SNP discovery in the sheep milk casein genes using transcriptome sequencing

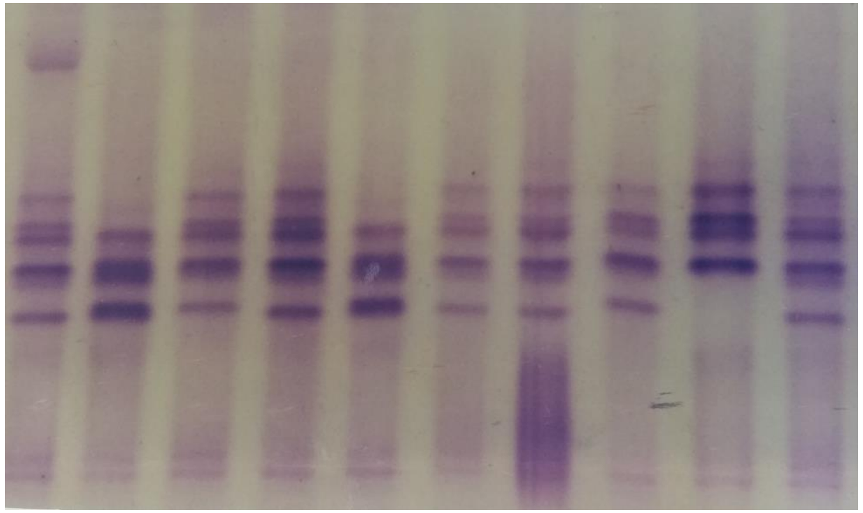
A. Suárez-Vega, B. Gutiérrez-Gil, C. Klopp, G. Tosser-Klopp and J.J. Arranz

Milk and Caseins

- Sheep milk is generally processed into dairy products, mostly cheese
- The knowledge of the expression levels of milk genes influencing cheese related traits becomes essential
- Cheese contains mainly the caseins and fat from milk
- In cattle and goat, caseins polymorphisms have been associated to quantitative and qualitative dairy parameters
- In sheep, the association studies between caseins polymorphisms and dairy traits are controversial

Sheep Caseins

- αs1-casein, αs2-casein, β-casein, κ-casein
- Encoded by *CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*
- Constitute about 76-83% of total milk protein
- 250 kb cluster located on the ovine chromosome 6 (OAR6: 85,087,000-85,318,000)
- Variability within caseins has been traditionally studied by:
  - Isoelectric focusing
  - PCR and Single strand conformation polymorphism



RNA-seq technology

- Next-generation sequencing (NGS): Changing the biological research view by enabling the comprehensive and relatively inexpensive analysis of genomes and transcriptomes

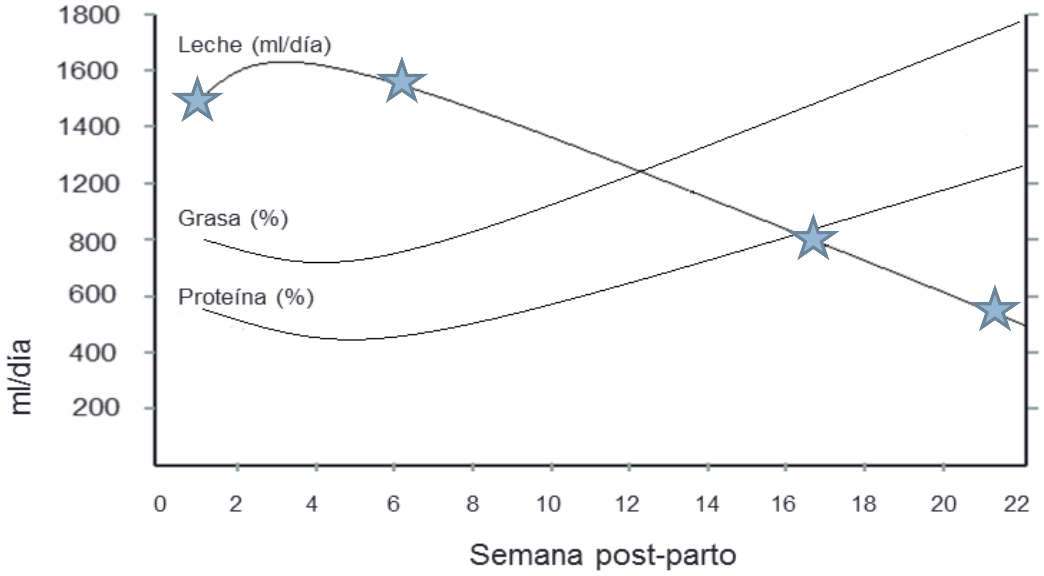
- RNA-seq
  - Quantify transcripts
  - Identify differential regulation
  - Detect transcriptomic variants

