



**Universidad de León**

**Facultad de Veterinaria**

**Departamento de Producción Animal**

**UTILIZACIÓN DE HERRAMIENTAS GENÓMICAS EN EL ESTUDIO  
DE LA ARQUITECTURA MOLECULAR DE CARACTERES DE  
IMPORTANCIA ECONÓMICA EN EL GANADO OVINO**

**USING GENOMIC TOOLS FOR THE STUDY OF THE MOLECULAR  
ARCHITECTURE OF ECONOMICALLY IMPORTANT TRAITS IN SHEEP**

**Aroa Suárez Vega**

**León, Septiembre de 2015**



*Las investigaciones de esta Memoria de Tesis Doctoral han sido financiadas por los proyectos AGL2009-07000 y AGL2012-34437 del Ministerio de Economía y Competitividad.*

*La autora de esta Memoria ha sido beneficiaria de una beca de posgrado correspondiente al Programa de Formación de Profesorado Universitario (FPU) del Ministerio de Educación, Cultura y Deporte con referencia AP2009-0909.*





*"It always seems impossible until it's done."*

Nelson Mandela



## **AGRADECIMIENTOS**

Después de todo el trabajo, de los momentos vividos, me siento afortunada al pensar en todas las personas a las que tengo que agradecer el apoyo brindado a lo largo de estos años en la elaboración de esta Tesis Doctoral.

En primer lugar me gustaría agradecer a Juanjo y a Bea, mis directores de tesis, la confianza que han depositado en mí para realizar este proyecto. Gracias por su dedicación, por su paciencia y por la ayuda que me han prestado siempre que lo he necesitado. Gracias por las oportunidades de desarrollo profesional que me habéis dado en este tiempo y también por ayudarme a crecer personalmente enseñándome que con constancia todo se puede aprender.

Je voudrais destiner une reconnaissance à Gwenola et Christophe pour me donner l'occasion de travailler avec eux à l'INRA de Toulouse, pour partager votre connaissance avec moi et pour me faire sentir comme si j'étais chez moi.

A Valentín y a Julio por su participación y colaboración lo largo de estos años en esta Tesis Doctoral. Por la ayuda y los conocimientos anatomopatológicos aportados a la misma.

A ANCHE, en especial a Sandra, a Jorge y a Miguel de S.A.T. SANVI. Sin su ayuda, dedicación y buen hacer en el trabajo muchos de los logros alcanzados en esta Tesis Doctoral no hubiesen sido posibles. A los operarios de la granja de la Universidad de León por su disponibilidad y amabilidad a lo largo de los nueve meses de intenso muestreo en la granja.

A los profesores y PAS del departamento de Producción Animal de la Universidad de León, por su contribución a mi formación y por su ayuda siempre que la he necesitado.

A mis compis de Departamento (tanto actuales como antiguas) por su apoyo, palabras de ánimo y por todos los momentos compartidos. Es una suerte poder decir que durante este tiempo en el Departamento he hecho verdaderos amigos.

A mis amig@s de toda la vida (Marta, Noe, Lore y Pike), porque aunque no tenga tiempo ni para un café sé que siempre están ahí cuando los necesito, que creen en mí

y que me apoyan; gracias por darme los “Buenos Días” todas las mañana cuando más lo necesitaba y por las visitas esté donde esté. A mi familia de León (Marta, Patri, Moni, Lucia y Luis) porque muchas veces confían más en mí que yo misma, y sé que siempre puedo contar con ellos para ir a por la “próxima locura”. A mis chicos y familia de Fonty, sobre todo a Álvaro y Lidia que sufren mis venturas y desventuras más de cerca y a Davicillo que cada vez que viajo por el mundo se acuerda de mí para darme ánimos. A Virgi, por ser como es, escucharme y porque sin ella este último año no habría sido lo mismo, te deseo lo mejor en esta nueva etapa. A María y a Inés, que empezaron viviendo conmigo esta aventura y que ahora vuelan solas, les deseo lo mejor. To Sarah, Stina, Nicole, Belén and Chiara and all the people I met in Toulouse, thank you very much for your support and for make my stage there more comfortable.

A mi familia: a mis abuelos, tíos y primos. Gracias porque aunque muchas veces no tengáis muy claro que es lo que realmente hago siempre estáis ahí preocupándoos por mí, porque este bien y alegrándoos de mis pequeños logros.

A mis padres, José Ramón y M<sup>a</sup> Esther, por estar ahí siempre de manera incondicional y ayudarme paso a paso a conseguir todos mis objetivos. Por tener siempre un buen consejo y una palabra de cariño, porque gracias a vosotros soy lo que soy ahora.

A Dieguichuelas, por creer en mí y apoyarme, por preocuparte por mí y darme consejos como sí tú fueses el hermano mayor, por ser el mejor hermano del mundo.

A Alberto, por tu cariño y comprensión a lo largo de estos años, por estar siempre ahí, tanto en los buenos como en los malos momentos, por creer que puedo con todo, llenarme de tu optimismo y sacarme siempre una sonrisa.

A todos ellos y a los que no he nombrado que han puesto su granito de arena para hacer estos años de trabajo más llevaderos **MUCHAS GRACIAS!**

## ÍNDICE DE CONTENIDOS

|   |    |
|---|----|
| 1. Planteamiento y Objetivos .....  | 1  |
| 2. Introducción general .....   | 7  |
| 2.1. Evolución de la genómica .....   | 9  |
| 2.2. El genoma ovino .....  | 13 |
| 2.2.1. El genoma virtual ovino y el chip de SNPs de 1.5K .....  | 13 |
| 2.2.2. El chip de SNPs ovino de 50K .....   | 15 |
| 2.2.3. El proyecto genoma ovino y el chip de SNPs de 600K.....  | 17 |
| 2.3. Uso de herramientas genómicas para el control de enfermedades hereditarias.....  | 18 |
| 2.3.1 Métodos clásicos de detección de mutaciones responsables de enfermedades.....   | 18 |
| 2.3.2 Utilización del chip de SNP en la identificación de mutaciones en enfermedades de herencia mendeliana .....               | 20 |
| 2.3.3 Utilización de herramientas de secuenciación de nueva generación para la detección de enfermedades mendelianas.....       | 23 |
| 2.3.4 Enfermedades de herencia mendeliana en el ganado ovino.....   | 24 |
| 2.4. Uso de la transcriptómica para el estudio de la lactación.....   | 26 |
| 2.4.1. Herramientas utilizadas en el estudio del transcriptoma.....   | 26 |
| 2.4.2. Aplicaciones de la técnica RNA-Seq en el estudio de la lactación .....   | 28 |
| 2.5. La producción láctea ovina en Castilla y León .....  | 30 |
| 3. Metodología .....  | 31 |
| 3.1. Clonado posicional con el chip de SNPs de 50K ovino de los genes responsables de enfermedades de herencia mendeliana ..... | 33 |

|   |     |
|---|-----|
| 3.1.1. Etapa 1: Detección de la región portadora de genes responsables de enfermedades de herencia mendeliana.....                            | 33  |
| 3.1.2. Etapa 2: Mapeo fino y redefinición del fragmento portador del gen causal .....   | 34  |
| 3.2. Análisis de datos de NGS: flujo de trabajo para el análisis de variantes genéticas en el diagnóstico de enfermedades mendelianas.....    | 35  |
| 3.3. Análisis de datos de RNA-seq: flujo de trabajo para el análisis de expresión diferencial .....   | 36  |
| 4. Resultados .....   | 41  |
| 4.1. Identification of a 31-bp deletion in the RELN gene causing lissencephaly with cerebellar hypoplasia in sheep.....                       | 43  |
| 4.2. Procedimiento y kit de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino.....   | 57  |
| 4.3. Combining GWAS and RNA-Seq approaches for detection of the causal mutation for hereditary junctional epidermolysis bullosa in sheep..... | 73  |
| 4.4. Procedimiento y kit de diagnóstico de epidermólisis bullosa juntural en ganado ovino.....  | 93  |
| 4.5. Initial RNA-seq analysis on the milk transcriptional profiling of two sheep breeds .....   | 109 |
| 4.6. Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing.....                  | 115 |
| 4.7. Estudio del perfil transcripcional de genes implicados en el metabolismo de los ácidos grasos de la leche en dos razas ovinas .....      | 145 |
| 5. Resumen de Resultados y Discusión General.....   | 151 |

|  |     |
|--|-----|
| 5.1. Utilización de herramientas genómicas en el estudio de la base genética de enfermedades monogénicas.....            | 153 |
| 5.1.1. Estudio de la base genética de la lisencefalia con hipoplasia cerebelar en el ganado ovino de raza Churra .....   | 153 |
| 5.1.2. Estudio de la base genética de la epidermólisis bullosa juntural en el ganado ovino de raza Churra .....          | 156 |
| 5.1.3. Utilidad de las herramientas genómicas en la caracterización de enfermedades monogénicas .....                    | 158 |
| 5.2. Utilización de la técnica RNA-Seq en el estudio de la lactación en el ganado ovino .....                            | 161 |
| 5.3. Perspectivas Futuras .....  | 169 |
| 6. Conclusiones .....  | 173 |
| 7. Conclusions .....   | 177 |
| 8. Resumen .....   | 181 |
| 9. Summary .....   | 189 |
| 10. Bibliografía .....   | 197 |
| 11. Anexo I: Artículos Adicionales .....   | 225 |
| 11.1. Hereditary lissencephaly and cerebellar hypoplasia in Churra lambs.....  | 227 |
| 11.2. Generalized severe junctional epidermolysis bullosa with congenital absence of skin in churra lambs.....           | 241 |
| 12. Anexo II: Material Suplementario de los artículos .....  | 255 |
| 12.1. Identification of a 31-bp deletion in the RELN gene causing lissencephaly with cerebellar hypoplasia in sheep..... | 257 |

|  |     |
|--|-----|
| 12.2. Combining GWAS and RNA-Seq approaches for detection of the causal mutation for hereditary junctional epidermolysis bullosa in sheep..... | 271 |
| 12.3. Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing.....                  | 321 |





## **1. Planteamiento y Objetivos**



La presente Tesis Doctoral surge dentro del marco de los cambios y la evolución que han sufrido las herramientas genómicas en los últimos diez años. La aparición de las técnicas de secuenciación de segunda generación (*Next Generation Sequencing*) en la primera década del siglo XXI revolucionó el campo de la genómica permitiendo, junto con desarrollos bioinformáticos paralelos, plantearse la secuenciación de los genomas de los organismos superiores de una manera eficaz y a un coste asequible. Este hecho favoreció que, para la mayoría de las especies domésticas, surgiesen proyectos de secuenciación de los genomas que permitieron el desarrollo de herramientas moleculares, como los chips de polimorfismos de un solo nucleótido o chips de SNPs (del inglés *Single-Nucleotide Polymorphism*) que surgen derivados de la información generada a partir de dichas secuencias. En este punto, la importancia que adquieren las herramientas genéticas en los programas de mejora aplicados en producción animal ha sido cada vez mayor. La disponibilidad de estas nuevas herramientas moleculares ha revolucionado la posibilidad de identificar muchas de las variantes genéticas relacionadas con caracteres simples y complejos, sustituyendo enfoques más tradicionales (estrategia gen candidato y clonado posicional con marcadores microsatélites), por su mayor eficiencia y menor coste.

El planteamiento de la presente Tesis Doctoral se realiza en el año 2009, coincidiendo con la comercialización del chip de SNPs de 50K ovino, que había sido utilizado de forma eficaz para el mapeo de mutaciones causales de enfermedades monogénicas en la oveja (Becker *et al.*, 2010). El grupo de investigación de Mejora Genética Animal, en el que se ha desarrollado la presente Tesis Doctoral, perteneciente al Departamento de Producción Animal de la Universidad de León, trabaja en la mejora genética del ganado ovino de leche, principalmente en la raza Churra. Este grupo colabora con la Asociación de Criadores de Raza Churra (ANCHE) en cuestiones de asesoramiento genético, incluyendo la coordinación de los diferentes programas de selección. Así mismo, otros grupos de investigación de la Facultad de Veterinaria de León, como el Departamento de Sanidad Animal, cooperan con la Asociación en cuestiones de asesoramiento para problemas de índole sanitario y patológico. En este entorno, surge un trabajo de colaboración con el objetivo de estudiar y caracterizar la base molecular de una serie de alteraciones patológicas congénitas que surgen dentro de

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

explotaciones del núcleo de selección de ANCHE cuya principal causa parece ser la elevada consanguinidad de dichas explotaciones. El objetivo inicial de la presente Tesis Doctoral consistía en la utilización del chip de media densidad ovino (50K) para el estudio y detección de mutaciones responsables de enfermedades genéticas de herencia mendeliana en la raza ovina Churra.

Sin embargo, la rápida evolución, eficacia y asequibilidad de las técnicas de secuenciación de segunda generación en los últimos años nos ha permitido aplicar el uso de estas nuevas metodologías en la presente Tesis Doctoral, modificando así el planteamiento inicial de la misma. En concreto, se ha utilizado la técnica de secuenciación masiva paralela del RNA (RNA-Seq) para el estudio tanto de enfermedades mendelianas como de caracteres relacionados con la producción láctea.

La técnica de RNA-Seq permite la secuenciación de todo el transcriptoma con un elevado rendimiento y a un coste relativamente bajo, proporcionando un nuevo escenario en el mapeo y cuantificación de transcriptomas (Wang *et al.*, 2009). Esta técnica ha demostrado ser útil en la identificación de variantes génicas (SNPs e *indels* cortas) en diferentes tejidos (Canovas *et al.*, 2010; Cox *et al.*, 2012; Djari *et al.*, 2013), pudiendo utilizarse, por tanto, en la detección de mutaciones causales.

Así mismo, la utilización de la técnica de RNA-Seq nos ha permitido su aplicación al estudio del transcriptoma mamario de la oveja y la comparación del mismo entre dos razas con diferentes rendimientos productivos, Churra y Assaf. Como ya hemos indicado, uno de los principales objetivos de nuestro grupo de investigación es la mejora de la producción láctea en el ganado ovino, principalmente en la raza Churra, por ello, el conocimiento del perfil transcripcional de la glándula mamaria en lactación es de especial interés, ya que permite la identificación de genes que controlan caracteres complejos relacionados con la producción y la composición de la leche y la detección de variantes génicas que pueden ser utilizadas en la mejora de la producción láctea. En el ganado vacuno, se han realizado diversos estudios a este respecto (Medrano *et al.*, 2010; Wickramasinghe *et al.*, 2012; Canovas *et al.*, 2013; Cui *et al.*, 2014).

En función de lo expuesto, el objetivo general de esta Tesis Doctoral ha sido utilizar las herramientas genómicas disponibles en el estudio de caracteres de importancia económica en el ganado ovino lechero. Este objetivo general se ha desarrollado en los siguientes objetivos concretos:

1. -Utilizar el chip de media densidad ovino (50K) en el estudio de caracteres mendelianos responsables de enfermedades genéticas.
2. -Utilizar las herramientas genómicas de secuenciación masiva paralela en el estudio y caracterización de mutaciones responsables de enfermedades genéticas de herencia mendeliana en el ganado ovino.
3. -Utilizar la técnica de secuenciación masiva paralela de RNA en el estudio y caracterización del transcriptoma mamario de la oveja durante la lactación.





## **2. Introducción General**





## **2.1. Evolución de la genómica**

La genómica es una ciencia relativamente joven. El término genómica fue acuñado por Tom Roderick, genetista del laboratorio Jackson (Maine, USA) durante un congreso sobre el mapeo del genoma humano en 1986 (Kuska, 1998) y fue utilizado por primera vez en la revista que lleva el mismo nombre (*Genomics*) en 1987 (McKusick and Ruddle, 1987). La Genómica se define como la rama de la genética que se ocupa del estudio integral del funcionamiento, contenido, evolución y origen de los genomas. La obtención de información en profundidad de la secuencia genómica de la mayoría de mamíferos y su análisis nos permitirían entender las variaciones genómicas relacionadas con determinados fenotipos, entre los que se incluyen susceptibilidad genética a las enfermedades, caracteres económicos e incluso la respuesta genética a determinados fármacos (Pareek *et al.*, 2011).

La evolución de las herramientas para el estudio de la genómica va ligada al proyecto de secuenciación del genoma humano. El término secuenciación del DNA se refiere a los métodos utilizados para determinar el orden de los nucleótidos (Adenina, Timina, Citosina y Guanina) en una molécula de DNA. En 1977, Fred Sanger y Alan R. Coulson publicaron dos artículos metodológicos sobre técnicas de secuenciación (Sanger *et al.*, 1977a; Sanger *et al.*, 1977b) que pasarían a transformar la biología en su conjunto, proporcionando una herramienta para descifrar genes completos y más tarde genomas enteros. La técnica *Sanger*, basada en el método enzimático de terminación de cadenas mediante la utilización de didesoxinucleótidos, se impuso sobre la alternativa publicada en el mismo año por Maxam y Gilbert (Maxam and Gilbert, 1977) debido a las ventajas que ofrecía a nivel de manejo de productos tóxicos. Como resultado, la “secuenciación *Sanger*” fue el método rutinario de secuenciación durante los siguientes treinta años, siendo la metodología utilizada para la obtención de la primera secuencia consenso del genoma humano (Lander *et al.*, 2001; Venter *et al.*, 2001) y de la primera secuencia diploide individual humana (J. Craig Venter) (Levy *et al.*, 2007).

Sin embargo, aunque la tecnología *Sanger* posee una elevada precisión en la determinación de las secuencias de DNA, el rendimiento de esta técnica es bajo. Coincidiendo con la finalización del primer boceto del genoma humano (International

### **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

Human Genome Sequencing Consortium, 2004) se puso en marcha un proyecto cuyo objetivo principal era la reducción del coste de la secuenciación del genoma humano a 1000 \$ en el periodo de diez años (Schloss, 2008). Este proyecto supuso el comienzo de una nueva era en la secuenciación del DNA, resultando en una mejora del desarrollo de las estrategias de secuenciación hacia metodologías de alto rendimiento, denominadas secuenciación de segunda generación, secuenciación masiva paralela o en inglés *Next Generation Sequencing* (NGS).

Las técnicas de secuenciación masiva paralela o NGS producen lecturas más cortas que oscilan entre los ~30 y ~500 pb, en comparación con la secuenciación *Sanger* (entre 700 y 1000 pb). Sin embargo, las NGS superaron la limitada capacidad de escalabilidad de la metodología *Sanger* mediante la creación de micro-reactores y/o superficies de fijación de moléculas de DNA que permiten que millones de reacciones de secuenciación ocurran en paralelo. Las NGS, por tanto, presentan dos ventajas fundamentales respecto a la metodología anterior: la miniaturización y la paralelización de las reacciones de secuenciación que permiten la secuenciación de genomas enteros a una velocidad sin precedentes y a un coste relativamente bajo (Pettersson *et al.*, 2009). Existen cuatro técnicas desarrolladas por diferentes laboratorios que han sido las más utilizadas en los últimos años: éstas son las desarrolladas por *Roche-454* (pirosecuenciación), *Ion Torrent* (secuenciación por semiconductores), *Illumina/Solexa* (secuenciación con terminadores reversibles) y *SOLiD* (secuenciación por ligamiento), todas ellas resumidas en la Tabla 1 y revisadas en las siguientes referencias bibliográficas (Metzker, 2010; Schadt *et al.*, 2010; Liu *et al.*, 2012). Los puntos en común de los diferentes métodos de secuenciación masiva paralela serían los siguientes: (i) fragmentación del material genético, (ii) incorporación de adaptadores en los extremos de los fragmentos, (iii) amplificación clonal (ej. PCR de emulsión o PCR tipo puente), (iv) secuenciación de los fragmentos (ej. secuenciación por síntesis o secuenciación por ligamiento), (v) recopilación de los datos brutos (ej. captura de imágenes o detección de iones), (vi) conversión de los datos brutos a secuencias de nucleótidos.

**Tabla 1.** Principales plataformas de secuenciación de segunda generación (Liu *et al.*, 2012; Desai and Jere, 2012; Diaz-Sanchez *et al.*, 2013; van Dijk *et al.*, 2014)

**Roche 454: pirosecuenciación**

**Técnica.** Fue la primera tecnología NGS que tuvo éxito comercial. Primero se añaden adaptadores específicos al DNA fragmentado (obtenido en la preparación de las genotecas) y cada uno de estos fragmentos es capturado por unas microesferas donde se realizará la PCR, dentro de gotas de una emulsión de lípidos-agua (*Emulsion-PCR*). Estas esferas se distribuyen en celdas en una placa de fibra óptica donde se realizará la pirosecuenciación. La pirosecuenciación se basa en la detección del pirofosfato liberado durante la incorporación de los nucleótidos. Se van añadiendo los nucleótidos (A, T, C, G) uno por uno. Cuando el nucleótido añadido es complementario de la cadena molde el pirofosfato liberado hace que se forme ATP y mediante una reacción enzimática se produce una transformación de luciferina (presente en el pocillo) en oxi-luciferina que hace que se emita luz.

**Pros.** Lecturas largas (1 Kb máximo) fáciles de mapear frente a un genoma de referencia, lo que supone una ventaja a la hora del ensamblado *de novo* y para experimentos de metagenómica. Tiempos de secuenciación relativamente rápidos (~23 h).

**Contras.** El rendimiento es menor que sus competidores (aprox. 1 millón de lecturas, 700 Mb de datos de secuenciación) y altos costes de reactivos. Altas tasas de error en repeticiones de homopolímeros.

**Illumina: secuenciación con terminadores reversibles**

**Técnica.** Al igual que en la técnica anterior, se añaden adaptadores a los fragmentos de DNA generados en el paso previo de creación de las genotecas. Estos fragmentos se desnaturalizan y se adhieren a la placa de secuenciación. Se produce una amplificación en modo puente (*bridge-PCR*) donde se generan grupos (*clusters*) de amplificados del mismo fragmento. Para la secuenciación los cuatro tipos de nucleótidos (A, T, C, G) son añadidos a la vez en cada paso de la secuenciación, cada uno de ellos está marcado con un fluoróforo distinto y dependiendo de cuál sea el complementario de cada clúster se emitirá una luz específica que será capturada secuencialmente por una cámara.

**Pros.** Actualmente lidera la industria de las NGS y la mayoría de los protocolos de preparación de genotecas son compatibles con el sistema de *Illumina*. Ofrece el rendimiento más alto al precio más bajo. Lecturas de hasta 300 pb, compatibles con la mayoría de experimentos NGS.

**Contras.** La carga de las muestras es técnicamente complicada debido a la dispersión al azar de los grupos de amplificación a lo largo de la celda de flujo. Concentraciones elevadas pueden provocar que se superpongan los grupos de amplificación y disminuya la calidad de la secuenciación. Muestras de baja complejidad deben diluirse o mezclarse con una genoteca de referencia (PhiX) para generar diversidad.

**SOLiD: secuenciación por ligamiento**

**Técnica.** Al igual que en 454 los fragmentos de DNA se amplifican en microesferas por *Emulsion-PCR*. La secuencia se determina mediante hibridación y ligamiento. Un *primer* hibrida con el adaptador quedando el extremo 5' libre. Una mezcla de octámeros marcados con color compiten por unirse a la cadena molde (cuatro colores codifican las bases de las posiciones 4 y 5). Una vez detectado el color, el octámero se rompe entre las posiciones 5 y 6. En la primera ronda de secuenciación se determinan las posiciones 4, 5, 9, 10, 14, 15 etc. El proceso se repite con un *primer* más corto para determinar las posiciones 3, 4, 8, 9, 13 y 14. Cada nucleótido se lee dos veces. En cada ciclo se añaden *primers* más cortos hasta que se lee la posición 0.

