon (*Ovis Musimon*) and sheep (*Ovis aries*) Provides novel insights into the genetic mechanisms for female reproductive success. *Heredity*, 122(2), 172-186. doi:10.1038/s41437-018-0090-1

- Yang, Z., Wang, J., Pan, Z. & Zhang, Y. (2018). miR-143-3p regulates cell proliferation and apoptosis by targeting IGF1R and IGFBP5 and regulating the Ras/p38 MAPK signaling pathway in rheumatoid arthritis. *Experimental and Therapeutic Medicine*, 5(4), 3781-3790. doi: 10.3892/etm.2018.5907
- Zhang, Z., Tang, J., Di, R., Liu, Q., Wang, X., Gan, S., ... Hu, W. (2020). Integrated hypothalamic transcriptome profiling reveals the reproductive roles of mRNAs And miRNAs in sheep. *Frontiers in Genetics*, 10. doi:10.3389/fgene.2019.01296
- Zhou, G., Wang, X., Yuan, C., Kang, D., Xu, X., Zhou, J., ... Chen, Y. (2017). Integrating miRNA and mRNA expression Profiling Uncovers Mirnas Underlying fat deposition in sheep. *BioMed Research International*, 2017, 1-11. doi:10.1155/2017/1857580