Lactating sheep mammary gland transcriptome study

- Global project of the lactating sheep mammary gland transcriptome

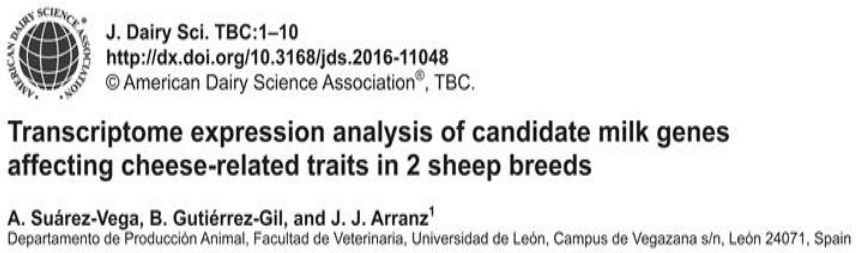
Milk yield (kg): 400
Milk fat %: 6.7
Milk protein %: 5.4

Milk yield (kg): 117
Milk fat %: 7.0
Milk protein %: 5.8



Lactating sheep mammary gland transcriptome study

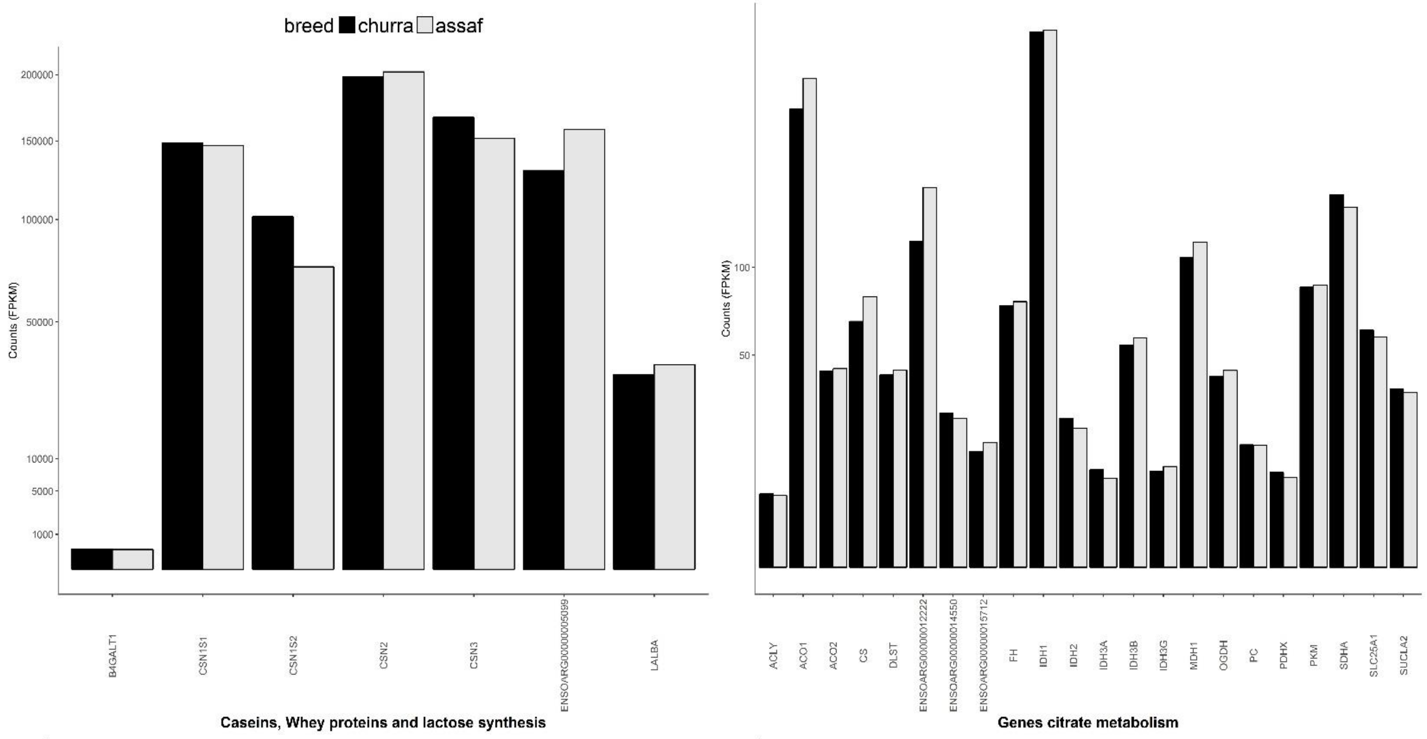
- **Quantify the abundance of transcripts related to cheese yield**
- Identify differential regulation
- Detect transcriptomic variants