**Pros.** Elevado rendimiento y tasas de error bajas, 99,94% de precisión.

**Contras.** Lecturas cortas (~75 pb) y tiempos de secuenciación elevados, por ello es más complicado utilizar estas lecturas en ensamblado *de novo*.

**Ion Torrent: secuenciación por semiconductores**

**Técnica:** Muy similar a 454 pero en lugar de un pirofosfato lo que se libera es un protón, que es detectado por un sensor de iones.

**Pros.** No requiere escáner óptico ni nucleótidos fluorescentes. Tiempos de secuenciación rápidos (horas). Lecturas compatibles con la mayoría de experimentos NGS.

**Contras.** Altas tasas de error en homopolímeros.

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

Las mejoras realizadas en los últimos años en las técnicas NGS y la reducción drástica de los precios de secuenciación (democratización de los precios de secuenciación) no sólo han permitido la obtención de genomas de otras especies de interés (Li *et al.*, 2010; Dalloul *et al.*, 2010; Dong *et al.*, 2013; Jiang *et al.*, 2014), sino que han propiciado la aplicación de estas tecnologías en ámbitos más amplios (Soon *et al.*, 2013), como (i) la resecuenciación de genomas o exomas para detectar variantes relacionadas con caracteres de interés, adquiriendo esta aplicación importancia tanto desde el punto de vista de la genética aplicada a la clínica (Gonzaga-Jauregui *et al.*, 2012) como a caracteres productivos (Larkin *et al.*, 2012; Daetwyler *et al.*, 2014), (ii) la secuenciación de transcriptomas que permiten identificar los transcritos expresados en un determinado tejido (RNA-Seq) (Mortazavi *et al.*, 2008; Wang *et al.*, 2009), (iii) el mapeo de regiones con marcas epigenéticas (Methyl-seq) (Laird, 2010) o (iv) el mapeo regiones de unión de factores de transcripción (ChIP-seq) (Park, 2009).

Cabe señalar que el incremento de la utilización de las técnicas NGS también se ha debido en parte a los avances significativos en el manejo de los datos generados. El desarrollo de las técnicas NGS ha provocado la aparición de forma paralela de una nueva generación de herramientas bioinformáticas como requisito esencial para el manejo e interpretación de la gran cantidad de datos generados. Además, la generación de secuencias de una longitud relativamente corta, comparada con la secuenciación *Sanger*, ha precipitado el desarrollo de nuevos algoritmos de alineamiento. Se han desarrollado algoritmos para el manejo específico de las lecturas cortas (Hatem *et al.*, 2013), así como para el ensamblado *de novo* (Zhang *et al.*, 2011), la detección de variantes (Li *et al.*, 2013), los análisis ChIP-seq (Rougemont and Naef, 2012) y los análisis RNA-Seq (Kvam *et al.*, 2012). Así mismo, se han desarrollado algoritmos para corregir los posibles sesgos producidos durante la preparación de las genotecas (Hansen *et al.*, 2010). La combinación de las mejoras tanto técnicas como computacionales ha favorecido el aumento en la utilización de las técnicas NGS maximizando la cantidad de información que es posible obtener de los datos de secuenciación masiva paralela.

Actualmente, *Illumina* lidera el mercado de las tecnologías NGS ofreciendo el rendimiento más alto y el precio más bajo por base secuenciada (van Dijk *et al.*, 2014). Sin embargo, la secuenciación de tercera generación está empezando a aparecer. La principal característica de la secuenciación de tercera generación es la lectura directa de moléculas individuales de ácidos nucleicos (Schadt *et al.*, 2010). Estas técnicas intentan solventar las flaquezas de las NGS utilizando una sola molécula de DNA, es decir, prescindiendo de una etapa de amplificación previa (PCR). Esto supone ventajas ya que se aumenta la tasa de secuenciación, los rendimientos son más altos, las lecturas son más largas (~10-30 Kb), disminuye la complejidad de la preparación de las muestras y se reducen los costes del análisis (Schadt *et al.*, 2010). Dos ejemplos de estas tecnologías de tercera generación son: el secuenciador *PacBio RS II* de *Pacific Biosciences* que se basa en la técnica SMRT (*Single-Molecule Real-Time*), en la que una sola molécula de DNA ligada a una polimerasa es insertada en una cámara ZMW (guía de onda de "modo cero"). Esta cámara capta la luz emitida por cada nucleótido marcado con un fluoróforo que la polimerasa incorpora durante la síntesis de la cadena complementaria, permitiendo conocer la secuencia (Eid *et al.*, 2009). Por otro lado, *Oxford Nanopore Technologies* ha desarrollado tres dispositivos *MinION*, *PromethION* y *GridION*, que se basan en el paso de una corriente iónica a través de nanoporos y la medición de los cambios en la corriente cuando el material genético pasa a través del poro, lo que permite identificar las distintas bases y generar la secuencia (Clarke *et al.*, 2009). Aunque existen trabajos publicados que se basan en estas estrategias (Chaisson *et al.*, 2015; Loman *et al.*, 2015), ambas metodologías están en una fase preliminar en las que las tasas de error son elevadas y deben corregirse aumentando la profundidad de secuenciación o combinando estas lecturas con lecturas cortas de alta calidad de secuenciadores NGS (Koren *et al.*, 2012; Chin *et al.*, 2013; Madoui *et al.*, 2015).

## 2.2. El genoma ovino

### 2.2.1. El genoma virtual ovino y el chip de SNPs de 1.5K

La secuenciación del genoma humano (Lander *et al.*, 2001; Venter *et al.*, 2001) fue un importante hito científico e histórico que abrió el camino para la secuenciación del

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

genoma de otras especies de mamíferos. El elevado coste que suponía en aquel momento la secuenciación de un genoma (tres mil millones de dólares el genoma humano; van Dijk *et al.*, 2014), hizo que únicamente surgiesen proyectos de secuenciación genómica para aquellas especies ganaderas con producciones de gran repercusión económica que justificase la inversión del proyecto. Las primeras especies ganaderas con un borrador secuenciado del genoma fueron la gallina (International Chicken Genome Sequencing Consortium, 2004) y la vaca (<https://www.hgsc.bcm.edu/other-mammals/bovine-genome-project>, (Bovine Genome Sequencing and Analysis Consortium *et al.*, 2009), ambos en el año 2004.

En el año 2002, se constituye el *International Sheep Genomics Consortium* (ISGC) compuesto por científicos y entidades financiadoras de numerosos países (<http://www.sheephapmap.org/>). En el caso del ganado ovino, no se disponía de recursos para financiar un proyecto de secuenciación genómica, por lo que esta entidad surgió con el propósito central de realizar la secuenciación parcial del genoma ovino para la identificación de variantes de un solo nucleótido o SNPs (de la siglas en inglés *Single-Nucleotide Polymorphism*) y el desarrollo, a partir de estos datos, de herramientas genómicas útiles para la mejora de caracteres productivos y sanitarios en la especie ovina (International Sheep Genomics Consortium *et al.*, 2010).

El primer paso, fue crear un genoma virtual que permitiese el posicionamiento y orientación de las secuencias para que se pudiesen seleccionar regiones uniformemente espaciadas para la resecuenciación y detección de SNPs. Para ello se secuenciaron mediante la técnica *Sanger* extremos de una genoteca de plásmidos tipo BAC del genoma ovino de un macho de raza Texel (CHORI-243). Posteriormente, estas secuencias se mapearon frente a los genomas bovino (Btau2.0), canino (Lindblad-Toh *et al.*, 2005) y humano (International Human Genome Sequencing Consortium, 2004), que eran las especies filogenéticamente más próximas a la oveja con genomas secuenciados. Este mapeo permitió la ordenación y la creación de secuencias de mayor longitud por comparación entre especies. Este protocolo proporcionó una cobertura del 91,2% del genoma humano (Dalrymple *et al.*, 2007). Las secuencias generadas se reorganizaron para formar los cromosomas ovinos basándose en las posiciones

conocidas de los marcadores en el mapa de ligamiento ovino disponible (versión 4.6 del mapa australiano). Como resultado se obtuvo el genoma virtual de la oveja (*Virtual Sheep Genome*), que cubría el 76% del genoma de esta especie y proporcionaba la primera secuencia de referencia disponible en el ganado ovino (Dalrymple *et al.*, 2007).

Con la información proporcionada por el *Virtual Sheep Genome*, se puso en marcha el proyecto piloto para el desarrollo del *chip* de SNPs de 1.5K. El objetivo principal de este proyecto era evaluar la utilidad de un *array* de SNPs para la investigación genómica en ovino. Para la elaboración del *chip* de 1.5K se secuenciaron, por la metodología *Sanger*, nueve animales de diferentes razas (Awasi, Gulf Coast Naive, Katahdin, Lacaune, Merino, Poll Dorset, Red Masai, Rommey y Texel) en total 2.644 loci genómicos; 350 asociados con regiones codificantes (ESTs) y 2.294 seleccionados de la genoteca de las secuencias terminales de los clones BAC. Previamente se había comprobado su localización en el genoma y que no eran secuencias repetidas (Kijas *et al.*, 2009). En total se detectaron 6.021 SNPs con una densidad media de 4,9 SNPs por Kb. Para la elaboración del *chip* se seleccionaron 1.536 SNPs que representaban a la mayoría (952/1.142) de los *contigs* creados en la elaboración del genoma virtual ovino (Kijas *et al.*, 2009; Kijas *et al.*, 2012).

### 2.2.2. El *chip* de SNPs ovino de 50K

Paralelamente a la elaboración del *chip* de SNPs de 1.5K, surgen las técnicas de secuenciación de segunda generación. La aparición de estas técnicas hizo que, como se ha dicho en el apartado anterior, se incrementasen los rendimientos y que los tiempos y costes de secuenciación disminuyesen, lo que permitió pasar de un genoma virtual de la oveja a un genoma real. El objetivo era obtener un genoma real de baja cobertura (3x) que permitiese generar un *chip* de SNPs de 50K. Para su elaboración se secuenció a una cobertura 0,5x el genoma de seis animales pertenecientes a las razas Awassi, Merino Australiano, Poll Dorset, Romney, Scottish Black Face y Texel, utilizando la técnica de secuenciación de segunda generación *Roche 454 FLX*, basada en pirosecuenciación. Las secuencias obtenidas fueron alineadas frente al genoma bovino y se reorganizaron utilizando el genoma virtual de la oveja. La técnica de

### **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

secuenciación 454 hizo que se produjesen más de 9 Gb de datos, permitiendo la detección de más de 590.000 SNPs dentro de regiones genómicas definidas, de los cuales más de 270.000 se clasificaron como SNPs de “clase A” (ambos alelos han sido detectados en más de dos animales). Para la creación del chip *OvineSNP50 BeadChip*, comercializado por la empresa *Illumina* en el año 2009, se seleccionaron 54.241 SNPs de los 270.000 de “clase A” previamente identificados. Para esta selección se marcaron una serie de criterios de selección: los SNPs deberían tener una distribución uniforme a lo largo de todo el genoma, la frecuencia del alelo menor (MAF) debía ser superior a 0,2, el *probe score* debía ser superior a 0,8, el SNP debía presentar una secuencia flanqueante única de 200 pb y se aplicaron además criterios de control de calidad respecto a la técnica de detección y respecto al oligonucleótido fijado al *array*.

La utilización de estos chips de SNPs en las distintas especies domésticas permite a los científicos realizar estudios de asociación en todo el genoma a un relativo bajo coste, hecho que habría sido impensable con la utilización de los marcadores tradicionales. El éxito de la utilización de estos *arrays* reside en el hecho de que presentan una fuerte capacidad de procesamiento en paralelo, miniaturización y una notable capacidad de automatización (Eggen, 2012). En el caso del ganado ovino, el chip de SNPs de 50K se utilizó por primera vez en el proyecto internacional *Sheep Hapmap*, coordinado por el ISGC. En él se genotiparon un total de 2.819 animales de 74 razas ovinas (Kijas *et al.*, 2012). En los últimos años, varios estudios han sido publicados demostrando la efectividad del uso del chip de 50K en el ganado ovino:

- para el mapeo de mutaciones causales en enfermedades genéticas mediante la utilización de un número reducido de casos y controles (Becker *et al.*, 2010; Momke *et al.*, 2011),
- para el mapeo de *loci* relacionados con caracteres cuantitativos (QTL) (García-Gamez *et al.*, 2011; García-Gamez *et al.*, 2012a),
- para la caracterización de la estructura de la población, así como para documentar mejor y entender la evolución de las distintas especies domésticas (Kijas *et al.*, 2012).



### 2.2.3. El proyecto genoma ovino y el chip de SNPs de 600K

Aunque existían borradores del genoma ovino, el proyecto de secuenciación del genoma de la oveja propiamente dicho surge en el año 2009. La aparición de las técnicas de secuenciación de segunda generación y la idoneidad en la utilización de lecturas cortas en la secuenciación de genomas complejos demostrada en el borrador de la secuencia del panda gigante (Li *et al.*, 2010) impulsó al consorcio a planificar un proyecto para elaborar el genoma referencia de la especie ovina. Para ello se secuenció el genoma de dos individuos de raza Texel utilizando la técnica NGS desarrollada por *Illumina*, de secuenciación con terminadores reversibles. Las secuencias se reorganizaron utilizando mapas de ligamiento y de híbridos irradiados. La versión final del borrador del genoma ovino (Oar\_v3.1), publicada en el 2014 (Jiang *et al.*, 2014), tiene una buena calidad de secuencia ya que el parámetro N50 tiene un valor aproximado de 40 Kb y una longitud de ensamblado total de 2,62 Gb, de la que aproximadamente el 99% se reorganizó en los 26 autosomas ovinos y el cromosoma X (Jiang *et al.*, 2014). La anotación del genoma ovino, disponible en el repositorio público de genomas *Ensembl* (<http://www.ensembl.org/>), ha sido realizada con una metodología mixta. Por un lado, se han generado 66.797 modelos de proteínas basados en la similitud que las proteínas ovinas tienen con proteínas ortólogas de otros vertebrados de la base de datos *UniProtKB*. Por otro lado, se utilizaron datos procedentes de experimentos de RNA-Seq, obtenidos a partir de muestras de 94 tejidos, proporcionados por el ISGC. En total en la base de datos de *Ensembl* hay anotados 20.921 proteínas codificantes, 3.985 RNAs cortos no codificantes y 291 pseudogenes ([http://www.ensembl.org/Ovis\\_aries/Info/Annotation](http://www.ensembl.org/Ovis_aries/Info/Annotation)).

A pesar de la importancia que supuso el chip de 50K para el desarrollo del genoma ovino, se ha demostrado que debido a la menor extensión del desequilibrio de ligamiento en esta especie (en comparación con el ganado de raza Holstein), se necesitaría una mayor densidad de marcadores tanto para el mapeo eficiente de QTL de caracteres complejos así como para el uso eficiente de estos *arrays* en selección genómica (García-Gamez *et al.*, 2012b). Esto hace que por parte del ISGC surja la necesidad de desarrollar un chip de marcadores de alta densidad (600K). El ISGC en colaboración con *FarmIQ* ([www.farmiq.co.nz](http://www.farmiq.co.nz)) han desarrollado el *Ovine Infinium® HD*

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

*SNP BeadChip* comercializado por *Illumina* en 2013. Este chip tiene un total de 603.350 SNPs, entre los que se incluyen la mayor parte de los marcadores del chip *OvineSNP50 BeadChip*. Para su desarrollo se han utilizado la secuencia actual del genoma ovino (Oar\_v3.1) y los datos de resecuenciación de 73 animales (Anderson, 2014).

### **2.3. Uso de herramientas genómicas para el control de enfermedades hereditarias**

Tradicionalmente, las asociaciones de criadores han centrado sus esfuerzos en la selección de aquellos animales con mejores índices productivos. La productividad de ciertas razas ovinas ha mejorado considerablemente debido a la implementación de programas de mejora basados en la genética cuantitativa. Sin embargo, aunque la selección incrementa la productividad, el tamaño efectivo de la población se reduce, particularmente cuando unos pocos machos mejorantes se utilizan de forma intensiva. El aumento de la consanguinidad provoca la aparición de enfermedades genéticas recesivas que causan pérdidas económicas y una disminución del bienestar animal (Greber *et al.*, 2013).

Es importante que los criadores lleven a cabo estrategias que les permitan controlar el incremento de la consanguinidad y el número de defectos recesivos que se derivan de este incremento. Seleccionar animales no portadores de enfermedades es interesante desde varios puntos de vista, ya que permite la elección de aquellos progenitores que aseguren una descendencia libre de taras genéticas, así como la disminución de la frecuencia de tratamientos, aumentando el estado de salud de los animales y mejorando su nivel de bienestar. Una de las vías que nos permite lograr este objetivo es la identificación de los genes y mutaciones que originan enfermedades o malformaciones (Luhken, 2012).

#### **2.3.1 Métodos clásicos de detección de mutaciones responsables de enfermedades**

En el pasado la mayor parte de las mutaciones causales de enfermedades eran identificadas mediante dos estrategias: secuenciación de genes candidatos funcionales o mediante análisis de ligamiento.

La estrategia del gen candidato consiste en el estudio detallado de un gen del que, por su función biológica, se sospecha que puede contener la mutación responsable de la aparición de la enfermedad (Kwon and Goate, 2000). Esta aproximación requiere un conocimiento previo de fisiología de la enfermedad. En el caso del ganado ovino, los genes seleccionados como candidatos funcionales de fenotipos específicos proceden, en general, de estudios en otras especies, como la humana, la bovina o el ratón. En el estudio de la hipotricosis en el ganado ovino se seleccionó el gen candidato *hairless* (*HR*) (Finocchiaro *et al.*, 2003) en base a estudios previos que relacionaban un fenotipo similar con mutaciones en el gen *HR* en humano y ratón (Ahmad *et al.*, 1998; Cachon-Gonzalez *et al.*, 1999). Esta aproximación requiere la secuenciación de todo el gen objeto de estudio (exones, intrones y regiones promotoras) en busca de variantes (SNPs, inserciones, deleciones) entre casos y controles. Se trata de una técnica laboriosa y cara; por ello para la secuenciación de algunos genes de gran tamaño se opta por comenzar la secuenciación por las regiones codificantes para la proteína (Hirschhorn and Daly, 2005). Otra limitación a tener en cuenta es que, en muchos casos, una misma enfermedad puede ser genéticamente heterogénea, por lo que podríamos seleccionar un gen como candidato cuando la mutación causal se encuentra en otro gen cuya proteína tiene una función fisiológica parecida o complementaria.

Los estudios de análisis de ligamiento, se basan en el mapeo de loci concretos mediante la utilización de marcadores genéticos en individuos emparentados. Esta técnica, al contrario que la anterior, no requiere un conocimiento previo de la fisiología de la enfermedad. Para el mapeo del locus, se selecciona una serie de familias donde la enfermedad está segregando, genotipándose esas familias para un determinado número de marcadores genéticos informativos distribuidos por todo el genoma, en la estrategia clásica se utilizan generalmente marcadores microsatélite. El análisis de ligamiento nos permitirá calcular la distancia entre el marcador y el locus responsable de una enfermedad al analizar la cosegregación entre los marcadores y el fenotipo. En general, en este tipo de estudios, las regiones obtenidas como significativas exceden los 10 cM (unos 10 millones de pares de bases) por lo que se requieren estudios posteriores empleando mayor número de animales y marcadores genéticos (mapeo fino) para la localización del gen responsable de la enfermedad (Hirschhorn and Daly,

### **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

2005). Aunque se trata de una tarea compleja y costosa, esta estrategia ha permitido el mapeo de genes relacionados tanto con caracteres productivos como con enfermedades en el ganado ovino. Como ejemplo, gracias a esta técnica se logró identificar la mutación causal del fenotipo *Booroola (FecB)* que incrementa la tasa de ovulación en el ganado ovino (Montgomery *et al.*, 1994; Mulsant *et al.*, 2001). También se localizaron las mutaciones responsables de algunas enfermedades de herencia simple como las variantes *CLN5* y *CLN6* de la lipofuscinosis ceroides neuronal ovina (Broom *et al.*, 1998; Frugier *et al.*, 2008) o el síndrome *Spider lamb* (Cockett *et al.*, 1999).

#### **2.3.2 Utilización del chip de SNP en la identificación de mutaciones en enfermedades de herencia mendeliana**

El genotipado con los chips de SNPs en combinación con las estructuras poblacionales típicas de nuestros animales domésticos permite acelerar la localización e identificación de las mutaciones causales en enfermedades de herencia mendeliana (Charlier *et al.*, 2008). El desarrollo de chips de SNPs viene ligado a la creación de los borradores de las secuencias genómicas de las distintas especies animales. La aparición de los *arrays* de SNPs ha permitido el desarrollo de un tipo de estrategia para el mapeo de mutaciones causales distinta a las utilizadas tradicionalmente (estrategia del gen candidato y análisis de ligamiento con marcadores microsatélite): el estudio de asociación a nivel del genoma completo (en inglés, GWAS (*Genome-wide association study*)). Los estudios de asociación se basan básicamente en la comparación de genotipos entre dos grupos de individuos: los casos (afectados de un determinado fenotipo) y los controles (los animales con el fenotipo “salvaje”). En el caso de los caracteres mendelianos, al realizar el análisis estadístico de los genotipos obtendremos evidencia de asociación en la región concreta del genoma en la que las variantes alélicas estén sobre-representadas en los casos respecto a los controles. Esta región será la candidata a contener la mutación causal de la enfermedad. El alto número de variantes alélicas que pueden ser genotipadas a lo largo de todo el genoma gracias a los chips de SNPs ha hecho posible analizar todo el genoma de forma simultánea de forma que la localización primaria de la región responsable suele ser muy eficiente

(Hirschhorn and Daly, 2005). En el caso de estudio de enfermedades con una base genética más compleja que los clásicos caracteres mendelianos, el número de regiones que muestran evidencia estadística de asociación será superior y el tamaño muestral necesario para un diseño eficiente más elevado.

En general, en la búsqueda de mutaciones causales de enfermedades, los análisis de asociación suelen combinarse con otro tipo de análisis denominado de identidad por descendencia (Charlier *et al.*, 2008). En este análisis, los genotipos obtenidos con el chip de SNPs en los animales enfermos se analizan en busca regiones de homocigosis compartidas a lo largo de todo el genoma; la región en la que se localice el mayor bloque de homocigosis será la que más probablemente contenga la mutación responsable de la enfermedad.

Tanto los análisis de asociación como los de identidad por descendencia se basan en el principio de desequilibrio de ligamiento, en la que la variante causal y el alelo del marcador se heredarán juntos por encontrarse físicamente próximos. Estas metodologías, al igual que los análisis de ligamiento con marcadores microsatélite, no requieren conocimiento previo de la fisiología de la enfermedad. Sin embargo, presentan ventajas frente a este:

- No requieren el estudio de familias.
- Necesitan un menor tamaño muestral, llegándose a la determinación de la mutación causal con un modesto número de casos. Por ejemplo, para la determinación de la mutación causal de la epidermólisis bullosa en el ganado bovino de raza Rotes Höhenvieh sólo se utilizaron tres casos (Menoud *et al.*, 2012).
- El proceso de genotipado es menos laborioso (permite automatización) y por lo tanto es más rápido y económico (Eggen, 2012).