J. Dairy Sci. TBC:1–10
http://dx.doi.org/10.3168/jds.2016-11048
© American Dairy Science Association®, TBC.

Transcriptome expression analysis of candidate milk genes affecting cheese-related traits in 2 sheep breeds

A. Suárez-Vega, B. Gutiérrez-Gil, and J. J. Arranz[1]
Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain



Lactating sheep mammary gland transcriptome study

- Quantify the abundance
- **Identify differential regulation between breeds and across lactation**
- Detect transcriptomic variants

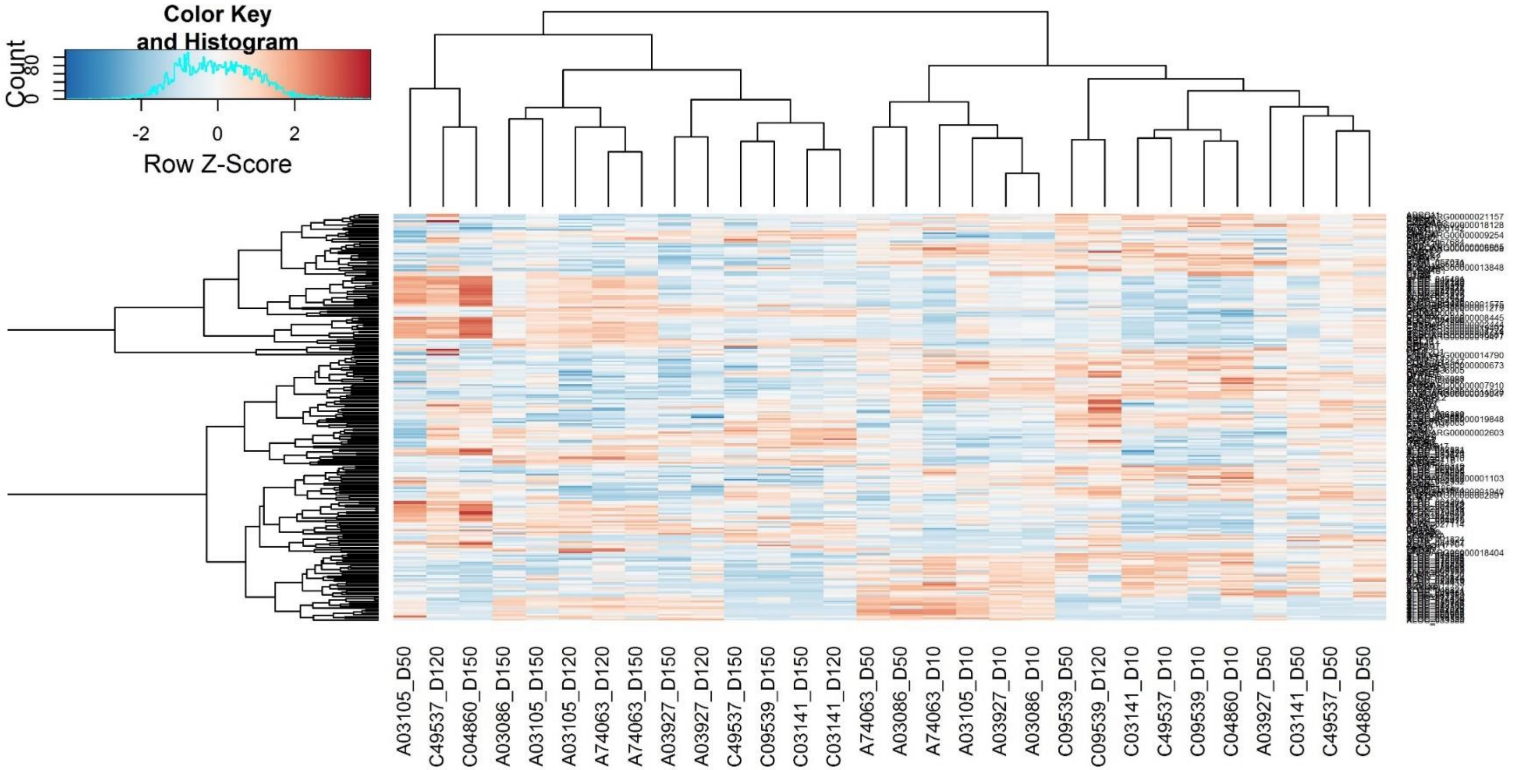
SCIENTIFIC REPORTS

OPEN

Characterization and Comparative Analysis of the Milk Transcriptome in Two Dairy Sheep Breeds using RNA Sequencing

Received: 20 July 2015
Accepted: 17 November 2015
Published: 18 December 2015

Aroa Suárez-Vega[1], Beatriz Gutiérrez-Gil[1], Christophe Klopp[5], Christèle Robert-Granie[2,3,4], Gwenola Tosser-Klopp[2,3,4] & Juan José Arranz[1]



Lactating sheep mammary gland transcriptome study

- Quantify the abundance of transcripts related to cheese yield
- Identify differential regulation
- **Detect transcriptomic variants**
  - Candidate genes → **Caseins**
  - QTL regions