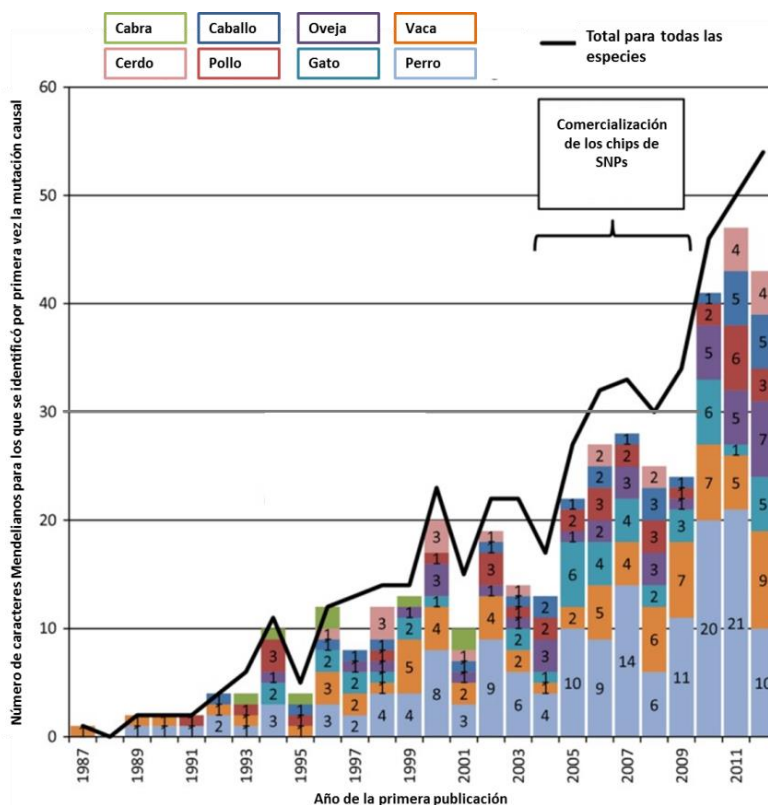
El uso de estas herramientas ha revolucionado la detección de mutaciones causales en las especies domésticas. En la Figura 1, obtenida de la revisión bibliográfica realizada por Nicholas and Hobbs (2014), se puede observar el rápido incremento producido en la detección de mutaciones causales en las especies domesticas a partir del año 2005, coincidiendo con la aparición de los primeros chips de SNPs, en pollo (Wong *et al.*,

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

2004), perro (Lindblad-Toh *et al.*, 2005), gato (Pontius *et al.*, 2007), vaca (Charlier *et al.*, 2008), caballo (Wade *et al.*, 2009), cerdo (Amaral *et al.*, 2009) y oveja (International Sheep Genomics Consortium *et al.*, 2010).

En el caso del ganado ovino, el chip de SNPs de 50K desarrollado por *Illumina* ha demostrado ser una herramienta útil para la identificación de mutaciones causales en enfermedades mendelianas como la microftalmia (Becker *et al.*, 2010), la epidermolísis bullosa (Momke *et al.*, 2011; Suarez-Vega *et al.*, 2015), el raquitismo (Zhao *et al.*, 2011), la condrodisplasia (Zhao *et al.*, 2012b) o la enfermedad de neurona motora inferior (Zhao *et al.*, 2012a).

**Figura 1.** Cronología de la detección de mutaciones causales de caracteres de herencia Mendeliana en las especies domésticas hasta el año 2012 (tomada de Nicholas and Hobbs, 2014).



### *2.3.3. Utilización de herramientas de secuenciación de nueva generación para la detección de enfermedades mendelianas*

La evolución de las técnicas NGS junto con la disminución de los costes de secuenciación ha hecho estas técnicas más accesibles, incrementado significativamente los ámbitos de su aplicación en el análisis de genomas. Si nos centramos en las aplicaciones diagnósticas, la resecuenciación de genomas permite el análisis simultáneo de todas las variantes de un genoma en una única reacción y con un número reducido de casos, demostrando ser una alternativa eficaz para establecer la base de genética de enfermedades de herencia mendeliana (Cirulli and Goldstein, 2010; Boycott *et al.*, 2013) y para su aplicación en el ámbito clínico (Saunders *et al.*, 2012; Willig *et al.*, 2015). Sin embargo, la secuenciación de genomas enteros, en ciertas ocasiones, es poco práctico e innecesario. El exoma, parte codificante del genoma, representa aproximadamente el 2% del genoma humano y en él se encuentran aproximadamente el 85% de las mutaciones causales conocidas (Hodges *et al.*, 2007). Esto hace que para la búsqueda de mutaciones causales en enfermedades monogénicas, la aproximación más comúnmente utilizada en seres humanos sea la secuenciación del exoma o en inglés, *exome-sequencing* (Bamshad *et al.*, 2011; Ku *et al.*, 2012; Yang *et al.*, 2013; Rabbani *et al.*, 2014). Una variante del *exome-sequencing* sería el estudio de un grupo de genes que se sabe que tienen relación directa con el carácter objeto de estudio, también denominado *target-sequencing* (Rehm, 2013). Los experimentos de secuenciación masiva paralela del transcriptoma (RNA-Seq) también han sido utilizados para la detección eficiente y rentable de variantes en diferentes tejidos y especies (Canovas *et al.*, 2010; Djari *et al.*, 2013). En el caso de especies animales en las que la anotación de los genomas está en una etapa temprana de desarrollo, una alternativa eficaz a la secuenciación del exoma para la detección de variantes en regiones codificantes del genoma podría ser la técnica RNA-Seq.

En el caso de las especies ganaderas, la disponibilidad de genomas de referencia sumada a unos costes de secuenciación más asequibles, han hecho que comiencen a aparecer publicados estudios que utilizan exitosamente las técnicas de secuenciación de segunda generación en la identificación de la causa genética de enfermedades mendelianas. Así han surgido estudios en el ganado bovino (Daetwyler *et al.*, 2014;

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**







Murgiano *et al.*, 2014; Peters *et al.*, 2015; Sartelet *et al.*, 2015), ganado porcino (Qiao *et al.*, 2015) o ganado ovino (Suarez-Vega *et al.*, 2015) en los que técnicas NGS han sido utilizadas en caracterización de mutaciones responsables de enfermedades monogénicas.

### **2.3.4. Enfermedades de herencia mendeliana en el ganado ovino**




En la Tabla 2 se resumen las enfermedades ovinas para las cuales la mutación causal ha sido identificada. Para la realización de la misma se ha tomado como referencia la base de datos online *Mendelian Inheritance in Animals* (OMIA) (<http://omia.angis.org.au/>), consultada el 15 de Julio de 2015. En el caso del ganado ovino, esta base de datos recoge 47 caracteres mendelianos para los que se conoce la mutación causal. De ellos, 22 hacen referencia a enfermedades (recogidos en la Tabla 2), 20 a caracteres productivos (fecundidad, hipertrofia muscular, ausencia de cuernos, grasa amarilla), tres al color y dos a la resistencia/susceptibilidad a enfermedades (scrapie y lentivirus). Como se puede observar en la Tabla 2, los estudios realizados mediante la secuenciación de genes candidatos han sido los más empleados de forma tradicional en la detección de mutaciones causales para enfermedades producidas por genes mayores. Sin embargo, tanto esta estrategia como el análisis de ligamiento se han visto desplazadas con la aparición del chip de SNPs, apareciendo el primer estudio en el año 2010 (Becker *et al.*, 2010) y habiéndose publicado siete más entre los años 2011 y 2015 (Momke *et al.*, 2011; Zhao *et al.*, 2011; Walker, 2012; Zhao *et al.*, 2012a; Zhao *et al.*, 2012b; Suarez-Vega *et al.*, 2013; Suarez-Vega *et al.*, 2015).



**Tabla 2.** Enfermedades de herencia mendeliana en el ganado ovino.

| Enfermedades de herencia mendeliana en el ganado ovino |                |                     |               |           |  |                                    |
|--|----------------|---------------------|---------------|-----------|--|------------------------------------|
| Enfermedad   | Gen            | Análisis ligamiento | Gen Candidato | Chip SNPs | Fenotipo   | Referencia                         |
| Acromatopsia   | <i>CNGA3</i>   |                     | X             |           | Ceguera diurna   | (Reicher <i>et al.</i> , 2010)     |
| Condrodisplasia (Spider Lamb Syndrome)                 | <i>FGFR3</i>   | X                   |               |           |   | (Cockett <i>et al.</i> , 1999)     |
| Condrodisplasia  | <i>SLC13A1</i> |                     |               | X         |   | (Zhao <i>et al.</i> , 2012b)       |
| Síndrome Ehlers-Danlos tipo VII (dermatoparaxis)       | <i>ADAMTS2</i> |                     | X             |           |    | (Zhou <i>et al.</i> , 2012)        |
| Epidermólisis bullosa juntural                         | <i>LAMC2</i>   |                     |               | X         |   | (Momke <i>et al.</i> , 2011)       |
| Gangliosidosis, GM1                                    | <i>GLB1</i>    |                     |               | X         | Disfunción neuromuscular y alteración del crecimiento                                | (Walker, 2012)                     |
| Gangliosidosis GM2, tipo I                             | <i>HEXA</i>    |                     | X             |           | Ataxia, ceguera  | (Torres <i>et al.</i> , 2010)      |
| Enfermedad de Gaucher tipo I                           | <i>GBA</i>     |                     | X             |           | Neuropatía, ictiosis, incapacidad de mantenerse en la estación                       | (Karageorgos <i>et al.</i> , 2011) |
| Enfermedad de almacenamiento de glicógeno V            | <i>PYGM</i>    |                     | X             |           | Mialgias y calambres musculares  | (Tan <i>et al.</i> , 1997)         |
| Hemofilia A  | <i>F8</i>      |                     | X             |           | Alteración en la coagulación de la sangre  | (Porada <i>et al.</i> , 2010)      |
| Raquitismo   | <i>DMP1</i>    |                     |               | X         |  | (Zhao <i>et al.</i> , 2011)        |
| Hipotricosis   | <i>HR</i>      |                     | X             |           | Pérdida progresiva de pelo   | (Finocchiaro <i>et al.</i> , 2003) |
| Microftalmia   | <i>PITX3</i>   |                     |               | X         |  | (Becker <i>et al.</i> , 2010)      |
| Síndrome de neurona motora inferior                    | <i>AGTPBP1</i> |                     |               | X         | Debilidad progresiva y tetraparesia  | (Zhao <i>et al.</i> , 2012a)       |
| Lipofuscinosis cerioide, neuronal 5                    | <i>CLN5</i>    | X                   |               |           | Retraso psicomotor progresivo, ceguera y muerte prematura                            | (Frugier <i>et al.</i> , 2008)     |
| Lipofuscinosis cerioide, neuronal 6                    | <i>CLN6</i>    | X                   |               |           | Retraso psicomotor progresivo, ceguera y muerte prematura                            | (Tammen <i>et al.</i> , 2006)      |

## Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino

|  |              |                |   |                                    |
|--|--------------|----------------|---|------------------------------------|
| Lipofuscinosis ceróide congénita, Enfermedad de Batten | <i>CTSD</i>  | X              | Retraso psicomotor progresivo, ceguera y muerte prematura   | (Tyynela <i>et al.</i> , 2001)     |
| Porfiria cutánea tarda                                 | <i>UROD</i>  | X              | Orina marrón, lesiones en piel por exposición a la luz solar  | (Nezamzadeh <i>et al.</i> , 2005)  |
| Síndrome Waardenburg, tipo 4A                          | <i>EDNRB</i> | X              |   | (Luhken <i>et al.</i> , 2012)      |
| Lisencefalia con hipoplasia cerebelar                  | <i>RELN</i>  | X              |   | (Suarez-Vega <i>et al.</i> , 2013) |
| Miotonía   | <i>CLCN1</i> | X              | Ante estímulos que en condiciones normales provocarían la huida, los animales se quedan rígidos y caen al suelo. Después de 6-25 s vuelven a moverse normalmente. | (Monteagudo <i>et al.</i> , 2015)  |
| Epidermólisis bullosa juntural                         | <i>ITGB4</i> | X<br>(RNA-Seq) |    | (Suarez-Vega <i>et al.</i> , 2015) |

### 2.4. Uso de la transcriptómica para el estudio de la lactación

#### 2.4.1. Herramientas utilizadas en el estudio del transcriptoma

El transcriptoma se define como el conjunto de transcritos expresados en una célula en un momento determinado. El estudio de los transcriptomas es importante ya que nos permite interpretar los elementos funcionales del genoma e identificar los genes que se están expresando en las distintas células y tejidos. Los objetivos principales de la transcriptómica incluyen la identificación de los distintos tipos de transcritos (mRNAs, RNAs no codificantes y RNAs pequeños), la determinación de la estructura transcripcional de los genes (extremos 5' y 3' y distintas isoformas) y la cuantificación de la expresión, identificando las posibles diferencias de expresión entre distintas condiciones fisiológicas (Wang *et al.*, 2009).

Las tecnologías tradicionalmente más utilizadas en el estudio de la expresión génica son los *microarrays* (Schena *et al.*, 1995); esta técnica se basa en la hibridación del cDNA marcado con fluorescencia frente a sondas que representan a los genes para los que queremos medir la expresión. Se trata de técnicas económicamente asequibles y

de alto rendimiento, siendo las más utilizadas en el análisis del perfil de transcriptoma completo de unas células o tejidos. Sin embargo, esta metodología también tiene inconvenientes ya que se necesita tener un conocimiento previo sobre la secuencia genómica, los altos niveles de ruido de fondo producidos debido a hibridaciones cruzadas, la incapacidad de la técnica para detectar transcritos nuevos y la cuantificación relativa que depende de la intensidad de hibridación (Morozova and Marra, 2008).

En contrapunto a los *microarrays*, surge la metodología de análisis en serie de la expresión génica (SAGE) que se basa en el principio de que un fragmento corto de cDNA (*tag*), de aproximadamente de 14-20 pb, contiene la suficiente información para identificar un transcrito determinado, siempre que sea aislado de una posición definida dentro del transcrito (Vilain and Vassart, 2004). La secuenciación, generalmente con la metodología *Sanger*, de *tags* de los transcritos presentes en la muestra permite la cuantificación de los niveles de expresión. Esta técnica presenta un elevado rendimiento y permite la cuantificación de todos los transcritos, además de permitir comparaciones más precisas que los *microarrays* entre muestras diferentes (Asmann *et al.*, 2009). Sin embargo, esta metodología también tiene sus inconvenientes, la secuenciación *Sanger* es cara, una proporción significativa de los transcritos no puede ser mapeada frente a la referencia y la detección de isoformas no es posible (Wang *et al.*, 2009).

La aparición de la técnica de secuenciación masiva paralela de RNA (RNA-Seq) revolucionó el análisis de transcriptomas. La técnica de RNA-Seq utiliza herramientas de secuenciación de segunda generación, siendo el primer método que permite la secuenciación de todo el transcriptoma con un elevado rendimiento y a un coste relativamente bajo, proporcionando un nuevo escenario en el mapeo y cuantificación de transcriptomas (Wang *et al.*, 2009). Esta técnica permite cuantificar todos los transcritos presentes en la muestra (en función de la redundancia de secuenciación) y no sólo los presentes en el *array*, como en el caso de los *microarrays* (Marioni *et al.*, 2008). Además, aunque las lecturas de las técnicas NGS son generalmente cortas (25-250 bp), el procesamiento en paralelo de millones de lecturas permite una cobertura alta del transcriptoma (Mardis, 2008), proporcionando información sobre la estructura

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

de los transcritos (existencia de fenómenos de *splicing*) y permitiendo la detección de variantes en regiones codificantes (Wickramasinghe *et al.*, 2014). Uno de los principales retos a los que se enfrenta la técnica de RNA-Seq está relacionado con el alto rendimiento y la gran cantidad de datos generados. Así, surgen desafíos desde el punto de vista bioinformático, que incluyen el desarrollo de métodos eficientes para el almacenamiento, recuperación y procesamiento de grandes cantidades de datos (Wang *et al.*, 2009).

### **2.4.2. Aplicaciones de la técnica RNA-Seq en el estudio de la lactación**

El conocimiento del perfil de expresión génica en la glándula mamaria durante la lactación tiene un especial interés ya que nos permite caracterizar a los genes implicados en la biología de la lactación, los cambios estructurales y metabólicos que se producen en la glándula mamaria durante este periodo y además, en el caso de la industria lechera, nos permite mejorar nuestros conocimientos en relación con la composición láctea y aquellos genes que pueden resultar interesantes desde el punto de vista productivo.

El transcriptoma mamario ha sido ampliamente estudiado en diversas especies mediante la utilización de *microarrays* de expresión (Lemkin *et al.*, 2000; Maningat *et al.*, 2009; Bionaz *et al.*, 2012; Shu *et al.*, 2012), sin embargo, la aparición de las técnicas de RNA-Seq nos han proporcionado una oportunidad única para cuantificar y anotar transcriptomas en un único experimento (Wang *et al.*, 2009).

En el caso del ganado vacuno, se han realizado diversos estudios sobre el perfil transcripcional de la glándula mamaria en lactación con la técnica de RNA-Seq. El primer estudio de RNA-Seq en el ganado vacuno fue realizado por Medrano *et al.* (2010), en él se realizaba un análisis comparativo entre el transcriptoma de las células somáticas de la leche y biopsias mamarias, demostrando la eficacia de esta técnica en el estudio global del perfil de expresión génica en tejidos de animales. En el año 2011 se realizó, por parte de Wickramasinghe *et al.* (2011), un análisis en el que se examinaba el perfil de expresión de genes que codificaban para enzimas relacionadas con la glicosilación en la glándula mamaria y se comparaban con el perfil de

oligosacáridos en la leche. En este estudio se identificaron genes candidatos que podrían ser utilizados en la modificación del contenido de oligosacáridos en leche. Más tarde Wickramasinghe *et al.* (2012) realizaron un estudio del transcriptoma de las células somáticas de la leche en diferentes momentos de la lactación, demostrando que en las distintas etapas de la lactación se expresan grupos de genes con funciones específicas de cada una de ellas. La técnica de RNA-Seq aplicada al estudio de la glándula mamaria también ha permitido la identificación de variantes (SNPs) (Canovas *et al.*, 2010). El estudio propuesto por Canovas *et al.* (2013) se basa en el análisis de la expresión génica y detección de variantes relacionadas con el contenido en citrato en la leche de vaca. La combinación de los resultados obtenidos con los conocimientos existentes de las rutas del citrato y la síntesis de ácidos grasos permitieron identificar marcadores significativos que podrían regular el contenido en citrato en la leche bovina (Canovas *et al.*, 2013). En el año 2014, se publicó un estudio en el que se comparaba, mediante la técnica de RNA-Seq, el transcriptoma mamario de vacas Holstein con contenidos extremos de grasa y proteína, demostrando que esta técnica puede ser útil en la identificación de genes candidatos relacionados con caracteres de composición láctea (Cui *et al.*, 2014). Un estudio de RNA-Seq publicado por Canovas *et al.* (2014) comparó cinco fuentes diferentes de obtención de RNA en la glándula mamaria (biopsia de tejido, células somáticas lácteas, glóbulos grasos de la leche, microdissección laser de células epiteliales mamarias y células epiteliales mamarias capturadas con anticuerpos), demostrando que las técnicas menos invasivas (RNA obtenido de células somáticas y glóbulos de grasa de la leche) son representativas del tejido mamario y pueden utilizarse en estudios de RNA-Seq en lugar de biopsias. La técnica de RNA-Seq también ha sido utilizada para la determinación de la composición y estructura de las poblaciones bacterianas de la leche en animales sanos y con mastitis (Medrano *et al.*, 2014).

La amplia variedad de estudios de expresión génica realizados en la glándula mamaria del ganado vacuno son el mejor ejemplo de la utilidad de la técnica de RNA-Seq en la identificación de genes clave en la determinación de caracteres complejos relacionados con la producción y la composición de la leche; así como, la detección de variantes génicas que pueden ser utilizadas en la mejora de la producción láctea, tanto

en la especie bovina como en otras especies productoras de leche, como la oveja o la cabra.

## **2.5. La producción láctea ovina en Castilla y León**

La producción mundial de leche de oveja es de unos 10 millones de toneladas anuales, siendo la cuarta en importancia a nivel mundial en el 2013 (FAOSTAT). La industria lechera ovina se concentra principalmente en Europa y los países de la cuenca mediterránea (Haenlein, 2001). En general, la leche de oveja es procesada para la elaboración de productos lácteos, principalmente queso. España se sitúa como el séptimo productor mundial de leche de oveja en el 2013 (FAOSTAT), siendo la comunidad de Castilla y León la que genera un 70% de esta producción láctea (352.501.000 litros en 2011) (MAGRAMA).

En el estudio del perfil de expresión génica de la glándula mamaria durante la lactación, presentado en esta Tesis Doctoral, se muestrearon ovejas de raza Churra y de raza Assaf. Las razas lecheras Assaf y Churra son las más utilizadas en la comunidad de Castilla y León, con un total de 77.896 y 36.054 ovejas en control lechero en 2014, respectivamente. La raza Churra es una raza autóctona española, caracterizada por su rusticidad. La raza Assaf, es una raza lechera especializada, creada como un cruce entre las razas Awassi (5/8) y Milschschaf (3/8). La raza Assaf se introdujo en España en el año 1977 y el número de animales se incrementó rápidamente debido a su alto potencial productivo y a su capacidad de adaptación a las condiciones ambientales de la región de Castilla y León (De la Fuente *et al.*, 2006). Ambas razas difieren en sus características productivas. La producción láctea de la raza Assaf es de una media de 400 kg de leche por lactación (normalizada a 150 días), mientras que la producción en la raza Churra es de 117 kg de leche por lactación (normalizada a 120 días). Sin embargo, la leche del ganado ovino de raza Churra tiene un mayor porcentaje de grasa (7,01 vs. 6,65) y proteína (5,79 vs. 5,40), por lo que el rendimiento quesero es mayor en la leche de oveja de raza Churra (MAGRAMA).



## 3. Metodología





En el presente apartado se describe, de forma general, los procedimientos bioinformáticos utilizados en el desarrollo de la presente Tesis Doctoral. En el capítulo de “Resultados” se detallan las metodologías específicas empleadas en el desarrollo de cada uno de los objetivos concretos.

### **3.1. Clonado posicional con el chip de SNPs de 50K ovino de los genes responsables de enfermedades de herencia mendeliana**

#### *3.1.1. Etapa 1: Detección de la región portadora de genes responsables de enfermedades de herencia mendeliana*

La detección de la región portadora de genes responsables de enfermedades de herencia mendeliana con el chip de SNPs de media densidad ovino (*OvineSNP50 BeadChip*) se basa en la realización de un análisis de asociación a nivel genómico (GWAS, del inglés *Genome-Wise Association Study*), siguiendo un esquema caso/control.

En primer lugar se realiza el genotipado de una serie de casos (animales afectados de un determinado fenotipo) y los controles (los animales con el fenotipo “salvaje”) con el *OvineSNP50 BeadChip* comercializado por *Illumina*. Como ya se ha descrito en la Introducción General, una de las ventajas de la utilización de esta metodología es el reducido número de casos necesarios para detectar de forma efectiva la región genómica implicada en la enfermedad. En el caso de las enfermedades analizadas en la presente Tesis doctoral, se genotiparon un total de 7 corderos afectados y 33 controles en el caso de la lisencefalia con hipoplasia cerebelar y 20 corderos afectados y 76 controles en el caso de la epidermólisis bullosa juntural.

El flujo general de trabajo una vez recibidos los datos de genotipado comienza con un control de calidad a dos niveles: primero, un control por animal y, posteriormente, un control de calidad por marcador. En la primera fase, se eliminaron todos los animales con un porcentaje de genotipos inferior al 90%. En la segunda se eliminaron del análisis todos los SNPs con una tasa de genotipado inferior o igual a 95% y frecuencia del alelo menos frecuente (MAF) inferior o igual a 0,01. En la presente Tesis Doctoral

### ***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

todo el flujo de análisis de los datos de genotipado se realizó con la herramienta bioinformática PLINK (Purcell et al., 2007).

Tras el control de calidad de los genotipos, se realiza el estudio de asociación a nivel genómico, que se basa en el análisis del desequilibrio de ligamiento de los marcadores analizados con el fenotipo en estudio. Al realizar el análisis estadístico de los genotipos obtendremos evidencia de asociación, en la región concreta del genoma, en la que las variantes alélicas estén sobre-representadas en los casos respecto a los controles.

#### ***3.1.2. Etapa 2: Mapeo fino y redefinición del fragmento portador del gen causal***

Con el fin de acotar la región de búsqueda de los genes responsables de enfermedades monogénicas, los GWAS se suelen combinar con otro tipo de análisis denominados de identidad por descendencia o “*homozygosity run*”. Este análisis se realiza con los genotipos obtenidos con el chip de SNPs en los animales enfermos con el objetivo de buscar regiones de homocigosis compartidas a lo largo de todo el genoma. El análisis se basa en la asunción de que los animales enfermos no comparten solo la mutación causal, sino una región de homocigosis más amplia debido al desequilibrio de ligamiento. Así, la región en la que se localice el mayor bloque de homocigosis compartido por todos los individuos afectados será la que más probablemente contenga la mutación responsable de la enfermedad. En la presente Tesis Doctoral, el análisis de identidad por descendencia, al igual que el GWAS, se realizó con el *software PLINK* (Purcell et al., 2007). Los análisis realizados nos permitirán acotar la región a unas pocas Mb en el genoma ovino.

Una vez identificada la región genómica portadora de la mutación causal, los genes candidatos posicionales localizados en esta región deben de ser evaluados en base a su función y a la homología de la enfermedad objeto de estudio con enfermedades descritas en otras especies de referencia, como la especie humana. El gen que se posicione como el mejor candidato funcional y posicional será el seleccionado para su secuenciación en el intento de identificar la mutación causal. Si no hubiese ningún candidato funcional entre los posibles candidatos posicionales se optaría por la secuenciación de la región de homocigosis completa.

### **3.2. Análisis de datos de NGS: flujo de trabajo para el análisis de variantes genéticas en el diagnóstico de enfermedades medelianas**

El objetivo de la aplicación diagnóstica de las NGS, tanto si hacemos resecuenciación de parte (*exome sequencing* o *target sequencing*) como de todo el genoma o transcriptoma, es la detección de las mutaciones reales que subyacen a las enfermedades mediante la secuenciación masiva paralela del material genético. El flujo general de trabajo para la detección de variantes una vez obtenidos los datos de secuenciación comienza con el control de calidad de nuestras muestras (*FastQC*; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y el alineamiento de las lecturas generadas frente al genoma de referencia. El alineamiento se puede realizar con diferentes *software* bioinformáticos, dos de los más utilizados en la detección de variantes son *BWA* (Li and Durbin, 2009), en experimentos de DNA-seq o *STAR* (Dobin *et al.*, 2013), si queremos alinear lecturas procedentes de experimentos de RNA-Seq (el algoritmo considera que existen fenómenos de *splicing*). Tras el alineamiento, se eliminan duplicados de PCR que pueden introducir sesgos en el proceso de detección de variantes; este paso se realiza con herramientas bioinformáticas de manejo de secuencias, como *Samtools* (Li *et al.*, 2009) o *Picard* (<http://picard.sourceforge.net>). El esquema de análisis continúa con la detección variantes: *GATK* (McKenna *et al.*, 2010) o *Samtools* (Li, 2011), son dos de los *software* más utilizados. Posteriormente se realiza la anotación funcional, una de las herramientas más utilizadas es el programa *Variant Effect Predictor (VEP)* (McLaren *et al.*, 2010), con el objetivo de obtener toda la información posible sobre la región genómica en la que se encuentran las mutaciones identificadas y su posible efecto funcional (codón *stop*, cambio de aminoácido, mutación sinónima, etc.). Una vez realizado estos pasos, el análisis concluye con la interpretación biológica e identificación de la mutación causal.

Existen cuatro estrategias diferentes en la detección de la mutación causal (Ku *et al.*, 2012). En la estrategia seguida de forma universal se siguen tres criterios principales: (1) eliminación de variantes comunes que ya han sido detectadas y descritas en la especie de estudio; (2) centrar el estudio en variantes deletéreas, como las mutaciones no sinónimas con cambio de sentido (*missense*), las que se producen codones de paro o terminación (*nonsense*), mutaciones en regiones de corte y empalme de exones

(regiones de *splicing*) o inserciones/deleciones en secuencias codificantes; (3) predecir los efectos funcionales de las variantes utilizando herramientas *in silico* como *SIFT* (Kumar *et al.*, 2009), *Polyphen-2* (Adzhubei *et al.*, 2010), o *ANNOVAR* (Wang *et al.*, 2010). Las otras estrategias se basan en el diseño del experimento, por ejemplo, si se han secuenciado individuos no relacionados con el caso o individuos emparentados con él. El filtrado de las variantes también dependerá de si existen datos de ligamiento o datos de regiones de homocigosis en las familias analizadas, ya que esto reducirá significativamente nuestros esfuerzos a la hora de localizar las mutaciones causales acotando la región de estudio. El modo de herencia del carácter mendeliano a estudiar también será otro criterio a tener en cuenta y determinará el filtrado de las variantes, por ejemplo, si el modo de herencia es recesivo nos centraremos en variantes en estado de homocigosis (Ku *et al.*, 2012).

### **3.3. Análisis de datos de RNA-Seq: flujo de trabajo para el análisis de expresión diferencial**

El flujo de trabajo que se sigue en el análisis de datos procedentes de RNA-Seq se describe en la Figura 2. Existen multitud de herramientas bioinformáticas que nos permitirían hacer el análisis de los datos de un experimento de RNA-Seq. A modo de resumen y para dar una idea de los diferentes pasos a seguir, se describen brevemente las etapas del análisis de expresión diferencial, que es uno de los análisis más comunes en los experimentos de RNA-Seq y que ha sido utilizado en esta Tesis Doctoral.

El primer paso en el análisis de expresión diferencial, una vez que se reciben los datos de secuenciación, es el control de calidad de los mismos. En este paso se comprueba la distribución general de las lecturas, su longitud, la calidad de las bases en cada posición de las lecturas, la calidad de la secuencia, el contenido de bases por secuencia, el contenido de Guanina-Citosina, el contenido de bases no leídas, la presencia de *kmeros*, los niveles de duplicados, de secuencias sobrerrepresentadas y la existencia de secuencias de adaptadores utilizados en la secuenciación. Una de las herramientas más utilizadas para probar la calidad de los datos brutos de secuenciación es *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Las lecturas de baja

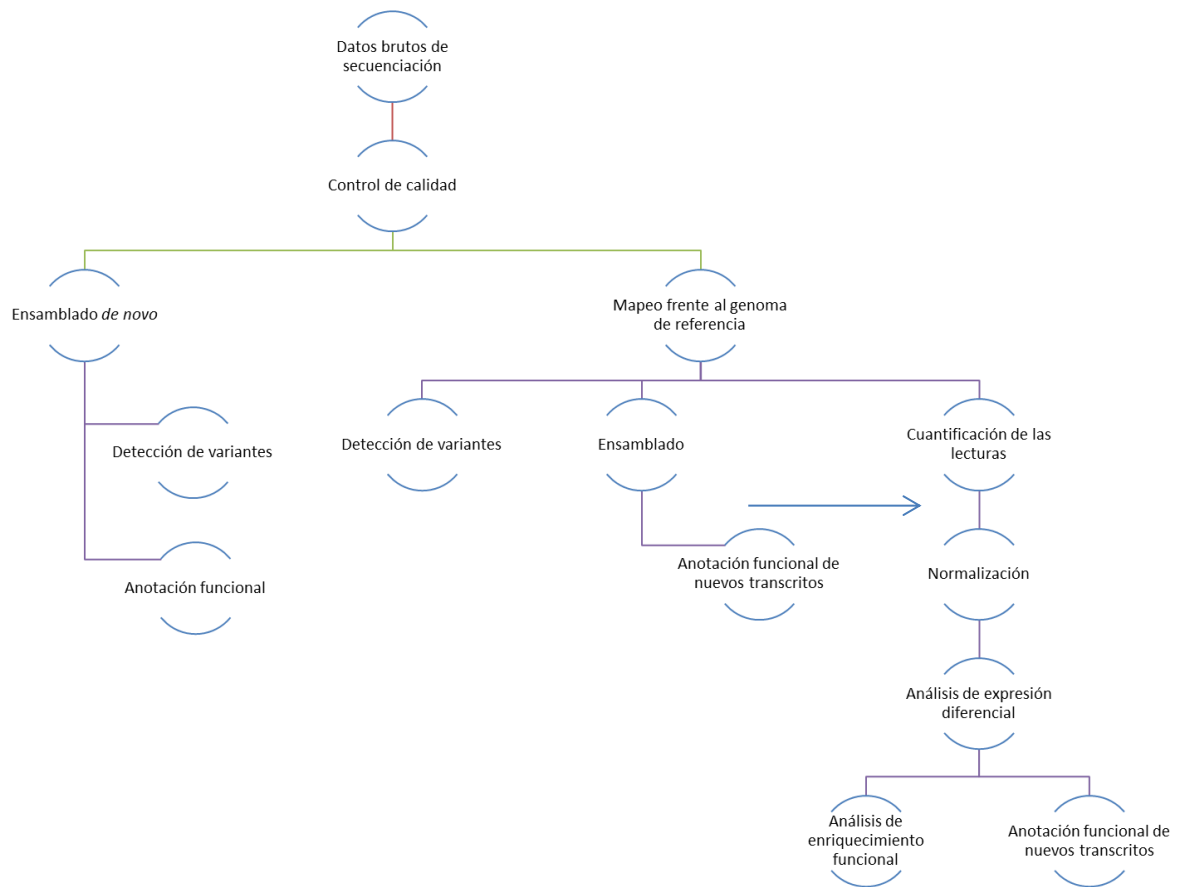
calidad son eliminadas, en su totalidad o en parte, en función de los criterios que fijemos cuando realizamos el control de calidad. Para realizar este paso existen numerosos programas, dos de ellos son *Fastx-Toolkit* ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) y *Trimmomatic* (Bolger *et al.*, 2014). A continuación se realiza el mapeo. En este paso millones de secuencias son alineadas en posiciones únicas del genoma de referencia. Existen programas especializados en el mapeo de lecturas procedentes de experimentos de RNA-Seq, que tienen en cuenta los fenómenos de corte y empalme que se producen en la transcripción del RNA en eucariotas. Dos de los programas más utilizados en el mapeo en experimentos de RNA-Seq son *TopHat2* (Kim *et al.*, 2013) y *STAR* (Dobin *et al.*, 2013). Después del mapeo se realiza un ensamblado de las secuencias, lo que nos permitirá identificar los diferentes transcritos presentes en nuestras muestras, creando una anotación de referencia que podremos utilizar en la cuantificación. El paquete de herramientas bioinformáticas *Cufflinks* (Trapnell *et al.*, 2010) es el más utilizado para realizar el ensamblado. El siguiente paso es la cuantificación de lecturas por gen, para ello podemos utilizar el archivo de anotación de referencia del genoma de la especie en estudio o el archivo de anotación creado a partir del ensamblado de nuestras lecturas. Para la cuantificación, podemos utilizar programas que nos den los datos normalizados, como es el caso de las herramientas de *Cufflinks* (Trapnell *et al.*, 2010) o podemos utilizar programas que proporcionen los recuentos brutos por gen como es el caso de *HTSeq-count* (Anders *et al.*, 2015). La normalización es un aspecto muy importante en el análisis ya que su objetivo es eliminar los efectos sistemáticos que ocurren en los datos para asegurar que los sesgos técnicos tengan un impacto mínimo en los resultados del análisis de expresión diferencial (Robinson and Oshlack, 2010). Uno de los procedimientos de normalización más utilizados es el método FPKM (*Fragments per Kilobase of Exon per Million Fragments Mapped*) que tiene en cuenta el tamaño de la genoteca y la longitud del gen analizado, ya que los genes más grandes tendrían un número más elevado de lecturas que alinean frente a ellos. Sin embargo, se ha demostrado que la corrección en base a la longitud del gen puede introducir sesgos en las varianzas por gen, sobre todo en los que están poco expresados (Oshlack and Wakefield, 2009). Existen otros dos métodos de normalización que realizan una corrección en función del tamaño de la genoteca, son el TMM (*Trimmed-Mean of M-values*) aplicado por *edgeR* (Robinson *et*

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

*al.*, 2010) y el aplicado por *DESeq2* (Love *et al.*, 2014). Ambos se basan en la hipótesis de que la mayoría de los genes no están diferencialmente expresados y proponen una normalización fundamentada en el cálculo de un factor escalar basado en un ratio de la media o mediana de las lecturas por gen en todas las muestras (Dillies *et al.*, 2013). Una vez normalizados los datos se realiza el análisis de expresión diferencial; tres de los programas más utilizados son *Cuffdiff2* del paquete *Cufflinks* (Trapnell *et al.*, 2012), *edgeR* (Robinson *et al.*, 2010) y *DESeq2* (Love *et al.*, 2014). El paso final consistiría en la interpretación biológica de los genes que han resultado diferencialmente expresados en nuestro experimento, para ello se pueden realizar análisis de enriquecimiento funcional utilizando herramientas como el *Ingenuity Pathway Analysis* (IPA, Qiagen), *WebGestalt* (Wang *et al.*, 2013) o *FatiGO* (Al-Shahrour *et al.*, 2007). También se podría realizar una anotación funcional de los genes/transcritos nuevos, que hemos identificado en el análisis pero que no están anotados, una de las herramientas más utilizadas para anotación funcional es *Blast2GO* (Conesa *et al.*, 2005).

Del mismo modo existen diferentes programas comerciales que permiten realizar el flujo de análisis completo de datos de RNA-Seq. Uno de más populares es el *CLC Bio Genomic Workbench* (Aahurs, Dinamarca).

Figura 2. Flujo de trabajo en el análisis de datos de RNA-Seq.







## 4. Resultados

**Suárez-Vega A., Gutiérrez-Gil B., Cuchillo-Ibáñez I., Sáez-Valero J., Pérez V., García-Gámez E., Benavides J. & Arranz J.J.** (2013) Identification of a 31-bp Deletion in the *RELN* Gene Causing Lissencephaly with Cerebellar Hypoplasia in Sheep. *PLoS ONE* 8(11): e81072. doi: 10.1371/journal.pone.0081072

**Suárez-Vega A., Gutiérrez-Gil B. & Arranz J.J.** (2013) Procedimiento y Kit de Diagnóstico de Lisencefalia con Hipoplasia Cerebelar en Ganado Ovino. Oficina Española de Patentes y Marcas: ES 2421207 B1.

**Suárez-Vega A., Gutiérrez-Gil B., Benavides J., Perez V., Tosser-Klopp G., Klopp C., Keenel, S.J. & Arranz, J.J.** (2015) Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. *PLoS ONE* 10(5): e0126416. doi: 10.1371/journal.pone.0126416.

**Suárez-Vega A., Gutiérrez-Gil B. & Arranz J.J.** (2015) Procedimiento y Kit de Diagnóstico de Epidermolísis Bullosa Juntural en Ganado Ovino. Solicitud de Patente Nacional en España Nº 201431635.

**Suárez-Vega, A., Gutiérrez-Gil, B., Esteban, C. & Arranz, J. J.** (2014) Initial RNA-seq analysis on the milk transcriptional profiling of two sheep breeds. *Proceedings, 10th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Vancouver, Canada. 17-22, Agosto, 2014.

**Suárez-Vega A., Gutiérrez-Gil B., Klopp C., Robert-Granie C., Tosser-Klopp G. & Arranz, J.J.**(2015) Characterization and Comparative Analysis of the Milk Transcriptome in Two Dairy Sheep Breeds using RNA Sequencing. *Scientific Reports* (reference number: SREP-15-21660). Under review.

**Suarez-Vega, A., Gutierrez-Gil, B. & Arranz, J.J.** (2015) Estudio del perfil transcripcional de genes implicados en el metabolismo de los ácidos grasos de la leche en dos razas ovinas. En XVI Jornadas sobre Producción Animal, Tomo II (ed. Asociación Interprofesional para el Desarrollo Agrario) 462-464 (INO Reproducciones S.A.).





**Identification of a 31-bp deletion in the *RELN* gene causing lissencephaly with cerebellar hypoplasia in sheep.**

Aroa Suárez-Vega<sup>1</sup>, Beatriz Gutiérrez-Gil<sup>1</sup>, Inmaculada Cuchillo- Ibáñez<sup>2,3</sup>, Javier Sáez-Valero<sup>2,3</sup>, Valentín Pérez<sup>4</sup>, Elsa García-Gómez<sup>1</sup>, Julio Benavides<sup>5</sup> and Juan José Arranz<sup>1</sup>.

<sup>1</sup> Dpto. Producción Animal, Universidad de León, 24071, León, Spain; <sup>2</sup>Instituto de Neurociencias de Alicante, Universidad Miguel Hernández-CSIC, Sant Joan d'Alacant; <sup>3</sup>Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Spain. <sup>4</sup> Dpto. de Sanidad Animal, Universidad de León, 24071, León, Spain. <sup>5</sup> Instituto de Ganadería de Montaña, Universidad de León CSIC.

PLoS ONE 2013, 8(11): e81072. doi: 10.1371/journal.pone.0081072.



# Identification of a 31-bp Deletion in the *RELN* Gene Causing Lissencephaly with Cerebellar Hypoplasia in Sheep

Aroa Suárez-Vega<sup>1</sup>, Beatriz Gutiérrez-Gil<sup>1</sup>, Inmaculada Cuchillo-Ibáñez<sup>2,3</sup>, Javier Sáez-Valero<sup>2,3</sup>, Valentín Pérez<sup>4,5</sup>, Elsa García-Gámez<sup>1</sup>, Julio Benavides<sup>5</sup>, Juan Jose Arranz<sup>1\*</sup>

**1** Dpto. Producción Animal, Universidad de León, León, León, Spain, **2** Instituto de Neurociencias de Alicante, Universidad Miguel Hernández, CSIC, Sant Joan d'Alacant, Alicante, Spain, **3** Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Universidad Miguel Hernández, Sant Joan d'Alacant, Alicante, Spain, **4** Dpto. de Sanidad Animal, Universidad de León, León, León, Spain, **5** Instituto de Ganadería de Montaña, Universidad de León CSIC, Grulleros, León, Spain

## Abstract

Lissencephaly is an inherited developmental disorder in which neuronal migration is impaired. A type of lissencephaly associated with cerebellar hypoplasia (LCH) was diagnosed in a commercial flock of Spanish Churra sheep. The genotyping of 7 affected animals and 33 controls with the OvineSNP50 BeadChip enabled the localization of the causative mutation for ovine LCH to a 4.8-Mb interval on sheep chromosome 4 using genome-wide association and homozygosity mapping. The *RELN* gene, which is located within this interval, was considered a strong positional and functional candidate because it plays critical roles in neuronal migration and layer formation. By performing a sequencing analysis of this gene's specific mRNA in a control lamb, we obtained the complete CDS of the ovine *RELN* gene. The cDNA sequence from an LCH-affected lamb revealed a deletion of 31 bp (c.5410\_5440del) in predicted exon 36 of *RELN*, resulting in a premature termination codon. A functional analysis of this mutation revealed decreased levels of *RELN* mRNA and a lack of reelin protein in the brain cortex and blood of affected lambs. This mutation showed a complete concordance with the Mendelian recessive pattern of inheritance observed for the disease. The identification of the causal mutation of LCH in Churra sheep will facilitate the implementation of gene-assisted selection to detect heterozygous mutants, which will help breeders avoid at-risk matings in their flocks. Moreover, the identification of this naturally occurring *RELN* mutation provides an opportunity to use Churra sheep as a genetically characterized large animal model for the study of reelin functions in the developing and mature brain.

**Citation:** Suárez-Vega A, Gutiérrez-Gil B, Cuchillo-Ibáñez I, Sáez-Valero J, Pérez V, et al. (2013) Identification of a 31-bp Deletion in the *RELN* Gene Causing Lissencephaly with Cerebellar Hypoplasia in Sheep. PLoS ONE 8(11): e81072. doi:10.1371/journal.pone.0081072

**Editor:** Glenn Francis Browning, The University of Melbourne, United States of America

**Received:** July 10, 2013; **Accepted:** October 9, 2013; **Published:** November 19, 2013

**Copyright:** © 2013 Suarez-Vega et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the Spanish Ministry of Science (Project AGL2009-07000) and by the European Commission 3SR Project (Sustainable Solutions for Small Ruminants; <http://www.3srbreeding.eu>) (to JJA). Fondo de Investigaciones Sanitarias (PS09/00684; PI12/00593), Fundación Ramón Areces, and CIBERNED, Instituto de Salud Carlos III from Spain (to JSV). Aroa Suarez-Vega is funded by an FPU contract from the Spanish Ministry of Education. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

\* E-mail: [jjarrs@unileon.es](mailto:jjarrs@unileon.es)

## Introduction

Lissencephaly (LIS), which literally means “smooth brain”, refers to a group of rare malformations that share a common feature: absent or abnormal brain convolutions caused by the aberrant migration of postmitotic neurons to the developing cortex. Although different forms of LIS have been described, there is still no final consensus on their classification. The classification system proposed by Jissendi-Tchofo et al. [1] divides this disease into four different groups: classic lissencephaly (cLIS), variant lissencephaly (vLIS), cobblestone complex, and related muscular dystrophy syndrome.

Lissencephalies are a genetically heterogeneous group of disorders. There are seven different human LIS-related phenotypes described in the Online Mendelian Inheritance in Man (OMIM) database (#607432, #300067, #300215, #257320, #611603, #614019 and #615191). These phenotypes have been linked to different aberrant proteins related to the cytoskeleton of neural cells (Platelet-activating factor acetylhydrolase IB subunit alpha (PAFAH1B1 or LIS1), Doublecortin (DCX), Tubulin alpha-1A (TUBA1A)), signaling molecules (Reelin (RELN), Very-low-density lipoprotein receptor (VLDLR)), molecules that modulate stop signals for migrating neurons (Protein-O-mannosyltransferase 1 (POMT1), Protein O-linked-mannose beta-1,2-N-

acetylglucosaminyltransferase 1 (POMGnT1), Fukutin (FKTN) and other factors shown to modulate neuronal migration (Aristaless related homeobox (ARX), Laminin alpha 1 (LAMA1)) [2].

Lissencephaly with cerebellar hypoplasia (LCH) is a type of LIS included within the vLIS group [1]. In this form of LIS, the thickened cortex is associated with significant cerebellar underdevelopment. Six subtypes of LCH have been described [3]. Several genes that are involved in the gestational migration of neurons have been linked to human LCH: *PAFAH1B1* [3], *DCX* [3], *RELN* [4], *VLDLR* [1,5] and *TUBA1A* [6].

Although rarely observed clinically, inherited forms of LCH are well known in humans and mouse mutant models like the *reeler* phenotype [7] due to its importance in understanding brain development. However, beyond its known forms in rodents, LCH is rarely described in veterinary medicine. Two litters of Wire Fox Terriers and Irish Setters [8] and one two-year-old cat [9] represent the only cases of LCH described in domestic animals, to our knowledge.

Between 2004 and 2012, an inherited form of LCH was identified in a commercial flock of Spanish Churra sheep. The affected lambs exhibited severe ataxia, were unable to stand by themselves and died several days after birth. Pathological examination showed alterations in the brain characterized by agyria, pachygyria and cerebellar hypoplasia. The segregation of this disease in the affected pedigrees was consistent with a recessive mode of Mendelian inheritance [10].

Recently, high-density single-nucleotide polymorphism (SNP) arrays have provided an opportunity to explore the genomes of livestock species to identify genes and mutations that cause inherited defects [11]. In sheep, the Illumina OvineSNP50 BeadChip has proven to be a useful tool for the identification of causal mutations underlying the genetic control of diseases with Mendelian inheritance patterns [12-14]. Using this genomic tool, we performed genome-wide association and homozygosity mapping analyses to map the ovine locus associated with the form of LCH identified in Churra sheep. These analyses allowed the identification of a strong positional and functional candidate gene, which was subjected to later analyses with the aim of identifying the causal mutation underlying the studied developmental malformation.