Abundance

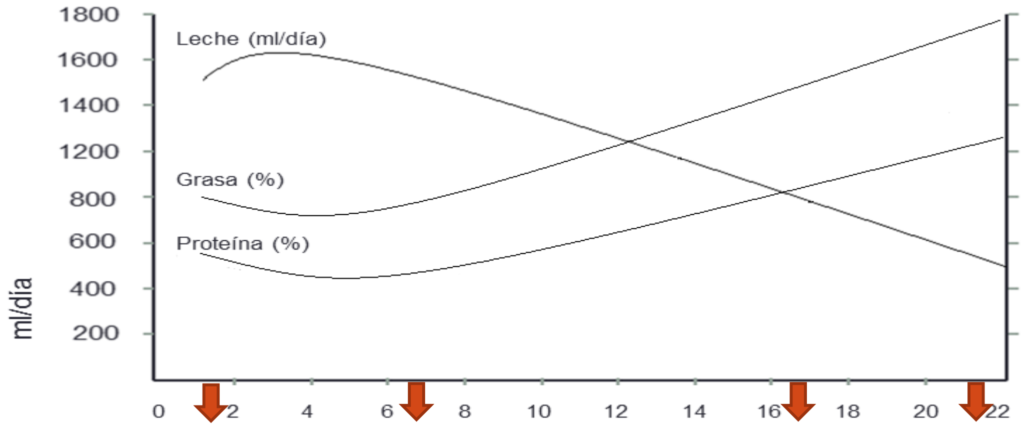
Technological properties

Cheese making

Objective

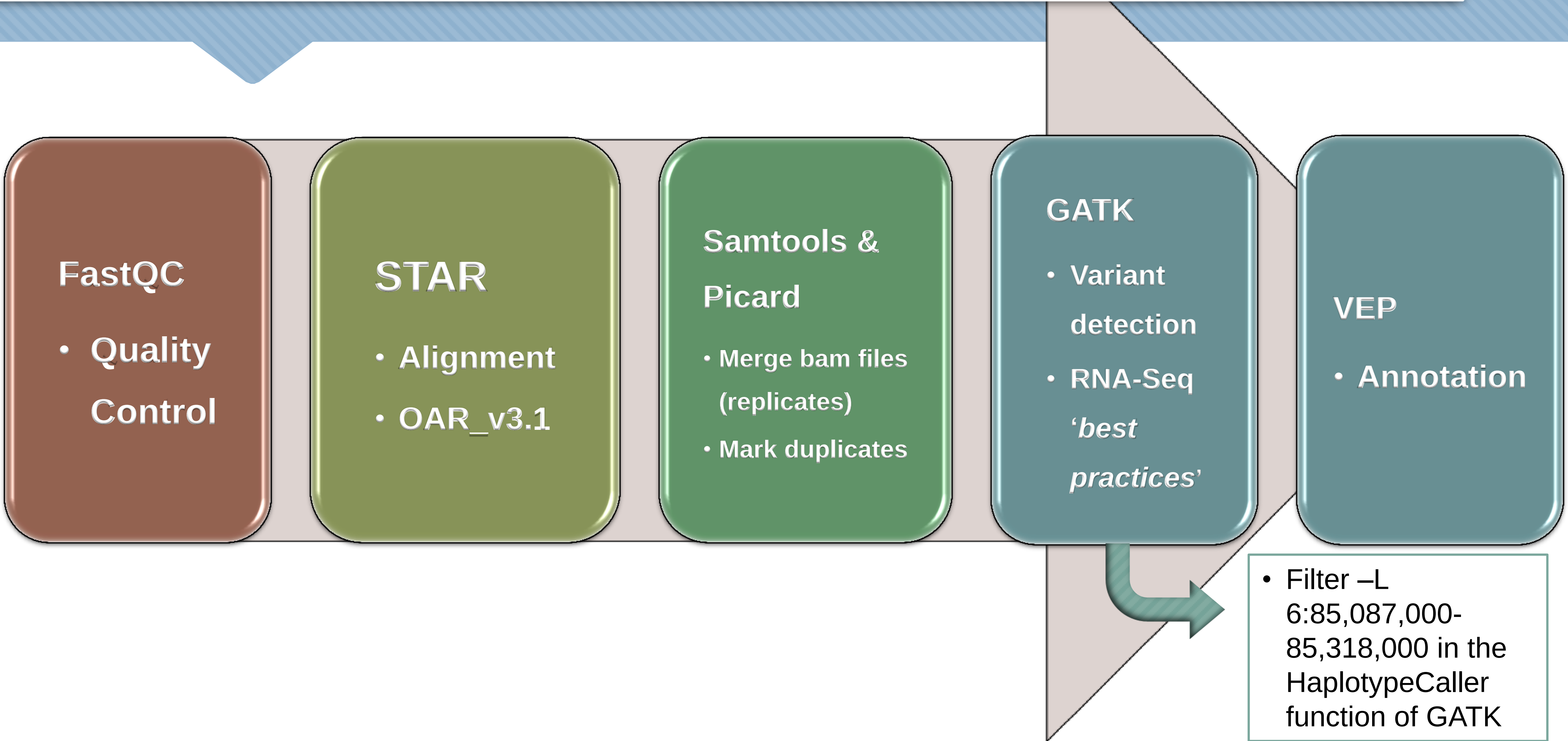
The use of the transcriptome sequencing technology on milk somatic cells with the aim of characterizing the variability within the genomic region encoding the sheep caseins in two common dairy sheep breeds

Experimental design, RNA extraction and Sequencing



Day 10	Day 50	Day 120	Day 150
 4 samples	 4 samples	 3 samples	 4 samples
 4 samples	 4 samples	 3 samples	 4 samples