Hence, the present study identified the causal mutation for LCH in sheep, which will allow the direct implementation of gene-assisted selection into breeding practices by enabling the detection of phenotypically normal carriers. By deciphering the genetic basis of this ovine disease, the current work also provides a potential large animal model for human LCH, the study of brain development and the development of possible gene therapy treatments.

## Materials and Methods

### Ethics statement

Blood and tissue samples were collected from rams, ewes and lambs by qualified veterinarians following standard procedures and conducted under license issued in accordance with European Union legislation (European Community Directive, 86/609/EC and Directive 2010/63/EU of the

European Parliament and of the Council). All animals were managed in accordance with the guidelines for the accommodation and care of animals.

The DNA samples used in this study were extracted from blood leucocytes. For this purpose 4 mL of blood was obtained by jugular venepuncture.

Tissues were collected immediately after euthanasia (performed by a qualified veterinarian with an intravenous injection of veterinary euthanasia drug (T-61, Intervet)). As the samples were from a commercial flock that underwent veterinary examination, we were in a special situation in veterinary medicine and there was no “animal experiment” according to the legal definitions in Spain (Animal care legislation “Ley 32/2007”). According to the Ethics Commission of the University of Leon, formal ethical approval is not required under these circumstances.

### Animals

Several LCH-affected animals were found in a commercial flock of Spanish Churra sheep belonging to the Spanish Churra sheep breeders' association (ANCHE). For many years, the breeding strategy for the flock has been based on the use of rams from within this flock. Although a larger number of possible LCH cases had been detected in the flock, we examined seven LCH-affected animals, all of which exhibited the same clinical features at birth. All these animals could not stand by themselves and showed severe ataxia and muscular hypertonia. The affected lambs had difficulty suckling from their mothers and died a few days after birth. The post-mortem examination showed agyria with only a few rudimentary sulci and gyri, as well as marked cerebellar hypoplasia. The microscopically normal layering of the cerebral cortex was disorganized, and immunohistochemical staining of the neurofilaments revealed a three-layered cortex instead of the six layers that appear in normal brains. All these findings were consistent with LCH [10].

### Samples and pedigree information

Blood samples were collected from 63 animals from the LCH-affected flock. Seven of these animals were affected lambs. DNA was extracted from blood using the salting out procedure [15]. To confirm the familial relationships recorded in the flock register, all the samples were analyzed for a set of 19 microsatellite markers in a single multiplex PCR reaction [16]. To construct the pedigree of the sampled animals information about the ram parents, grandparents and great-grandparents was obtained from the ANCHE database.

The pedigree was constructed to confirm the monogenic autosomic mode of inheritance of this ovine disease reported [10] and to identify the common founder responsible of the establishment of the disease in this flock. For this purpose, we used CraneFoot software [17].

### Mapping the causative gene for lissencephaly

A total of 40 DNA samples were genotyped using the Illumina OvineSNP50 BeadChip. These samples included 20 unrelated, healthy Churra individuals from different flocks of the Churra Selection Nucleus and 20 animals from the affected

flock. Seven of the sheep from the affected flock were LCH lambs, and the rest were related to them (sires, dams, siblings or half-siblings of the affected lambs). Raw data will be shared upon request.

The results were analyzed using PLINK software [18]. Firstly, a quality control procedure was performed to eliminate animals with call rates lower than 0.95 and SNPs with genotyping rate values lower than 0.05 and minor allele frequencies lower than 0.01. After the frequency and genotype pruning, 47,864 SNPs were considered in the subsequent analysis. The `--assoc` option of the PLINK software was used to perform a case-control genome-wide association (GWA) analysis. The empirical p-values were corrected by performing a permutation procedure implemented in PLINK with 100,000 permutations.

Finally, an analysis of runs of homozygosity was carried out. The LCH cases were filtered to identify allele-sharing regions. The SNPs flanking the consensus region according to this analysis were used to locate this region in the Ovine Genome Assembly v3.1 browser (<http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv3.1/>). Genes located within the homozygosity block were evaluated as putative positional candidate genes. A second analysis of runs of homozygosity was performed on only 10 unrelated healthy Churra sheep to confirm that these control individuals did not share any common homozygous regions with the affected lambs.

### Isolation of the complete coding sequence of the ovine *RELN* gene and mutation scanning

Because of its association with human LCH [4], the *RELN* gene, located within the homozygosity block (4.8 Mb), was identified as the most promising candidate. Due to the large size of the *RELN* gene, which encompasses 450 kb distributed across 65 exons in the human genome [19], we first examined the *RELN* coding sequence. Brain tissues from a control lamb and an affected lamb were collected for RNA extraction. The use of the animals was in compliance with the guidelines approved by the University of Leon Ethics Commission. The brain tissue samples were harvested immediately after euthanasia and were preserved in RNA Stabilization Reagent (RNAlater, Ambion). Slices of up to 500 mg of brain tissue were processed with the RNeasy Lipid Tissue Midi Kit (Qiagen) to extract total RNA.

For the primer design, due to the absence of a published RNA sequence for the ovine *RELN* gene, we first constructed a virtual mRNA sequence based on exons obtained by a BLAST comparison of the GenBank human (NM\_005045) and mouse (NM\_011261) mRNA sequences against the genomic *RELN* sequence obtained from the ovine genome assembly v3.1. For this procedure, we used the web-based Spidey software, available at <http://www.ncbi.nlm.nih.gov/spidey/>.

The virtual mRNA sequence thus generated was used to design 32 primer pairs to cover all the *RELN* coding sequence, using the Primer3 software [20]. The information from the human and mouse sequences was considered to avoid designing primers in regions with significant differences across species.

One-step RT-PCR was performed to amplify aliquots of 60–80 ng of RNA using the Qiagen OneStep RT-PCR Kit. The

sequences and annealing temperatures of all the primer pairs are provided in Table S1. Finally, the sequence information generated was assembled and analyzed for polymorphisms.

After the mRNA sequence was established, the predicted protein sequence was compared with the human and mouse *RELN* protein sequences using the align option of the web-based UniProt software (<http://www.uniprot.org/>). The effects on protein sequence and secondary structures of the mutations that had been identified by the sequencing analysis were predicted using the SWISS-MODEL software [21–23].

### Confirmation and genotyping of the causal mutation

After identifying the putative causal mutation in the mRNA sequence, we confirmed its direct association with the studied phenotype by analyzing additional samples at the genomic DNA level. DNA was extracted from blood samples collected from six cases, six controls and six mothers of affected animals.

A pair of primers (*RELNovine\_ex36up*: 5'-TTGCCTTCTCCGGTTTAATG and *RELNovine\_ex36dn*: 5'-AGGGATTTGTGATGCTGGAC) were designed to amplify a 498-bp fragment of the gene sequence containing the 31-bp deletion that was identified as the possible causal mutation. The amplicons were purified by ExoSAP-IT (UBS Corporation) treatment and were dideoxy-sequenced in both directions with the Big Dye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems) with the same primers used for fragment amplification. The sequence data were analyzed with SeqScape v2.5 software (Applied Biosystems).

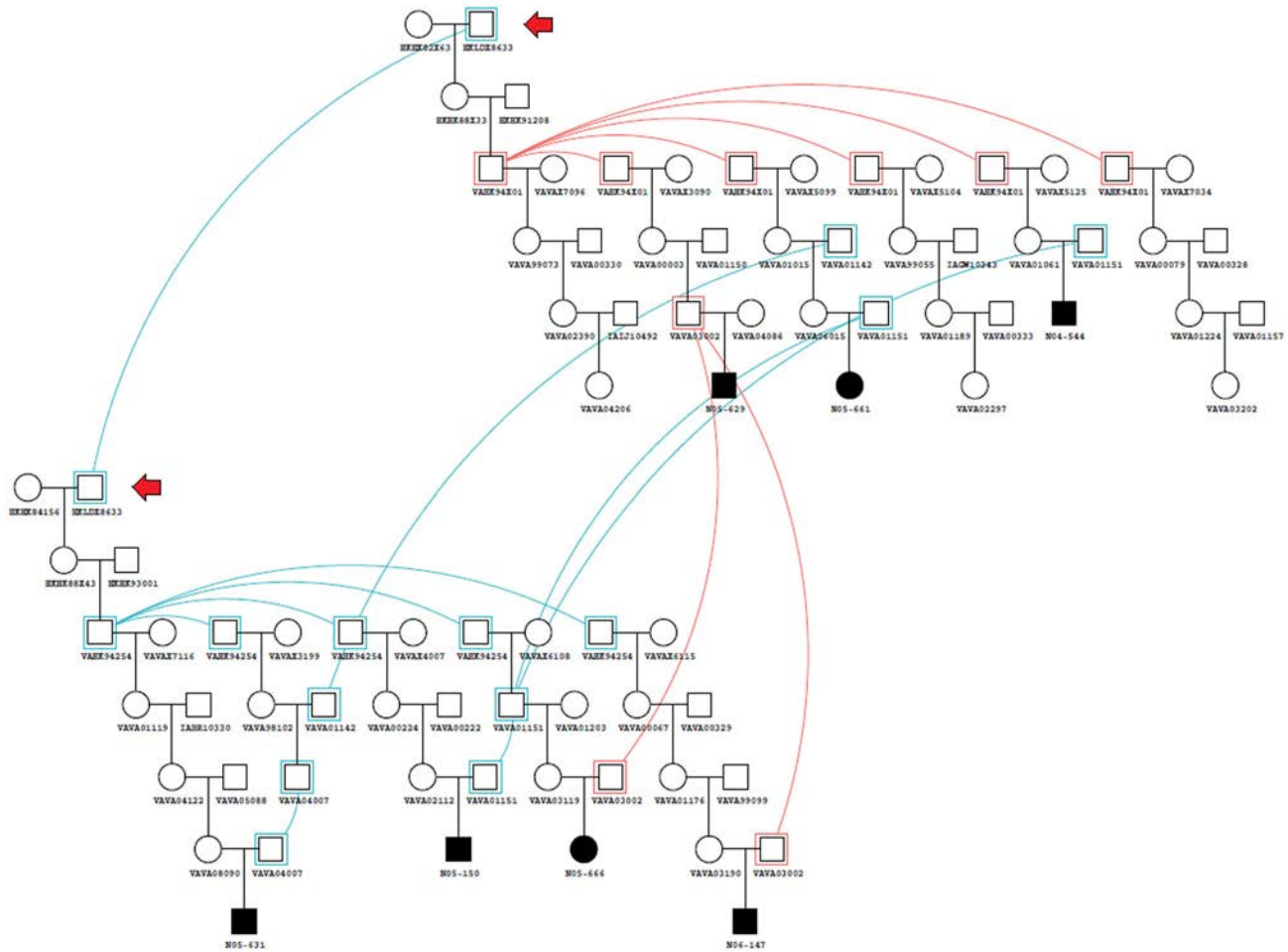
### Functional analysis of the 31-bp deletion: quantification of *RELN* mRNA by qRT-PCR

To quantify the levels of *RELN* mRNA in affected lambs and to compare them with unaffected animals, we performed two qRT-PCR amplifications in six LCH cases and six controls. The qRT-PCR assay was designed to amplify two different regions of the ovine *RELN* gene: a shared region which included 139 bp of the third reelin repeat, and the mutated region which include the four reelin-specific repeat encompassing the deletion.

Total RNA was isolated from the brains of six controls and six cases using TRIzol® Reagent and the PureLink™ RNA Mini Kit (Invitrogen) according to the manufacturer's instructions. The high capacity cDNA Reverse Transcription kit (Applied Biosystems) was used to synthesize DNA according to the manufacturer's instructions. Quantitative reverse transcription polymerase chain reaction was performed using the StepOne™ Real-Time PCR System (Applied Biosystems) with Power SYBR® Green PCR Master Mix according to the manufacturer's instructions (see Table S2 for the primer sequences). The *RELN* gene transcript levels were calculated using the relative standard curve method normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

### Functional analysis of the 31-bp deletion: Western blot analysis

A Western blot was performed to confirm that the 31-bp deletion identified in the affected individuals affected reelin



**Figure 1. Pedigree displaying LCH in Churra sheep.** The arrow indicates the ram identified as the common founder animal for the LCH disease. The lines indicate the inbreeding lops. Cases are indicated as solid symbols.

doi: 10.1371/journal.pone.0081072.g001

expression. We analyzed the protein in brains and sera from affected and unaffected animals. Samples (0.1 g) of sheep frontal cortex were homogenized (10% w/v) in 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% Triton X-100 and 0.5% Nonidet P-40 containing a cocktail of protease inhibitors [24]. The homogenates were sonicated and centrifuged at 20,000×g for 20 min at 4°C, and the supernatants collected and frozen at -80°C. Sheep brain extracts (30 μg) and serum samples (0.8 μl) were incubated with 6× SDS-PAGE sample buffer at 98°C for 3 min and then resolved using 6% SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The proteins were blotted onto nitrocellulose membranes and incubated with the monoclonal mouse anti-Reelin 142 antibody (1:500 dilution, Merck Millipore) and with a secondary antibody (1:4000 dilution, horseradish peroxidase (HRP)-conjugated anti-mouse IgG, (Sigma Chemicals)). The signals were visualized with the HRP chemiluminescent substrate Luminata™ Forte (Merck Millipore) in a Luminescent Image Analyzer LAS-1000 Plus (Fujifilm), and analyzed using the Science Lab Image Gauge v3.0 software (Fujifilm).

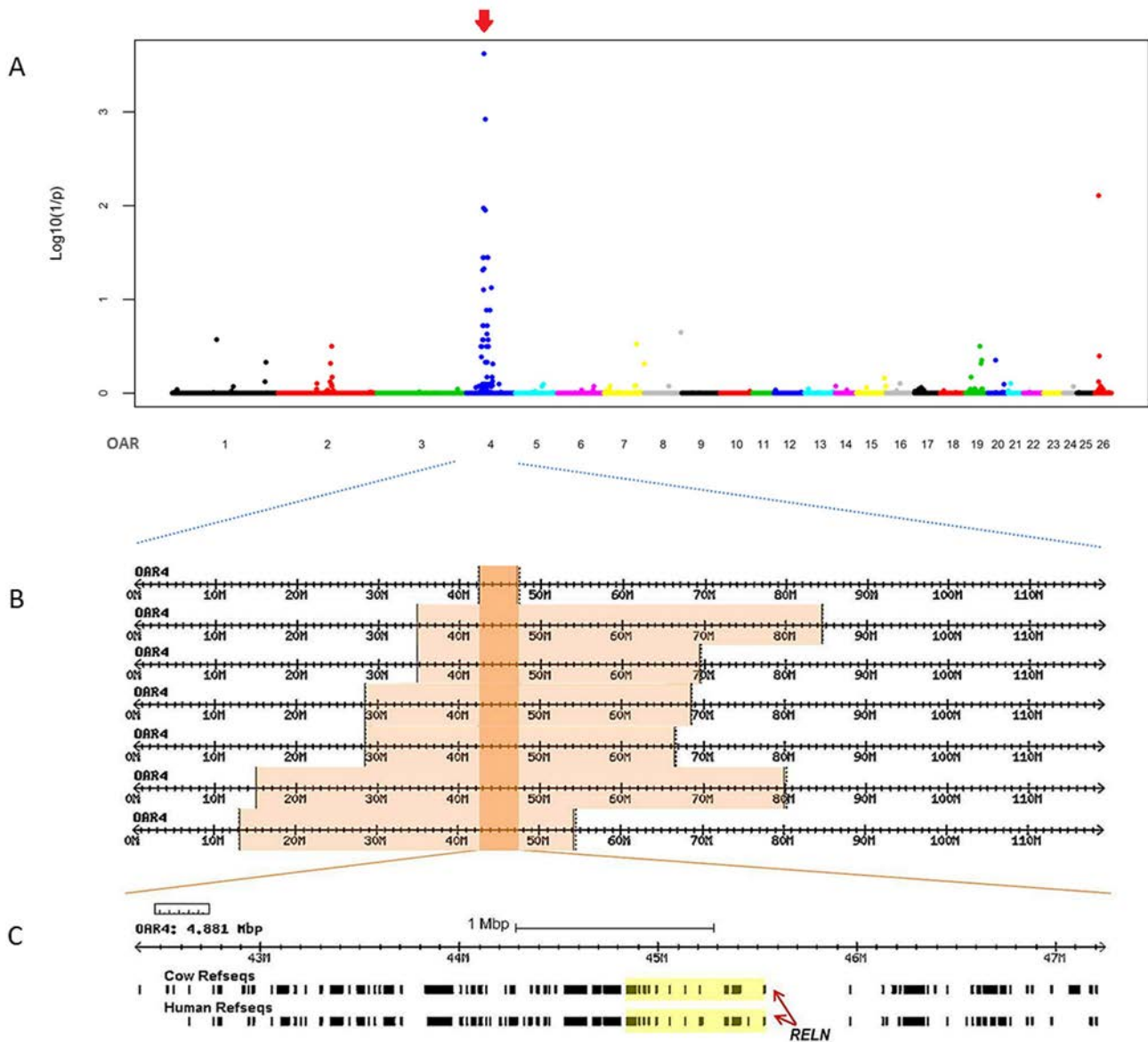
## Results

### Pedigree analysis

The pedigree of the affected individuals (Figure 1) was constructed to confirm the mode of inheritance previously described [10] and to search for the common founder of LCH in this flock. The breeding history of the affected families was consistent with a monogenic autosomal recessive pattern of inheritance. The parents and half-siblings of the affected animals did not show clinical signs of the disease. The parents of the affected animals were classified as obligate carriers.

An analysis of the pedigree data revealed that the LCH-affected individuals were descended from a single ram founder (HKLD8633) via either the paternal or maternal line (Figure 1). This male was born in 1986, and the disease was first recognized in 2004, four generations later.





**Figure 2. GWA and homozygosity mapping analyses for ovine LCH.** (A) Manhattan plot resulting from the case-control association analysis performed with 7 affected animals and 33 controls. The X-axis shows the positions of the genome analyzed across the 26 ovine autosomes, whereas the Y-axis represents the  $-\log_{10}$  P-values obtained after 100,000 permutations using PLINK. Alternating colors mark the limits between autosomes. The arrow indicates the most significant association identified, which was located on OAR4 (at 42,810,217 bp position). (B) Homozygosity mapping of the LCH mutation. The analysis of SNP genotypes from affected lambs indicated that they all shared an extended overlapping homozygous region on OAR 4 (indicated by orange blocks). The common haplotype block, where the causative mutation is located, expands between 42.369 and 47.251 Mb on OAR4 (indicated by the red box). (C) Gene content of the 4.8-Mb homozygosity block interval shared by the affected individuals based on the comparative genomic map with the orthologous regions in human and cow and including the *RELN* gene.

doi: 10.1371/journal.pone.0081072.g002

### Genome-wide association and runs of homozygosity studies

The GWA study revealed that the strongest association with the LCH phenotype localized to sheep chromosome 4 (OAR4) (Figure 2). The SNP showing the most highly significant

association was OAR4\_45088426, which is located at position 42,810,217 bp in the OARv3.1 ovine genome sequence. After 100,000 permutations were performed, this SNP showed a genome-wide corrected P-value of  $2.4 \times 10^{-4}$ .

The homozygosity mapping approach that was performed later to narrow the region containing the LCH mutation was chosen on the assumption that the affected lambs were identical by descent (IBD) for the causative mutation and the flanking chromosomal segments. The seven genotyped cases showed a single, large homozygous consensus region on OAR4 that contained 91 SNP markers corresponding to a 4.8-Mb interval from 42.369–47.251 Mb (Figure 2). Analyzing the data from the 10 unrelated healthy Churra controls did not reveal any homozygous shared regions across the genome that were greater than 300 Kb. Because of the limited annotation of the ovine genome assembly (v 3.1), we inferred the gene annotation of the mapped interval from the corresponding bovine and human orthologous intervals. The ovine LCH interval corresponds to segments on bovine chromosome (BTA) 4 and human chromosome (HSA) 7. A careful inspection of the BTA4 and HSA7 genes and database searches for their presumed functions indicated that the gene that encodes the reelin protein was the most promising functional candidate gene. *RELN* is located within the critical identified intervals: at 44.6 Mb on OAR4, at 44.8 Mb on BTA4 and at 103.1 Mb on HSA7. The reelin protein plays critical roles in neuronal migration and layer formation [25] and has been associated with the human LIS forms LIS2 and Norman-Roberts syndrome [4].

### Complete coding sequence of the ovine *RELN* gene

Because no reference sequence of the ovine *RELN* gene was initially available, the sequencing analysis of the specific mRNA of this gene in a control lamb allowed us to obtain the complete cDNA of the ovine *RELN* gene (GenBank acc. no. KC590614).

The BLAST comparison of the sequenced cDNA and the genomic sequence obtained from the ovine genome assembly v3.1 using the Spidey program showed an overall identity of 99% with 95% of mRNA coverage. The comparison between the cDNA and genomic DNA sequences allowed us to predict the structure of the coding sequence at the genomic level, resulting in 64 exons.

The sheep *RELN* cDNA sequence encodes a predicted protein of 3460 amino acids. The UniProt software (<http://www.uniprot.org/>) was used to compare the predicted amino acid sequence from the ovine *RELN* gene with several available sequences, including human (P78509), mouse (Q60841) and rat (P58751) (Figure S1). The ovine *RELN* protein sequence showed 97% homology with the human protein and 95% homology with the mouse and rat sequences, suggesting high conservation of this protein across species.

### DNA variant scanning and identification of the causal mutation

To identify the causal mutation of the studied disease, the entire *RELN* coding sequence from an affected lamb was sequenced. We amplified 9986 bp, corresponding to 96.21% of the coding region previously amplified in the control sample.

The sequence comparison between the control and affected samples revealed the presence of six SNPs and one 31-bp deletion (Table 1). For the two SNPs that produced amino acid

**Table 1.** DNA variants found by cDNA sequencing of the ovine *RELN* gene in a case and a control.

| Mutation ID    | Control | Case | Reference* | Protein level |
|----------------|---------|------|------------|---------------|
| c.44ta         | t/a     | t/a  | t          | p.Leu15Gln    |
| c.2676ag       | a/a     | g/g  | a          | p.=           |
| c.5410_5440del | -       | del  | -          | p. 0          |
| c.5862ct       | c/t     | c/c  | c          | p.=           |
| c8010gc        | g/c     | g/g  | g          | p.Met2670Ile  |
| c8754tc        | t/c     | t/t  | t          | p.=           |
| c.9153ag       | a/g     | a/a  | a          | p.=           |

The effect of each DNA variant on the protein sequence is indicated in the last column.

\* Reference based on the nucleotide found for this position in the Ovine Genome Assembly v3.1.

doi: 10.1371/journal.pone.0081072.t001

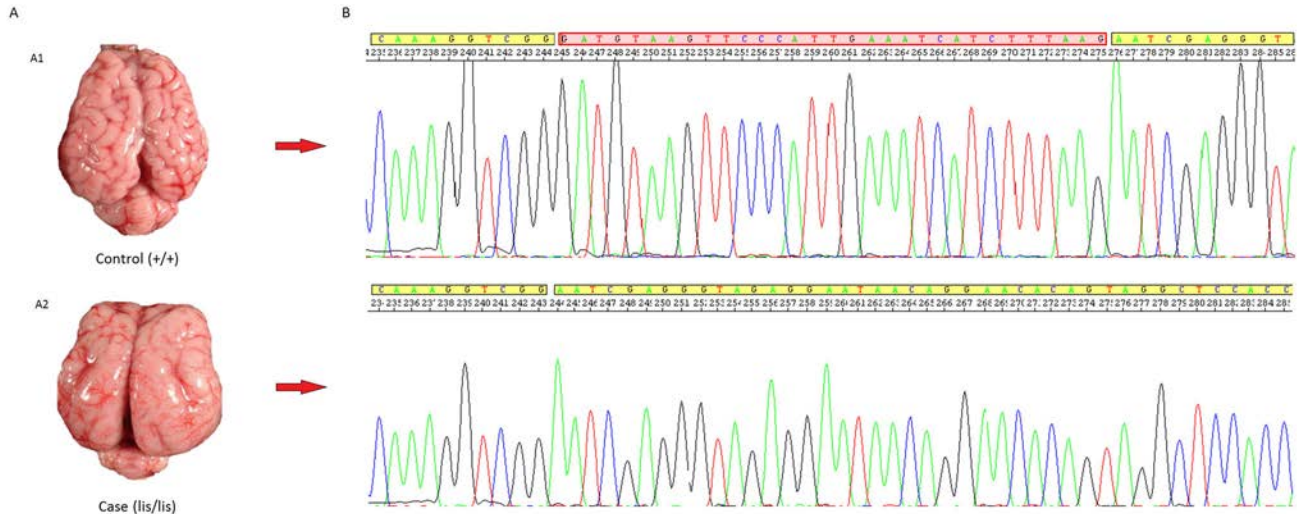
changes, the LCH-affected lamb shared the heterozygous state with the normal lamb (p.Leu15Gln) or was homozygous for the wild-type allele exhibited by 8 normal animals and the reference sequence of the ovine assembly at this position (p.Met2670Ile). In contrast, the identified 31-bp deletion (c.5410\_5440del) was present in a homozygous form in the LCH lamb (c.5410\_5440del) but not in the control animal. Based on the genomic structure obtained by a comparison of the cDNA and the gDNA, this mutation occurred within the sequence of predicted exon 36 of the ovine *RELN* gene. This mutation was predicted to cause a shift in the open reading frame of the *RELN* coding sequence and create a stop codon at position 1817. This mutation would lead to a truncated protein of 1817 amino acids (1803 amino acids of normal reelin followed by 14 missense amino acids and a premature termination codon) vs. the normal 3460-amino-acid-long *RELN* protein. Due to its highly disruptive effect on the composition and structure of the resulting protein, this 31-bp deletion was identified as possibly directly responsible for the disease under study.

### Confirmation and genotyping of the causal mutation

To confirm a direct association between the c.5410\_5440del allelic variant and the LCH phenotype, we searched for a correspondence between the genotype of this deletion and the gDNA sequences of affected and carrier animals. Using DNA samples, we performed PCR amplifications of putative exon 36 for six cases, six unrelated controls and six mothers of affected lambs. The resulting amplicon was sequenced and analyzed with the SeqScape v2.5 software (Applied Biosystems). All the samples from the LCH-affected animals were homozygous for the identified 31-bp deletion (c.5410\_5440del/c.5410\_5440del or Lis/Lis); the mothers were heterozygous for this mutation (Lis/+); and the controls from different flocks were homozygous (+/+) for the wild-type allele (Figure 3).

### Functional analysis of the 31-bp deletion

The levels of *RELN* mRNA in the affected lambs were markedly reduced, with means of 0.08 for the shared region and 0.04 for the mutated region compared with the mean



**Figure 3. Identification of the causative mutation for LCH in sheep.** (A) LCH phenotype in Spanish Churra sheep. (A1) Normal newborn lamb brain. (A2) LCH-affected lamb brain. (B) Electropherograms including the mutated region at exon 36 of *RELN* in the normal and LCH-affected lambs. The bases within the red square are included in the 31-bp deletion (c.5410\_5440del) detected in the LCH-affected lambs.

doi: 10.1371/journal.pone.0081072.g003

values observed for the control animals (1.52 and 0.40 for the shared and mutated fragments, respectively) (Figure 4A). In both cases, a one-tailed Student's t-test revealed significant differences in the *RELN* mRNA expression level between the cases and controls ( $P < 0.001$ ). We also analyzed the protein levels of reelin in the cortex and serum using Western blots (Figure 4B). Three typical reelin-immunoreactive bands of 420 kDa, 310 kDa and 180 kDa with different densities corresponding to the full-length and truncated N-terminal fragments [24] were observed in both the cortex and the serum of healthy lambs. The homozygous Lis/Lis animals showed a lack of reelin in both, tissue and fluids.

## Discussion

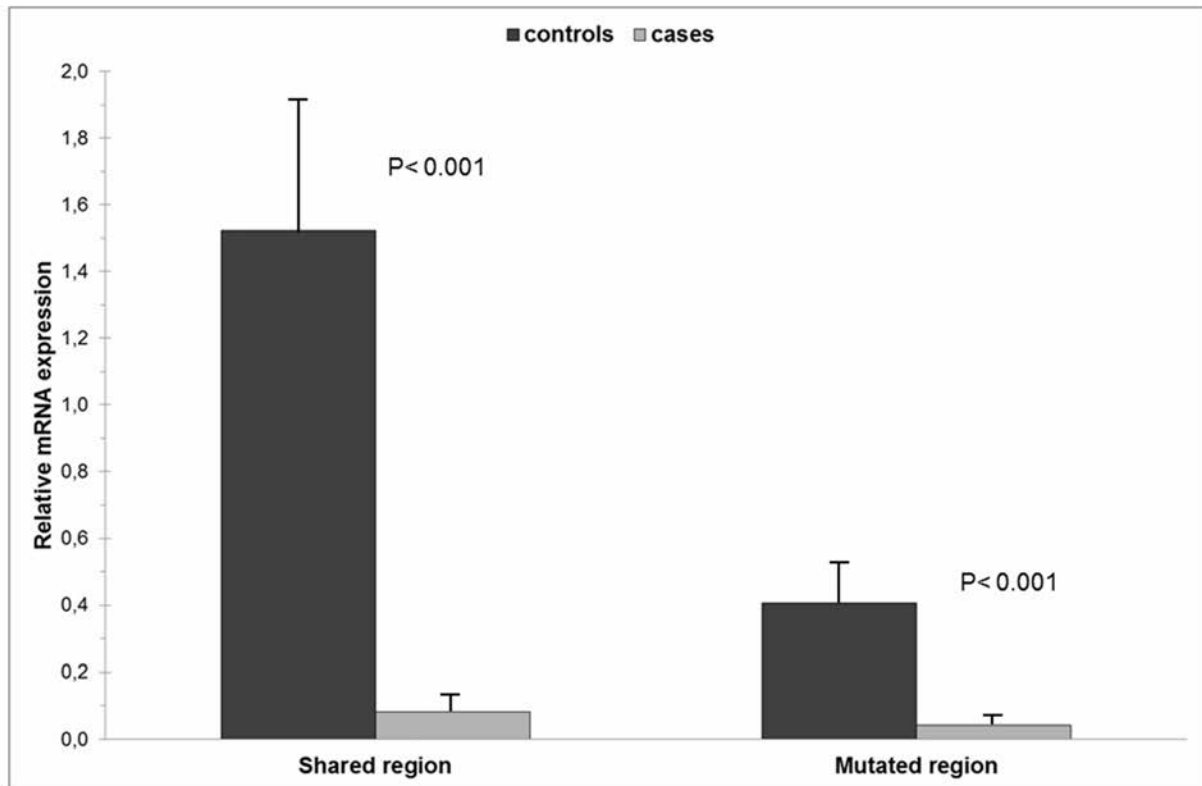
We herein demonstrate that LCH in sheep is due to a 31-bp deletion encompassing predicted exon 36 of the ovine *RELN* gene. This mutation was the only abnormality detected by sequencing the entire CDS of the *RELN* gene, which was the best positional and functional candidate in the region highlighted by GWA and homozygosity analyses. In this study, we revealed that homozygous individuals for this mutation lacked the extracellular protein reelin, which has been shown to play a key role in neuronal migration during the embryonic stage and whose absence causes LCH in humans and rodents [4,26]. The discovery of this mutation has allowed screening in the affected flock to identify phenotypically normal individuals carrying the 31-bp deletion to avoid at-risk matings. The implementation of this gene-assisted selection approach has solved the problem of LCH in Churra sheep flocks.

The results of this work support previous results in different livestock species, in which the use of genomic methodologies has had important practical consequences. High-throughput

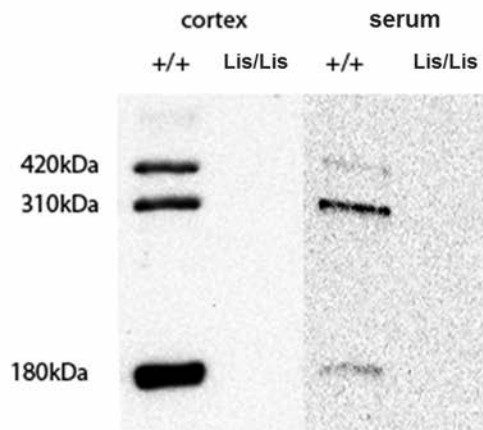
genotyping platforms allow the rapid control of emerging recessive defects with otherwise major economic and animal welfare implications [11,14]. In this study, 7 LCH cases, 13 related animals and 20 unrelated controls were genotyped using the OvineSNP50 BeadChip. Both the case-control association analysis and the run of homozygosity performed with the SNP-chip derived dataset suggested that the same region on chromosome OAR4 was the most likely to harbor the causal mutation responsible for LCH in Churra sheep. Within the homozygous consensus interval on OAR4 (42.4–47.2 Mb), we identified the *RELN* gene as a good functional and positional candidate for the studied phenotype. Based on the recessive mode of inheritance of the disease, which had been deduced based on a segregation analysis of affected pedigrees, a loss-of-function mutation in the coding sequence of the *RELN* gene was predicted.

Sequencing the cDNA of the *RELN* gene allowed us to obtain the complete coding sequence of this gene in a control sample. For the case sample, 96.21% of the sequence was obtained, revealing a deletion of 31 bp in putative exon 36 of the *RELN* gene (c.5410\_5440del), which results in the premature termination of the predicted protein and the loss of the final 1643 amino acids of the protein, including the loss of the highly basic C-terminal region [27,28]. The mutation was confirmed at the genomic level, and three genotypes cosegregate with disease status: homozygous mutants (Lis/Lis) were the affected lambs, heterozygous animals (Lis/+) corresponded to the phenotypically normal carriers (ewes and rams of the affected lambs), and the wild-type genotype (+/+) was identified in all the analyzed control individuals. These genotypes completely agreed with the autosomal recessive mode of Mendelian inheritance previously inferred for LCH in Churra sheep [10].

A



B



**Figure 4. Functional evaluation of the 31-bp deletion (c.5410\_5440del) in the ovine *RELN* gene.** (A) Expression of *RELN* mRNA analyzed by qRT-PCR. The figure shows the relative expression obtained from two regions of the mRNA: the “Shared region” includes 139 bp of the third reelin repeat, and the “Mutated region” amplifies a fragment that includes the fourth repeat, where the 31-bp deletion was located. The samples were analyzed in triplicate. The values were calculated using relative standard curves and were normalized to GAPDH from the same cDNA preparations and are expressed as the means  $\pm$  SEM. The specificity of the PCR products was confirmed by melting curve analysis after the qRT-PCR. In both regions significant differences ( $P < 0.001$ ) in mRNA level were evidenced using the one-tailed Student’s t-test. (B) Western blot immunoprobed for reelin of brain extracts (30  $\mu$ g) and serum samples (0.8  $\mu$ l) from control lamb (+/+) and lissencephaly-affected lamb homozygous for the deletion (Lis/Lis).

doi: 10.1371/journal.pone.0081072.g004

The low amount of mRNA in the LCH brain samples compared with the control samples (identified through the qRT-PCR analysis) may explain the difficulties experienced in the amplification of some regions of the coding sequence in the affected animals. The decreased levels of *RELN* mRNA may be the result of nonsense-mediated mRNA decay (NMD). NMD is a mechanism whereby mRNAs harboring premature termination codons are selectively degraded. This quality-control machinery prevents the production of truncated proteins with dominant-negative or deleterious gain-of-function activities [29] and is the likely reason for the absence of a protein signal in the Western blots of the LCH-affected lambs. Western blots were resolved with the anti-reelin antibody 142, an antibody raised to the N-terminus of reelin, and thus presumably should detect a hypothetical truncated C-terminal reelin.

Lissencephaly is a neuronal migration disorder characterized by a paucity or absence of cerebral sulcation and gyration accompanied by abnormal architecture of the cerebral cortex [30]. This disease comprises a group of brain developmental disorders with a heterogeneous genetic basis [31-33]. Human LCH, an autosomal recessive form of lissencephaly, has been described in two consanguineous pedigrees (OMIM #257320), and two different variants of the *RELN* gene were discovered as causal mutations in these patients [4]. Mutations that inactivate the *RELN* gene produce the majority of the identified mouse mutants with a phenotype similar to LCH, known as *reeler* [34-36]. However, mutant mice with phenotypes indistinguishable from *reeler* have also been identified by disrupting genes such as *Dab1*, *ApoER2* and *Vldlr*, which are involved in the reelin signaling pathway [2,37].

Among the known *RELN* mutations, the mouse Albany2 (*Reln*<sup>fl-Alb2</sup>) [35], the creeping rat (*Reln*<sup>cre</sup>) [38] and the human mutation found in a Saudi Arabian [4] pedigree resemble the ovine *RELN* mutation (Figure 5). All these mutations are expected to truncate the fourth specific reelin repeat of the predicted protein. As in the Albany2 *reeler*, the LCH-affected lambs exhibited reduced RNA concentrations, although no experimental protein analysis data was available from this mutant mouse [35]. Western blot analyses of sera from human LCH patients have shown a loss of reelin protein [4] similar to that occurring in the sheep cases reported here.

The resemblance between the mutations associated with LCH in the four species mentioned above supports the causative nature of the *RELN* gene 31-bp deletion reported here in relation to the LCH phenotype observed in Churra sheep and suggests that this is likely a null allele that is directly related to the LCH disease.

Development of the human brain requires a complex dynamic process of neuronal migration and positioning. A crucial problem in developmental neurobiology is to understand the genetic and biochemical pathways that regulate production

and migration of immature neurons. A detailed understanding of the mechanism of action of genes involved is necessary. Animal models have efficiently been used to understand the mechanisms of action and the pathophysiological abnormalities that result from mutations in genes involved in human lissencephaly [37,38]. A potential long-term goal such the development of target therapies for Human Lissencephaly would not be possible without a detailed understanding of the neuronal migration pathway.

Reelin protein has been identified as a major determinant of neuronal migration and plays a critical role in neuronal cell maturation during the development of central nervous system. *Reeler* mice mutants have been used as an opportunity to dissect the complex mechanism that underlies the molecular basis of the brain development [34]. Nonetheless, the adult primate brain displays remarkable differences with the rodent brain in the reelin-immunoreactive pattern [39]; therefore accessibility to large animal models maybe of interest in order to a better approach to human disease.

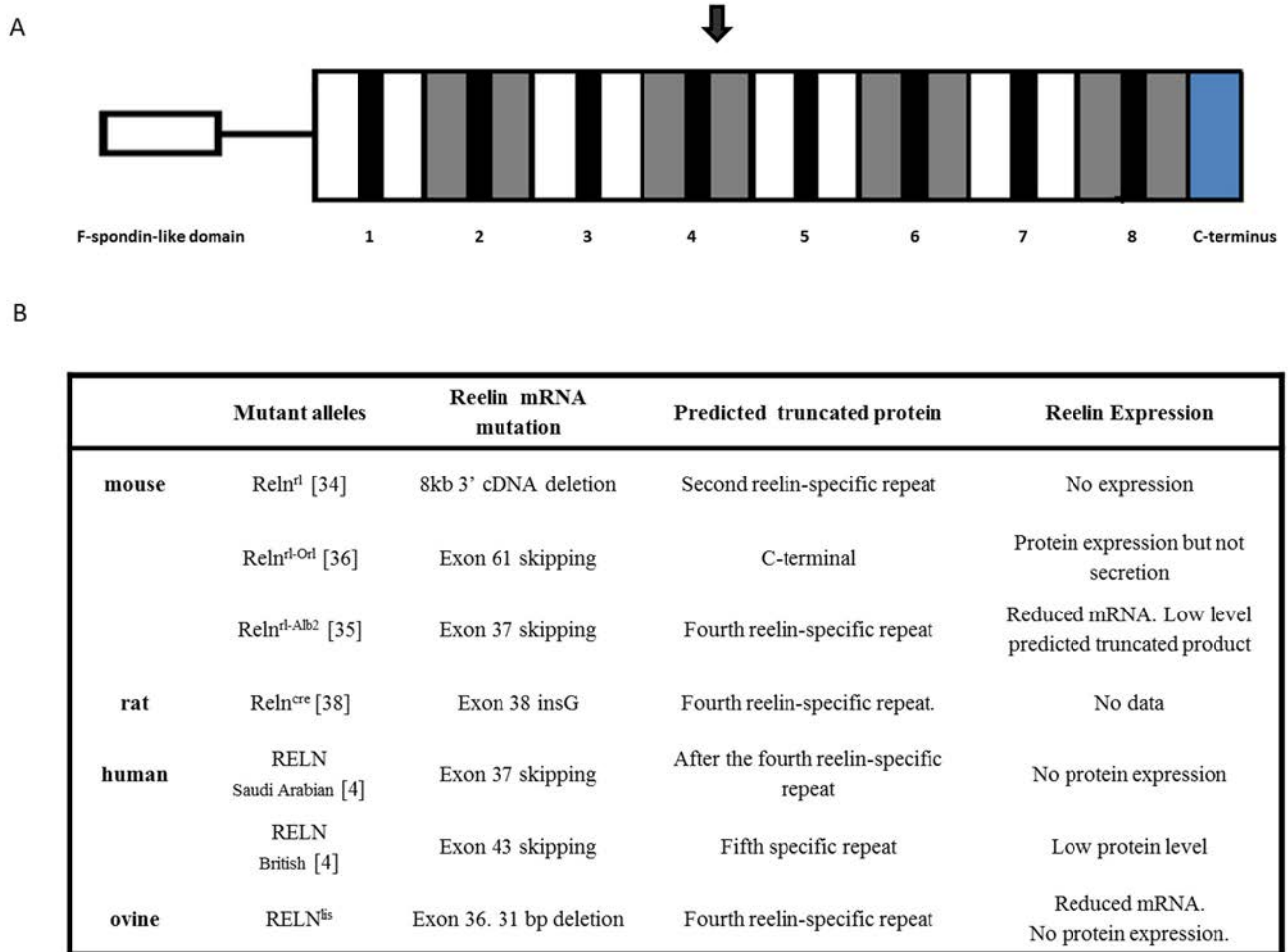
To our knowledge, the LCH sheep is the first large animal model with a confirmed mutation in *RELN*. A large *RELN* mutant model would be a valuable and powerful tool for studies of reelin function in neuronal migration and brain development. Large animal models present many advantages compared with small animal models, such as their greater physiological similarity to human patients and their larger body size [40].

Furthermore, animal models in which reelin expression is altered have provided important information about the actions of this protein in the mature brain. Recent evidence has shown that reelin is involved in modulating synaptic function in adults [41-45]. Several disorders, such as autism, schizophrenia, bipolar disorder, major depression and Alzheimer's disease, have been associated with abnormal levels of reelin in the serum and brain [46]. In this context, the longer life span and brain physiology of humans show a higher level of resemblance with the ovine model described here than with other small mutant models. Hence, the sheep *RELN* mutant model described in this study would allow a better understanding of reelin synaptic functions and their implications in the neuronal disorders described in the mature brain.

In conclusion, we have identified the causal mutation for ovine LCH in Churra sheep. This mutation (c.5410\_5440del) is located in putative exon 36 of the *RELN* gene and leads to the formation of a premature termination codon without protein expression.

The identification of the causal mutation for LCH in Churra sheep allows us to identify carriers and help breeders avoid at-risk matings between heterozygous animals. Moreover, the identification of *RELN* mutant sheep provides a suitable large model for the study of reelin functions in the developing and mature brain.





**Figure 5. Schematic representation of *RELN* mutant alleles.** (A) Predicted wild-type reelin protein consists of an F-spondin-like domain at the amino terminus, followed by 8 reelin repeats, each consisting of an internal EGF-like motif (black) flanked by reelin-specific sequence. The C-terminus includes highly basic aminoacids that appear to be required for secretion. The location of the predicted protein termination due to the ovine deletion is indicated by a black arrow. (B) Table comparing the *RELN* mutant alleles detected in different species indicating the effect of each mutation on reelin activity. The reference for each allele is indicated between square brackets.

doi: 10.1371/journal.pone.0081072.g005

## Supporting Information

### Figure S1. Predicted ovine Reelin aminoacid sequence.

Aminoacid sequences of sheep (upper sequence), human (P78509), mouse (Q60841) and rat (P58751) Reelin are aligned using the UniProt software (<http://www.uniprot.org/align>). The sequence comparison showed a high conservation of this protein across species. (PDF)

**Table S1. Primers used in the amplification cDNA of the ovine *RELN* gene.** The length of the amplified product and the melting temperature ( $T_m$ ) of the RT-PCR are also indicated. (DOCX)

**Table S2. Primers used to quantify the levels of *RELN* mRNA by qRT-PCR.** The length of the amplified product and the melting temperature ( $T_m$ ) are also indicated.