Variant Detection Workflow



FastQC
- Quality Control

STAR
- Alignment
- OAR_v3.1

Samtools & Picard
- Merge bam files (replicates)
- Mark duplicates

GATK
- Variant detection
- RNA-Seq *‘best practices’*

VEP
- Annotation

- Filter –L 6:85,087,000-85,318,000 in the HaplotypeCaller function of GATK

Alignment results

	Sheep genome (Oar_v3.1)	Casein cluster (OAR6: 85,087,000-85,318,000)
Million Mapped Reads	119.33	22.27
Tags/Kb in Exons	2008.93	3.48 million
Tags/Kb in Introns	16.86	7011.22

High transcription level

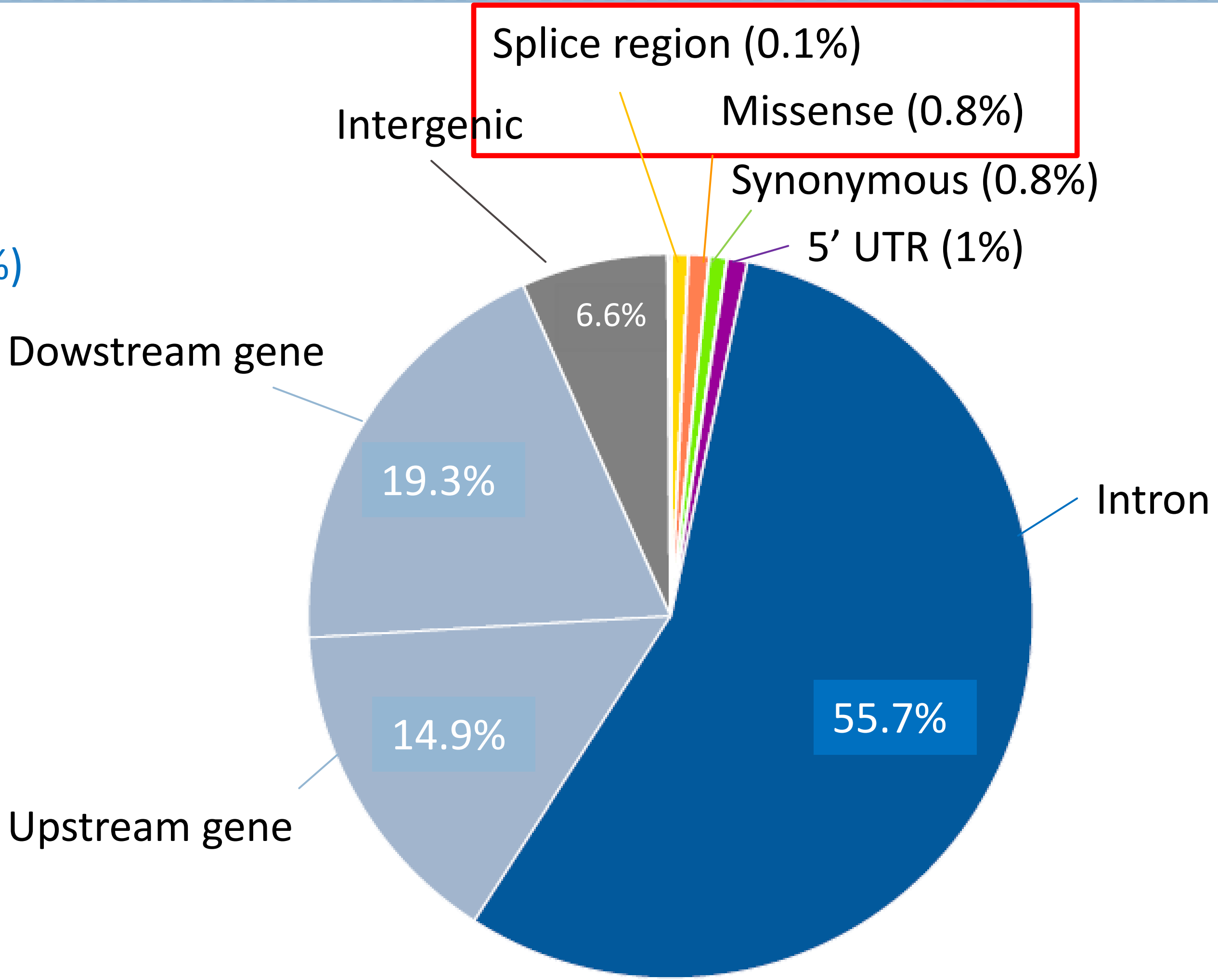
Inmature RNA

Variant detection

735 variants

Novel / existing variants: 65 (8.8%) / 670 (91.2%)