## References

- Jissendi-Tchofo P, Kara S, Barkovich AJ (2009) Midbrain-hindbrain involvement in lissencephalies. *Neurology* 72: 410-418. doi: 10.1212/01.wnl.0000333256.74903.94. PubMed: 19020296.
- Passemard S, Kaindl AM, Levreche V, Dournaud P, Gressens P (2010) Neuronal migration. In: H Lagercrantz MA Hanson LR Ment DM Peebles. In *The Newborn Brain: Neuroscience and Clinical Applications*. Cambridge: Cambridge University Press. pp. 55-70.
- Ross ME, Swanson K, Dobyns WB (2001) Lissencephaly with cerebellar hypoplasia (LCH): A heterogeneous group of cortical malformations. *Neuropediatrics* 32: 256-263. doi:10.1055/s-2001-19120. PubMed: 11748497.
- Hong SE, Shugart YY, Huang DT, Shahwan SA, Grant PE et al. (2000) Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human *RELN* mutations. *Nat Genet* 26: 93-96. doi: 10.1038/79246. PubMed: 10973257.
- Boycott KM, Flavell S, Bureau A, Glass HC, Fujiwara TM et al. (2005) Homozygous deletion of the very low density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification. *Am J Hum Genet* 77: 477-483. doi:10.1086/444400. PubMed: 16080122.
- Kumar RA, Pilz DT, Babatz TD, Cushion TD, Harvey K et al. (2010) TUBA1A mutations cause wide spectrum lissencephaly (smooth brain) and suggest that multiple neuronal migration pathways converge on alpha tubulins. *Hum Mol Genet* 19: 2817-2827. doi:10.1093/hmg/ddq182. PubMed: 20466733.
- Lambert de Rouvroit C, Goffinet AM (1998) The reeler mouse as a model of brain development. *Adv Anat Embryol Cell Biol* 150: 1-106. doi:10.1007/978-3-642-72257-8\_1. PubMed: 9816727.
- Summers BA, Cummings JF, de Lahunta A (1995) Malformations of the central nervous system. In: BA Summers JF Cummings A de Lahunta. *Veterinary Neuropathology*. St. Louis: Mosby. pp. 68-94.
- Herrmann A, Hecht W, Herden C (2011) Lissencephaly and microencephaly combined with hypoplasia of corpus callosum and cerebellum in a domestic cat. *Tierarztl Prax Ausg K Kleintiere Heimtiere* 39: 116-120. PubMed: 22143590.
- Perez V, Suárez-Vega A, Fuertes M, Benavides J, Delgado L et al. (2013) Hereditary lissencephaly and cerebellar hypoplasia in churra lambs. *BMC. Vet Res* 9: 156. doi:10.1186/1746-6148-9-156.
- Charlier C, Coppie W, Rollin F, Desmecht D, Agerholm JS et al. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40: 449-454. doi:10.1038/ng.96. PubMed: 18344998.
- Shariflou MR, Wade CM, Kijas J, McCulloch R, Windsor PA et al. (2012) Brachygnathia, cardiomegaly and renal hypoplasia syndrome (BCRHS) in merino sheep maps to a 1.1-megabase region on ovine chromosome OAR2. *Anim Genet* 44: 231-233. doi:10.1111/j.1365-2052.2012.02392.x. PubMed: 22762779.
- Mömke S, Kerkmann A, Wöhlke A, Ostmeier M, Hewicker-Trautwein M et al. (2011) A frameshift mutation within *LAMC2* is responsible for herlitz type junctional epidermolysis bullosa (HJEB) in black headed mutton sheep. *PLOS ONE* 6: e18943. doi:10.1371/journal.pone.0018943. PubMed: 21573221.
- Becker D, Tetens J, Brunner A, Bürstel D, Ganter M et al. (2010) Microphthalmia in texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (*PITX3*) gene. *PLOS ONE* 5: e8689. doi:10.1371/journal.pone.0008689. PubMed: 20084168.
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16: 1215. doi:10.1093/nar/16.3.1215. PubMed: 3344216.
- Glowatzki-Mullis ML, Muntwyler J, Gaillard C (2007) Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. *Anim Genet* 38: 86-88. doi:10.1111/j.1365-2052.2006.01550.x. PubMed: 17257200.
- Mäkinen VP, Parkkonen M, Wessman M, Groop PH, Kanninen T et al. (2005) High-throughput pedigree drawing. *Eur J Hum Genet* 13: 987-989. doi:10.1038/sj.ejhg.5201430. PubMed: 15870825.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575. doi:10.1086/519795. PubMed: 17701901.
- DeSilva U, D'Arcangelo G, Braden VV, Chen J, Miao GG et al. (1997) The human reelin gene: Isolation, sequencing, and mapping on chromosome 7. *Genome Res* 7: 157-164. doi:10.1101/gr.7.2.157. PubMed: 9049633.
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365-386. PubMed: 10547847.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL repository and associated resources. *Nucleic Acids Res* 37: D387-D392. doi:10.1093/nar/gkn750. PubMed: 18931379.
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201. doi:10.1093/bioinformatics/bti770. PubMed: 16301204.
- Peitsch MC (1995) Protein modeling by E-mail. *Nat Biotechnol* 13: 658-660. doi:10.1038/nbt0795-658.
- Botella-López A, Burgaya F, Gavín R, García-Ayllón MS, Gómez-Tortosa E et al. (2006) Reelin expression and glycosylation patterns are altered in Alzheimer's disease. *Proc Natl Acad Sci U S A* 103: 5573-5578. doi:10.1073/pnas.0601279103. PubMed: 16567613.
- Frotscher M (2010) Role for reelin in stabilizing cortical architecture. *Trends Neurosci* 33: 407-414. doi:10.1016/j.tins.2010.06.001. PubMed: 20598379.
- D'Arcangelo G, Miao GG, Chen SC, Soares HD, Morgan JI et al. (1995) A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* 374: 719-723. doi:10.1038/374719a0. PubMed: 7715726.
- Nakano Y, Kohno T, Hibi T, Kohno S, Baba A et al. (2007) The extremely conserved C-terminal region of reelin is not necessary for secretion but is required for efficient activation of downstream signaling.

(DOCX)

## Acknowledgements

We are grateful to the Churra sheep breeders association (ANCHE) for providing material and pedigree data and Lorena Alvarez for technical support. The support and availability of the computing facilities of the Foundation of Supercomputing Center of Castile and León (FCSC) (<http://www.fcsc.es>) are gratefully acknowledged.

## Author Contributions

Conceived and designed the experiments: JJA. Performed the experiments: ASV BGG ICI VP JB. Analyzed the data: EGG JSV. Contributed reagents/materials/analysis tools: JJA VP JSV. Wrote the manuscript: ASV BGG JJA.

- J Biol Chem 282: 20544-20552. doi:10.1074/jbc.M702300200. PubMed: 17504759.
28. de Bergeyck V, Nakajima K, Lambert de Rouvroit C, Naerhuyzen B, Goffinet AM et al. (1997) A truncated reelin protein is produced but not secreted in the 'orleans' reeler mutation (*reln<sup>rl-orl</sup>*). Brain Res. Mol Brain Res 50: 85-90. doi:10.1016/S0169-328X(97)00166-6. PubMed: 9406921.
  29. Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76: 51-74. doi: 10.1146/annurev.biochem.76.050106.093909. PubMed: 17352659.
  30. Sohal AP, Montgomery T, Mitra D, Ramesh V (2012) TUBA1A mutation-associated lissencephaly: Case report and review of the literature. Pediatr Neurol 46: 127-131. doi:10.1016/j.pediatrneurol.2011.11.017. PubMed: 22264709.
  31. Barkovich AJ, Guerrini R, Kuzniecky RI, Jackson GD, Dobyns WB (2012) A developmental and genetic classification for malformations of cortical development: Update 2012. Brain 135: 1348-1369. doi:10.1093/brain/aws019. PubMed: 22427329.
  32. Gleeson JG, Walsh CA (2000) Neuronal migration disorders: From genetic diseases to developmental mechanisms. Trends Neurosci 23: 352-359. doi:10.1016/S0166-2236(00)01607-6. PubMed: 10906798.
  33. Walsh CA (1999) Genetic malformations of the human cerebral cortex. Neuron 23: 19-29. doi:10.1016/S0896-6273(00)80749-7. PubMed: 10402190.
  34. D'Arcangelo G, Curran T (1998) Reeler: New tales on an old mutant mouse. Bioessays 20: 235-244. doi:10.1002/(SICI)1521-1878(199803)20:3. PubMed: 9631651.
  35. Royaux I, Bernier B, Montgomery JC, Flaherty L, Goffinet AM (1997) *Rein(rl-Alb2)*, an allele of reeler isolated from a chlorambucil screen, is due to an IAP insertion with exon skipping. Genomics 42: 479-482. doi: 10.1006/geno.1997.4772. PubMed: 9205121.
  36. Takahara T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N et al. (1996) Dysfunction of the orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. Hum Mol Genet 5: 989-993. doi:10.1093/hmg/5.7.989. PubMed: 8817336.
  37. Trommsdorff M, Gotthardt M, Hiesberger T, Shelton J, Stockinger W et al. (1999) Reeler/disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. Cell 97: 689-701. doi:10.1016/S0092-8674(00)80782-5. PubMed: 10380922.
  38. Yokoi N, Namae M, Wang HY, Kojima K, Fuse M et al. (2003) Rat neurological disease creeping is caused by a mutation in the reelin gene. Brain Res. Mol Brain Res 112: 1-7. doi:10.1016/S0169-328X(02)00650-2. PubMed: 12670697.
  39. Martínez-Cerdeño V, Galazo MJ, Cavada C, Clascá F (2002) Reelin immunoreactivity in the adult primate brain: intracellular localization in projecting and local circuit neurons of the cerebral cortex, hippocampus and subcortical regions. Cereb Cortex 12: 1298-1311. doi:10.1093/cercor/12.12.1298. PubMed: 12427680.
  40. Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF et al. (2011) A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in corriedale sheep. PLOS ONE 6: e21739. doi:10.1371/journal.pone.0021739. PubMed: 21747952.
  41. Ventruti A, Kazdoba TM, Niu S, D'Arcangelo G (2011) Reelin deficiency causes specific defects in the molecular composition of the synapses in the adult brain. Neuroscience 189: 32-42. doi:10.1016/j.neuroscience.2011.05.050. PubMed: 21664258.
  42. Groc L, Choquet D, Stephenson FA, Verrier D, Manzoni OJ et al. (2007) NMDA receptor surface trafficking and synaptic subunit composition are developmentally regulated by the extracellular matrix protein reelin. J Neurosci 27: 10165-10175. doi:10.1523/JNEUROSCI.1772-07.2007. PubMed: 17881522.
  43. Qiu S, Weeber EJ (2007) Reelin signaling facilitates maturation of CA1 glutamatergic synapses. J Neurophysiol 97: 2312-2321. doi:10.1152/jn.00869.2006. PubMed: 17229826.
  44. Herz J, Chen Y (2006) Reelin, lipoprotein receptors and synaptic plasticity. Nat Rev Neurosci 7: 850-859. doi:10.1038/nrn2009. PubMed: 17053810.
  45. Beffert U, Weeber EJ, Durudas A, Qiu S, Masiulis I et al. (2005) Modulation of synaptic plasticity and memory by reelin involves differential splicing of the lipoprotein receptor Apoer2. Neuron 47: 567-579. doi:10.1016/j.neuron.2005.07.007. PubMed: 16102539.
  46. Folsom TD, Fatemi SH (2013) The involvement of reelin in neurodevelopmental disorders. Neuropharmacology 68: 122-135. doi: 10.1016/j.neuropharm.2012.08.015. PubMed: 22981949.



**Procedimiento y Kit de Diagnóstico de Lisencefalia con Hipoplasia  
Cerebelar en Ganado Ovino.**

Aroa Suárez-Vega, Beatriz Gutiérrez-Gil y Juan José Arranz.

Dpto. Producción Animal, Universidad de León, 24071, León, Spain.

Oficina Española de Patentes y Marcas 2013, ES 2421207 B1.

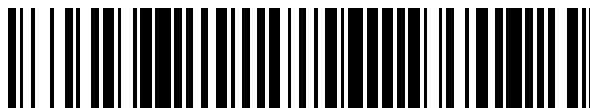


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 421 207**

21 Número de solicitud: 201330806

51 Int. Cl.:

**G01N 33/48** (2006.01)

12

## PATENTE DE INVENCION

B1

22 Fecha de presentación:

**31.05.2013**

43 Fecha de publicación de la solicitud:

**29.08.2013**

Fecha de la concesión:

**08.04.2014**

45 Fecha de publicación de la concesión:

**15.04.2014**

73 Titular/es:

**UNIVERSIDAD DE LEÓN (100.0%)  
Avda. de la Facultad, 25  
24071 León (León) ES**

72 Inventor/es:

**SUÁREZ VEGA, Aroa;  
GUTIÉRREZ GIL, Beatriz y  
ARRANZ SANTOS, Juan José**

74 Agente/Representante:

**CARVAJAL Y URQUIJO, Isabel**

54 Título: **Procedimiento y kit de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino**

57 Resumen:

Procedimiento y kit de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino. El procedimiento comprende la detección de la presencia o ausencia de un fragmento de ADN del gen RELN ovino, identificado por la secuencia SEQ ID NO: 1, en una muestra biológica obtenida de ganado ovino. Dicha identificación puede realizarse amplificando por PCR en dicha muestra un fragmento de ADN situado entre los cebadores con una identidad de al menos 80 % respecto a las secuencias SEQ ID NO: 2 y 3, determinando el número y tamaño de los fragmentos de ADN resultantes de dicha amplificación y realizando un diagnóstico de lisencefalia con hipoplasia cerebelar a partir de la información obtenida en el paso anterior.

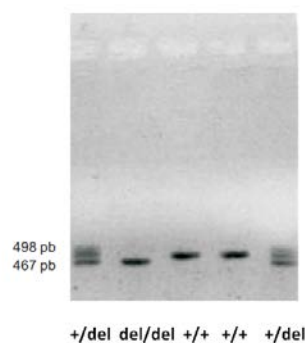


Fig. 1

ES 2 421 207 B1

## DESCRIPCIÓN

Procedimiento y kit de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino.

### 5 CAMPO DE LA INVENCION

El campo de la invención es el diagnóstico de enfermedades genéticas de herencia mendeliana simple (monogénica) en animales domésticos. En particular, la presente invención se refiere a la detección de la mutación responsable de la enfermedad genética denominada lisencefalia con hipoplasia cerebelar en el ganado ovino y al método de determinación del estatus genético de un individuo (sano, portador, enfermo) en relación con esta enfermedad.

### 15 ANTECEDENTES DE LA INVENCION

El término lisencefalia significa literalmente "cerebro liso" y engloba a un grupo de malformaciones severas del cerebro causadas por una migración anormal de las neuronas post-mitóticas al córtex en desarrollo. La lisencefalia se caracteriza por la ausencia total (agiria) o parcial (paquigiria) de circunvoluciones cerebrales. Se asocia con un córtex cerebral engrosado y heterotopia neuronal difusa.

Existen diversos tipos de lisencefalia, pero principalmente se clasifican en dos: lisencefalia clásica y lisencefalia pavimentosa. La lisencefalia con hipoplasia cerebelar (LHC) se constituye como una variante de estas, ya que a las malformaciones a nivel del córtex cerebral (que pueden ser de tipo clásica o pavimentosa) se asocia un subdesarrollo del cerebelo. Dentro de la LHC se pueden distinguir seis subtipos.

La base genética de la LHC ovina no ha sido descrita y por tanto no existe actualmente un marcador genético para esta enfermedad del ganado ovino que pueda utilizarse para predecir el estatus de portador o no portador de animales sanos en rebaños en los que se haya identificado la enfermedad.

El procedimiento diagnóstico de la LHC descrito en la presente invención se basa en que en la presente invención se ha identificado una mutación asociada a la enfermedad en un fragmento del gen *RELN*, que codifica para la proteína Reelina. La Reelina es una proteína extracelular implicada en la migración neuronal postmitótica durante el periodo embrionario que juega un papel clave en el desarrollo del sistema nervioso central (SNC).

En humanos, mutaciones en el gen *RELN* provocan una LHC conocida como lisencefalia tipo 2 o síndrome *Norman Roberts* (OMIM: 257320). En roedores, alteraciones en la reelina o en su cascada de señalización originan el fenotipo *reeler*. Los roedores *reeler* han sido ampliamente estudiados tanto por ser un modelo de la enfermedad humana como por su utilidad a la hora del estudio de la migración neuronal y desarrollo del SNC.

El gen *RELN* está situado en el cromosoma ovino 4 (OAR4), en la región 44.668.137- 45.204.609 pb según la versión Oarv3.1 del genoma ovino. La distribución de exones e intrones a lo largo de las 536.472 Kb del ADN que contiene el gen *RELN* no se ha determinado anteriormente.

### 40 DESCRIPCIÓN DE LA INVENCION

Una realización es un procedimiento de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino que comprende la detección de la presencia o ausencia de un fragmento de ADN del gen *RELN* ovino, identificado por la secuencia SEQ ID NO: 1, en una muestra biológica obtenida de ganado ovino, en adelante procedimiento de la invención.

Los inventores de la presente invención han determinado que el tamaño de la región codificante del gen es de 10383 pb distribuidos en un total de 64 exones.

50 La secuencia codificante del mRNA (CDS) del gen *RELN* ha sido depositada por los inventores de la presente invención en GenBank con fecha 07/02/2013, obteniéndose para ella el número de acceso KC590614.

El fragmento de ADN identificado por la secuencia SEQ ID NO: 1 corresponde a una delección de 31 pb entre las posiciones 5.410 y 5.440 de la secuencia codificante del mRNA del gen *RELN* ovino antes referida (GenBank Acc. no KC590614), correspondiente a las posiciones 44752377 y 44752407 pb en la secuencia de ADN del genoma ovino (Oarv3.1), en el exón 36 del gen. Esta delección se ha identificado por primera vez en la presente invención como responsable de la lisencefalia con hipoplasia cerebelar en oveja, mediante un estudio comparado de la secuencia codificante de este gen en animales enfermos y sanos.

60 El procedimiento de la invención permite detectar individuos sanos portadores de la enfermedad. Dichos portadores serán heterocigotos para la delección de 31 pb localizada en el exón 36 del gen, es decir tendrán una copia del gen completa y otra copia portadora de la delección.

65 Otra realización es el procedimiento de la invención, donde dicha detección se realiza con un biosensor o microarray.

Otra realización es el procedimiento de la invención, que comprende:

(a) amplificar por PCR en dicha muestra biológica un fragmento de ADN situado entre los cebadores con una identidad de al menos 80 % respecto a las secuencias SEQ ID NO: 2 y 3,

(b) determinar el número y tamaño de los fragmentos de ADN resultantes de dicha amplificación y

(c) realizar un diagnóstico de lisencefalia con hipoplasia cerebelar a partir de la información obtenida en el paso (b).

En la presente solicitud, dicho porcentaje de identidad en una secuencia determinada se calcula teniendo en cuenta que un 80% de identidad significa que un 80% de residuos de la secuencia completa de los cebadores identificados por las secuencias SEQ ID NO: 2 y 3 son idénticos a los residuos de la secuencia determinada.

Otra realización es el procedimiento de la invención, en el que antes de la etapa (a) se realiza una extracción de ADN de dicha muestra biológica.

La extracción del ADN a partir de estas muestras se puede realizar por métodos como el fenol-cloroformo descrito en Sambrook, 1989 (Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual. 2nd edition. New York: Cold Spring Harbor Laboratory Press); por el método "Salting out" descrito en Miller, 1988 (Miller, S.A. et al, 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research*, 16(3), pp. 1215) en el caso del semen o la utilización de Chelex-100 (BioRad Laboratories, Hercules, CA) para la extracción a partir de muestras de pelo. La extracción del ADN también se puede realizar utilizando Kits comerciales de extracción de ADN, sobre todo para muestras de saliva o de muestras de flujo nasal en las que se parte de una escasa cantidad de células nucleadas.

Otra realización es el procedimiento de la invención, donde dicha identidad de secuencia es de al menos 90%.

Otra realización es el procedimiento de la invención, donde dichos cebadores están identificados por las secuencias SEQ ID NO: 2 y SEQ ID NO: 3.

Otra realización es el procedimiento de la invención, donde dicha muestra biológica está seleccionada del grupo compuesto por pelo, tejido, célula o fluido biológico.

Otra realización es el procedimiento de la invención, donde dicho fluido biológico está seleccionado del grupo compuesto por saliva, leche, suero, plasma sanguíneo, sangre, semen y fluido nasal.

Otra realización es un kit para realizar el procedimiento de la invención, que comprende los cebadores identificados por las secuencias SEQ ID NO: 2 y SEQ ID NO: 3.

El procedimiento de la invención constituye un método sencillo, rápido y económico de detección de animales portadores, evitando que animales heterocigotos puedan seleccionarse para reposición en el rebaño y ser incluidos en el programa de mejora correspondiente, hechos que favorecerían la diseminación de la enfermedad.

El procedimiento de la invención puede utilizarse también para la detección de animales portadores en el caso de cruzamientos programados para obtener animales que sirvan como modelo animal en el estudio de la enfermedad.

El procedimiento de la invención puede aplicarse a corderos que presenten síntomas nerviosos compatibles con la enfermedad con el fin de descartarla sin necesidad de realizar una necropsia. Otras enfermedades con síntomas similares son la colibacilosis nerviosa, que en corderos recién nacidos produce delgadez y ataxia.

#### TEXTO LIBRE DE LA LISTA DE SECUENCIAS

A continuación se aporta una traducción del texto libre en inglés que aparece en la lista de secuencias.

SEQ ID NO: 1. Fragmento de delección del gen ovino *RELN*.

SEQ ID NO: 2. Cebador sentido para PCR.

SEQ ID NO: 3. Cebador antisentido para PCR.

SEQ ID NO: 4. Fragmento del gen salvaje ovino *RELN*.

188..365 Exón 36

277..307 Fragmento de delección

SEQ ID NO: 5. Fragmento del gen ovino *RELN* resultante de la delección.

188..334 Exón 36

#### BREVE DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Se muestran los resultados para 5 muestras, de izquierda a derecha, en la tercera y cuarta muestras se obtuvo una única banda de 498 pb, que indica genotipo homocigoto para el alelo salvaje (+/+), y por tanto animales sanos y no portadores;

en la segunda muestra se obtuvo una única banda de 467 pb, que indica genotipo homocigoto para la delección de 31 pb (del/del), únicamente identificable en individuos afectados por la enfermedad.

En la primera y quinta muestras se obtuvieron dos bandas de 498 y 467 pb, que indican genotipo heterocigoto para la delección (+/del), correspondiente a animales sanos pero portadores.

#### MODOS DE REALIZACIÓN PREFERENTE

5 Ejemplo 1. Diagnóstico genético de animales afectados por lisencefalia con hipoplasia cerebelar y animales portadores de dicha enfermedad

10 En una explotación ganadera nacieron corderos con incoordinación de movimiento, aunque con un estado de salud bueno. Se examinaron los corderos in vivo y se comprobó que sufrían un proceso neurológico que cursaba con ataxia e incapacidad para mantenerse en la estación.

15 En la necropsia se observaron alteraciones en el sistema nervioso central. En todos los casos se percibió una marcada ausencia de circunvoluciones cerebrales, con diferencias en cuanto a intensidad entre animales, pero siempre grave, tanto en la cara dorsal como en la ventral del cerebro así como una marcada hipoplasia cerebelar.

20 Se realizaron análisis anatomopatológicos correspondientes y se descartó la posible implicación de agentes infecciosos (como pestivirus y virus de lengua azul). A continuación se realizó un análisis del pedigrí de los animales afectados mediante marcadores microsatélite siguiendo el protocolo descrito en Glowatzki-Mullis, 2007 (Glowatzki-Mullis, M.L. et al, 2007. Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. *Animal Genetics*, 38(1), pp. 86-88). Tanto los análisis anatomopatológicos como los genéticos concluyeron que los animales sufrían lisencefalia con hipoplasia cerebelar.

#### Extracción del ADN a partir de la muestra biológica obtenida del animal en estudio.

25 En el presente ejemplo se utilizaron 4 de los corderos enfermos, las madres de los mismos (clasificadas como portadoras) y 41 machos de los que no se conocía el estatus genético. Las muestras utilizadas para la extracción de DNA fueron de sangre en todos los casos. La extracción de DNA se realizó por el método de "Salting out" descrito en Miller, 1988.

#### Diseño de los cebadores

35 Los cebadores se diseñaron específicamente con el programa *Primer3* (<http://frodo.wi.mit.edu/primer3>), a partir de la secuencia del genoma ovino v3.1, y tras la identificación en la misma de los exones identificados tras el análisis de la secuencia de mRNA del gen *RELN*. Así se partió de una región de 498 pb que incluía el Exon36, y 187 bp de la región intrónica anterior y 133 pb de la región intrónica posterior al mismo.

40 Se seleccionó una pareja de cebadores, cebador sentido, identificado con la secuencia SEQ ID NO: 2 y cebador antisentido, identificado con la secuencia SEQ ID NO: 3.

45 El tamaño del fragmento amplificado con esta pareja de cebadores fue 498 bp para la secuencia salvaje o no mutada, dicho fragmento incluye la secuencia completa del exón 36 (178 pb), 187 bp de la región intrónica anterior y 133 pb de la región intrónica posterior al mismo. En el caso de la presencia de la delección de 31 pb el tamaño del amplicón fue 467 pb.

#### Amplificación del ADN

50 Se realizó la amplificación del fragmento de ADN mediante la reacción en cadena de la polimerasa (PCR). En este estudio, las reacciones de PCR estuvieron constituidas por un 80% de lo que denominamos "mezcla", mientras que otro 20% fue de ADN total (10-30 ng/ml). La "mezcla" estuvo constituida por un 1X de tampón 10X (*Gold Applied Biosystem, AB*), MgCl<sub>2</sub> (de 1,5 µM a 2,5 µM), dNTPs (480 µM), los cebadores diseñados específicamente para la amplificación del fragmento de ADN de interés (0,5 µM), dH<sub>2</sub>O y la encima polimerasa (*AmpliTaQ Gold™ de AB*) (5 U/ml).