Variant class	Count
SNV	676
deletion	31
insertion	22
sequence_alteration	4
substitution	2



Variant	Position OAR6	Gene	Allele Freq		Effect	aminoacid	rs ID
	(Oar_v3.1)		Assaf	Churra			
c.626C>T	(6: 85102965)	*CSN1S1*	0.38	0.75	Missense-Tolerated	*p.Thr209Ile*	*rs420959261*
c.634C>A	(6: 85117294)	*CSN2*	0.625	0.25	Missense-Tolerated	*p.Leu212Ile*	*rs416941267*
c.595A>G	(6: 85117333)	*CSN2*	0	0.125	Missense-Tolerated	*p.Met199Val*	*rs430298704*
c.102+1G>A	6: 85186875	*CSN1S2*	0.625	0.625	Splice donor	⇦	
c.268G>T	(6: 85189841)	*CSN1S2*	0	0.125	Missense-Deleterious	*p.Asp90Tyr*	*rs430397133*
c.358A>G	(6: 85190192)	*CSN1S2*	0	0.25	Missense-Tolerated	*p.Ile120Val*	*rs424657035*
c.527G>A	(6: 85193644)	*CSN1S2*	0.125	0.75	Missense-Tolerated	*p.Arg176His*	*rs399378277*

CSN1S1

- One missense variant (*c.626C>T*; *rs420959261*) found in Churra and Assaf
- *p.Thr209Ile* substitution
- Differentiates protein variant CSN1S1*C’ (ancestral variant) from CSN1S1*C’’

CSN1S2

- **Splice donor variant**
  - Novel variant
  - Detected in the two breeds (allele frequency of 0.625)
  - Supposed to cause a change in the first base of the two base pair region at the 5’ end of the third intron of the *CSN1S2* gene (*c.102+1G>A*)
  - This SNP could cause intron retention and a novel isoform of *CSN1S2*
- Three missense variants related with known αs2-casein alleles

CSN1S2

- Splice donor variant
- **Three missense variants related with known αs2-casein alleles**
  - Deleterious variant ***c.268G>T*** (*rs430397133*)
    - Causes the *Asp90Tyr* substitution (CSN1S2*A→CSN1S2*B)
      - Advantageous effects of the CSN1S2*B protein allele for milk, fat and protein yield and protein content have been reported (Giambra *et al*., 2014)
    - Detected in one heterozygous Churra ewe
      - The same animal was heterozygous for the other two missense variants (*c.358A>G* and *c.527G>A*)-> CSN1S2*B’
  - Two tolerated variants responsible for αs2-CN protein alleles G (**c.358A>G**) and G’ (**c.358A>G** and **c.527G>A**)

CSN2

- ***c.595A>G* (*rs430298704*)**
  - Detected in one heterozygous Churra ewe
  - Missense variant causing the substitution *p.Met199Val*
  - Causes the CSN2*A and CSN2*G protein alleles of β-casein
  - Corral *et al.*(2010) found that in Merino sheep:
    - GG genotype is associated to an increase in milk production
    - AA genotype is associated to an increase in protein and fat percentage

- ***c.634C>A* (*rs416941267*)**
  - Detected in both breeds
  - Missense variant causing the substitution *p.Leu212Ile*
  - Responsible of the CSN2*X protein allele

Conclusions

- This preliminary analysis focused on the casein gene cluster region identified 735 variants
- High proportion of variants in introns (55.7%)
- We found one novel splice donor variant and six known missense variants in the genes *CSN1S1*, *CSN2* and *CSN1S2*
- Missense variants are related to known caseins protein alleles
- Further research is required to estimate the allele frequencies and determine the phenotypic effects of the variants found through this RNA-Seq approach in commercial sheep populations

Thank you for your attention

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
"Una manera de hacer Europa"

Project: AGL2015-66035-R