55 La amplificación se llevó a cabo en un termociclador automático (GeneAmp® PCR System 9700, AB). Los ciclos de temperatura siempre se vieron precedidos de un paso de desnaturalización a 95°C durante 5 minutos. Posteriormente se realizaron 35 ciclos, estando cada uno compuesto de 3 pasos:

- 60
- Desnaturalización del ADN: 94°C, 30s.
  - Hibridación de los cebadores con el ADN: 58°C, 40 s.
  - Síntesis de ADN: 72°C, 40 s.

El último ciclo finalizó en todos los casos con una extensión final de 10 minutos a 72°C.

### Separación y visualización del fragmento o fragmentos amplificados por electroforesis en gel de agarosa

5 Se utilizó un gel de agarosa al 2% (100 ml de TBE 1X y 2 g de agarosa) al que previamente se añadió 4 µl de RedSafe (iNtRON Biotechnology). El tampón TBE 1X se obtuvo a partir de una preparación previa de TBE 10X (Tris Base 54 g; Ácido Bórico: 27,5 g; EDTA 0.5 M 40 mL y agua mQ hasta 1L).

El gel se introdujo en una cubeta horizontal de electroforesis con 700 ml de tampón TBE 1X.

10 En cada pocillo del gel se cargaron 8 µl del producto de PCR y 4 µl de tampón de carga Ficoll. Asimismo, se utilizó uno de los pocillos de cada fila para cargar un marcador tamaño que nos permitió determinar con exactitud el tamaño de las muestras. El marcador utilizado fue el XIV de 100 pb (*Roche Diagnostic S.L.*) que contiene 15 fragmentos de ADN de doble cadena comprendidos entre 100 a 1500 pb y una banda adicional de 2642 pb.

15 En el inicio de la electroforesis se aplicó un voltaje de 100 V que se mantuvo constante durante 45 minutos.

Posteriormente se visualizaron las bandas amplificadas en un transiluminador de luz ultravioleta. Los resultados obtenidos permitieron agrupar las muestras en los siguientes grupos:

- 20
- muestras con una sola banda de 498 pb, correspondientes al genotipo homocigoto no mutado, es decir, a animales sanos y no portadores de la enfermedad.
  - muestras con una sola banda de 467 pb, correspondientes a animales homocigotos para la delección, es decir, a animales afectados por la enfermedad y
  - muestras con una doble banda de 467 y 498 pb, correspondientes al genotipo heterocigoto para la delección, es decir, a animales sanos pero portadores de la enfermedad.

25 En la Figura 1 se muestran un resultado representativo de este ejemplo, en el que se muestran los resultados obtenidos para muestras obtenidas de 5 animales diferentes.

## REIVINDICACIONES

1. Procedimiento de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino que comprende la detección de la presencia o ausencia de un fragmento de ADN del gen *RELN* ovino, identificado por la secuencia SEQ ID NO: 1, en una muestra biológica obtenida de ganado ovino.
- 5 2. Procedimiento según la reivindicación 1, caracterizado por que comprende:
  - (a) amplificar por PCR en dicha muestra biológica un fragmento de ADN situado entre los cebadores con una identidad de al menos 80 % respecto a las secuencias SEQ ID NO: 2 y 3,
  - (b) determinar el número y tamaño de los fragmentos de ADN resultantes de dicha amplificación y
  - (c) realizar un diagnóstico de lisencefalia con hipoplasia cerebelar a partir de la información obtenida en el paso (b).
- 10 3. Procedimiento según la reivindicación 2, caracterizado por que antes de la etapa (a) se realiza una extracción de ADN de dicha muestra biológica.
4. Procedimiento según la reivindicación 2, caracterizado por que dicha identidad de secuencia es de al menos 90%.
- 15 5. Procedimiento según la reivindicación 3, caracterizado por que dichos cebadores están identificados por las secuencias SEQ ID NO: 2 y SEQ ID NO: 3.
6. Procedimiento según la reivindicación 1, caracterizado por que dicha detección se realiza con un biosensor o microarray.
- 20 7. Procedimiento según cualquiera de las reivindicaciones 1 a 6, caracterizado por que dicha muestra biológica está seleccionada del grupo compuesto por pelo, tejido, célula o fluido biológico.
8. Procedimiento según la reivindicación 7, caracterizado por que dicho fluido biológico está seleccionado del grupo compuesto por saliva, leche, suero, plasma sanguíneo, sangre, semen y fluido nasal.
9. Kit para realizar el procedimiento según la reivindicación 2, caracterizado por que comprende los cebadores identificados por las secuencias SEQ ID NO: 2 y SEQ ID NO: 3.



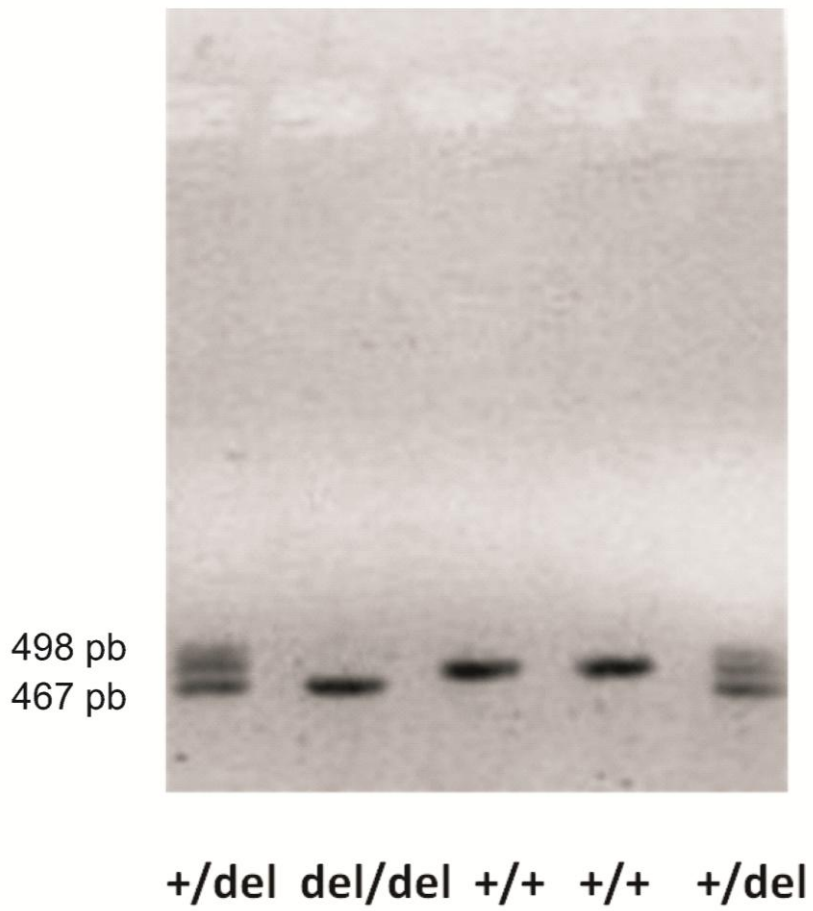


Fig. 1

Lista de Secuencias

<110> UNIVERSIDAD DE LEÓN

<120> PROCEDIMIENTO Y KIT DE DIAGNÓSTICO DE LISENCEFALIA CON HIPOPLASIA CEREBELAR EN GANADO OVINO

<130> P3745/2013

<160> 5

<170> BiSSAP 1.2

<210> 1

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<222> 1..31

<223> /mol\_type="unassigned DNA"  
/note="Ovine RELN gene deletion fragment"  
/organism="Artificial Sequence"

<400> 1

gatgtaagtt cccattgaaa tcatctttaa g

31

<210> 2

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<222> 1..20

<223> /mol\_type="unassigned DNA"  
/note="Sense PCR primer"  
/organism="Artificial Sequence"

<400> 2

ttgccttctc cggtttaatg

20

<210> 3

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<222> 1..20

<223> /mol\_type="unassigned DNA"  
/note="Antisense PCR primer"  
/organism="Artificial Sequence"

<400> 3

<210> 4  
 <211> 498  
 <212> DNA  
 <213> Ovis aries

<220>  
 <221> source  
 <222> 1..498  
 <223> /mol\_type="unassigned DNA"  
 /note="Wild type ovine RELN gene fragment"  
 /organism="Ovis aries"

<220>  
 <221> exon  
 <222> 188..365  
 <223> /note="Exon 36"

<220>  
 <221> variation  
 <222> 277..307  
 <223> /note="Deletion fragment"

<400> 4  
 ttgccttctc cggtttaatg cccaagtgat cataaaaaat gatctcagtc gatataagag 60  
 gaaacatgat atgtgttaga agacagtaac agtatataca ataatatatc atctgaacag 120  
 aatgtatct atgtggagca ttgagaaaag acaggaattt cagtgctagt gaacagtctt 180  
 atctgacccc tttaaaaatg agcgatgttc cagacttgat ggtttcacca ttcagattcc 240  
 ccctctctgc accatacact tcaggccaaa ggtcgggatg taagttccca ttgaaatcat 300  
 ctttaagaat cgagggtaga ggaataacag gaacacagta ggctccacca aagccccggt 360  
 cacacctaag gaaaaaatg agtggagaaa aggcattata tgtcatgaga tcaagatact 420  
 atctttcttt ttgcattaca aaaaaagaaa ctgccaatth cattaactta cacacagcgt 480  
 ccagcatcac aaatccct 498

<210> 5  
 <211> 467  
 <212> DNA  
 <213> Ovis aries

<220>  
 <221> source  
 <222> 1..467  
 <223> /mol\_type="unassigned DNA"  
 /note="Ovine RELN gene fragment resulting from deletion"  
 /organism="Ovis aries"

<220>

<221> exon  
<222> 188..334  
<223> /note="Exon 36"

<400> 5  
ttgccttctc cggtttaatg cccaagtgat cataaaaaat gatctcagtc gatataagag 60  
gaaacatgat atgtgtaga agacagtaac agtatataca ataatatatc atctgaacag 120  
aatgtatct atgtggagca ttgagaaaag acaggaattt cagtgctagt gaacagtctt 180  
atctgacccc tttaaaaatg agcgatgttc cagacttgat ggtttcacca ttcagattcc 240  
ccctctctgc accatacact tcaggccaaa ggtcggaatc gagggtagag gaataacagg 300  
aacacagtag gctccaccaa agccccggtc acacctaagg aaaaaaatga gtggagaaaa 360  
ggcattatat gtcatgagat caagatacta tctttctttt tgcattacaa aaaaagaaac 420  
tgccaatttc attaacttac acacagcgtc cagcatcaca aatccct 467



- ②<sup>1</sup> N.º solicitud: 201330806  
 ②<sup>2</sup> Fecha de presentación de la solicitud: 31.05.2013  
 ③<sup>2</sup> Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑤<sup>1</sup> Int. Cl.: **G01N33/48** (2006.01)

DOCUMENTOS RELEVANTES

| Categoría | ⑤ <sup>6</sup> Documentos citados   | Reivindicaciones afectadas |
|-----------|---|----------------------------|
| A         | WO 2012003353 A2 (UNIV YALE et al.) 05.01.2012, reivindicaciones.                     | 1-9                        |
| A         | WO 2008030065 A1 (MACROGEN INC et al.) 13.03.2008, reivindicaciones.                  | 1-9                        |
| A         | JP 2004081001 A (MITSUBISHI CHEM CORP et al.) 18.03.2004, recuperado de WPI, resumen. | 1-9                        |
| A         | WO 03091426 A1 (SPECTRAL GENOMICS INC et al.) 06.11.2003, reivindicaciones.           | 1-9                        |

Categoría de los documentos citados

X: de particular relevancia  
 Y: de particular relevancia combinado con otro/s de la misma categoría  
 A: refleja el estado de la técnica

O: referido a divulgación no escrita  
 P: publicado entre la fecha de prioridad y la de presentación de la solicitud  
 E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

**El presente informe ha sido realizado**

para todas las reivindicaciones

para las reivindicaciones n.º:

|   |                                      |                      |
|---|--------------------------------------|----------------------|
| <b>Fecha de realización del informe</b><br>08.08.2013 | <b>Examinador</b><br>I. Rueda Molíns | <b>Página</b><br>1/4 |
|---|--------------------------------------|----------------------|

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G01N

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, STN, TXT

Fecha de Realización de la Opinión Escrita: 08.08.2013

**Declaración**

|   |                      |           |
|---|----------------------|-----------|
| <b>Novedad (Art. 6.1 LP 11/1986)</b>            | Reivindicaciones 1-9 | <b>SI</b> |
|   | Reivindicaciones     | <b>NO</b> |
| <b>Actividad inventiva (Art. 8.1 LP11/1986)</b> | Reivindicaciones 1-9 | <b>SI</b> |
|   | Reivindicaciones     | <b>NO</b> |

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

**Base de la Opinión.-**

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

**1. Documentos considerados.-**

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

| Documento | Número Publicación o Identificación           | Fecha Publicación |
|-----------|---|-------------------|
| D01       | WO 2012003353 A2 (UNIV YALE et al.)           | 05.01.2012        |
| D02       | WO 2008030065 A1 (MACROGEN INC et al.)        | 13.03.2008        |
| D03       | JP 2004081001 A (MITSUBISHI CHEM CORP et al.) | 18.03.2004        |
| D04       | WO 03091426 A1 (SPECTRAL GENOMICS INC et al.) | 06.11.2003        |

**2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración****NOVEDAD Y ACTIVIDAD INVENTIVA (artículos 6 y 8 Ley 11/1986)**

En las reivindicaciones 1-8, de la solicitud de patente, se reivindica un procedimiento de diagnóstico de lisencefalia con hipoplasia cerebelar en ganado ovino, que comprende la detección de la presencia o ausencia de un fragmento de ADN del gen *RELN* ovino, identificado por la secuencia SEQ ID NO:1, en una muestra biológica obtenida de ganado. En la reivindicación 9, se reivindica un kit para la realización de dicho procedimiento.

Se ha efectuado una búsqueda de la secuencia de nucleótidos SEQ ID NO: 1, reivindicada en la solicitud de patente, esa secuencia no se ha encontrado en el estado de la técnica. Los documentos D01-D04 divulgan diferentes métodos de detección de lisencefalia. En ninguno de los citados documentos se refleja un procedimiento de diagnóstico de lisencefalia como el que se reivindica en la solicitud de patente. Por tanto, las reivindicaciones 1-9 presentan novedad y actividad inventiva, según lo establecido en los artículos 6 y 8 de la Ley 11/1986.





## Combining GWAS and RNA-Seq approaches for detection of the causal mutation for hereditary junctional epidermolysis bullosa in sheep.

Aroa Suárez-Vega<sup>1</sup>, Beatriz Gutiérrez-Gil<sup>1</sup>, Julio Benavides<sup>2</sup>, Valentín Pérez<sup>2</sup>, Gwenola Tosser-Klopp<sup>3,4,5</sup>, Christophe Klopp<sup>6</sup>, Stephen J. Keenel<sup>7</sup> and Juan José Arranz<sup>1</sup>.

<sup>1</sup> Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain; <sup>2</sup> Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE), Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071; <sup>3</sup> INRA, UMR1388 GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31326 Castanet-Tolosan, France; <sup>4</sup> Université de Toulouse, INP, ENSAT, GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31326 Castanet-Tolosan, France; <sup>5</sup> Université de Toulouse, INP, ENVIT, GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31076 Toulouse, France; <sup>6</sup> INRA, Plateforme bioinformatique Toulouse Midi-Pyrénées, UR875 Biométrie et Intelligence Artificielle, BP 52627, 31326 Castanet-Tolosan Cedex, France ; <sup>7</sup> Graduate School of Medicine, University of Tennessee, Knoxville, Tennessee, United States of America.

PLoS ONE (2015), 10(5): e0126416. doi: 10.1371/journal.pone.0126416.



RESEARCH ARTICLE

# Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep

Aroa Suárez-Vega<sup>1</sup>, Beatriz Gutiérrez-Gil<sup>1</sup>, Julio Benavides<sup>2</sup>, Valentín Pérez<sup>2</sup>, Gwenola Tosser-Klopp<sup>3,4,5</sup>, Christophe Klopp<sup>6</sup>, Stephen J. Keenel<sup>7</sup>, Juan José Arranz<sup>1\*</sup>

**1** Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071, León, Spain, **2** Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE), Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071, León, Spain, **3** INRA, UMR1388 GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31326, Castanet-Tolosan, France, **4** Université de Toulouse, INP, ENSAT, GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31326, Castanet-Tolosan, France, **5** Université de Toulouse, INP, ENVT, GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31076, Toulouse, France, **6** INRA, Plateforme bioinformatique Toulouse Midi-Pyrénées, UR875 Biométrie et Intelligence Artificielle, BP 52627, 31326, Castanet-Tolosan Cedex, France, **7** Graduate School of Medicine, University of Tennessee, Knoxville, 37920, Tennessee, United States of America

\* [jjarrs@unileon.es](mailto:jjarrs@unileon.es)



OPEN ACCESS

**Citation:** Suárez-Vega A, Gutiérrez-Gil B, Benavides J, Pérez V, Tosser-Klopp G, Klopp C, et al. (2015) Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. PLoS ONE 10(5): e0126416. doi:10.1371/journal.pone.0126416

**Academic Editor:** Gerhard Wiche, University of Vienna, AUSTRIA

**Received:** February 2, 2015

**Accepted:** April 1, 2015

**Published:** May 8, 2015

**Copyright:** © 2015 Suárez-Vega et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The transcriptomic data (RNA-Seq) are available from the European Nucleotide Archive database (<http://www.ebi.ac.uk/ena>) with the accession numbers: ERS631692 and ERS631693. The association data, genotypes and phenotypes have been submitted to the Dryad repository (doi:10.5061/dryad.4j7jh).

**Funding:** Funded by Spanish Ministry of Economy and Competitiveness (grant AGL2012-34437 to JJA); Aroa Suarez-Vega is funded by an FPU contract from Spanish Ministry of Education. Beatriz Guterrez-Gil is

## Abstract

In this study, we demonstrate the use of a genome-wide association mapping together with RNA-seq in a reduced number of samples, as an efficient approach to detect the causal mutation for a Mendelian disease. Junctional epidermolysis bullosa is a recessive genodermatosis that manifests with neonatal mechanical fragility of the skin, blistering confined to the lamina lucida of the basement membrane and severe alteration of the hemidesmosomal junctions. In Spanish Churra sheep, junctional epidermolysis bullosa (JEB) has been detected in two commercial flocks. The JEB locus was mapped to *Ovis aries* chromosome 11 by GWAS and subsequently fine-mapped to an 868-kb homozygous segment using the identical-by-descent method. The *ITGB4*, which is located within this region, was identified as the best positional and functional candidate gene. The RNA-seq variant analysis enabled us to discover a 4-bp deletion within exon 33 of the *ITGB4* gene (*c.4412\_4415del*). The *c.4412\_4415del* mutation causes a frameshift resulting in a premature stop codon at position 1472 of the integrin  $\beta 4$  protein. A functional analysis of this deletion revealed decreased levels of mRNA in JEB skin samples and the absence of integrin  $\beta 4$  labeling in immunohistochemical assays. Genotyping of *c.4412\_4415del* showed perfect concordance with the recessive mode of the disease phenotype. Selection against this causal mutation will now be used to solve the problem of JEB in flocks of Churra sheep. Furthermore, the identification of the *ITGB4* mutation means that affected sheep can be used as a large mammal animal model for the human form of epidermolysis bullosa with aplasia cutis. Our approach

funded by an Ramon y Cajal contract from the Spanish Ministry of Economy and Competitiveness. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

evidences that RNA-seq offers cost-effective alternative to identify variants in the species in which high resolution exome-sequencing is not straightforward.

## Introduction

Epidermolysis bullosa (EB) is a clinically and genetically heterogeneous group of inherited mechanobullous disorders caused by mutations in several structural skin proteins. EB is characterized by structural and mechanical fragility of the skin, leading to recurrent blistering and erosions that impair the life of EB affected individuals [1]. Following the latest classification [2], there are four major types of EB: EB simplex (EBS), junctional EB (JEB), dystrophic EB (DEB) and Kindler. This classification has been developed based on the level of blisters and the location of the aberrant proteins involved.

JEB encompasses all of the subtypes of EB that have mechanical fragility and blistering confined to the lamina lucida of the basement membrane and severe alteration of the hemidesmosomal junctions [1]. The hemidesmosomes are multi-protein complexes that play a critical role in the stable association of basal epithelial cells to the underlying basement membrane in stratified and pseudostratified epithelia [3]. Mutations in the genes encoding hemidesmosomal proteins and functionally associated structures have been shown to result in different subtypes of JEB [4, 5]. JEB occurs in various mammalian species [6–13]. In sheep, a Herlitz subtype of JEB has been described in German Black-headed Mutton sheep [14–16].

In Spanish Churra sheep, JEB has been detected in two commercial flocks belonging to the National Churra Sheep Breeders' Association (ANCHE). The affected animals show erosions and ulcers of the skin and mucous membranes (Fig 1). Initial analyses showed that the JEB in Churra sheep is a recessive autosomal inherited defect. Microscopically, JEB with congenital absence of skin have been diagnosed. In humans, the concurrence of JEB and congenital absence of skin is described in a subtype of JEB known as JEB with pyloric atresia (JEB-PA) [2]. Pyloric atresia has not been identified in the affected Churra lambs. Because of the clinical and molecular heterogeneity of EB, there are some rare cases of JEB-PA-affected humans who did not show pyloric atresia [17]. JEB-PA is related to mutations in the *ITGA6* and *ITGB4* genes [18–22].

In recent years, genome-wide mapping strategies have proven to be effective approaches for the identification of defective genes underlying diseases with Mendelian inheritance patterns [15, 23, 24]. The methods to identify the causal mutation in a Mendelian disorder generally involve the detection of a disease locus followed by candidate gene sequencing. Recently, Next Generation Sequencing (NGS) technologies have enabled high-throughput functional genomic research to be performed [25]. High-throughput sequencing of RNA (RNA-seq) has been developed to identify and quantify the gene expression in different tissues [26, 27]. RNA-seq also offers new opportunities for efficient and cost-effective detection of transcriptome variants (SNPs and short indels) in different tissues and species [28–30]. In humans, recent studies have highlighted the advantages of using NGS in EB diagnostics [31, 32].

In the present study, our aim was to find the causative mutation underlying JEB in Spanish Churra sheep. For this goal, a SNP array was used to map the causative region of JEB, and then, the disease-causing mutation was identified via an RNA-seq variant analysis. This study demonstrates the effectiveness of combining both approaches to map and identify the causal mutations in recessive Mendelian diseases.



**Fig 1. Clinical features of JEB in Churra sheep.** (A) Affected lamb showing skin lesions at the ears and limbs. (B) Erosions in the tongue. (C) Affected lamb showing severe skin lesions that involve the ears, limbs and caudal part of the body.

doi:10.1371/journal.pone.0126416.g001



## Results

### Mapping the causative gene for JEB

Approximately 44,800 evenly spaced SNPs were genotyped in 20 JEB-affected lambs and 28 related unaffected and 48 unrelated control sheep. A significant genome-wide association between the SNP genotypes and JEB was shown for sheep chromosome 11 (OAR11) (Fig 2A). The marker rs410387229, located at position OAR11:54,939,690 bp of the OARv3.1 sheep genome assembly, showed a genome-wide corrected p-value of 0.00001. There were three additional SNP markers on OAR11, rs418150485, rs425824336 and rs412661420, located at positions 56,897,973, 54,003,142 and 58,681,220 bp, respectively, with corrected p-values < 0.0001.

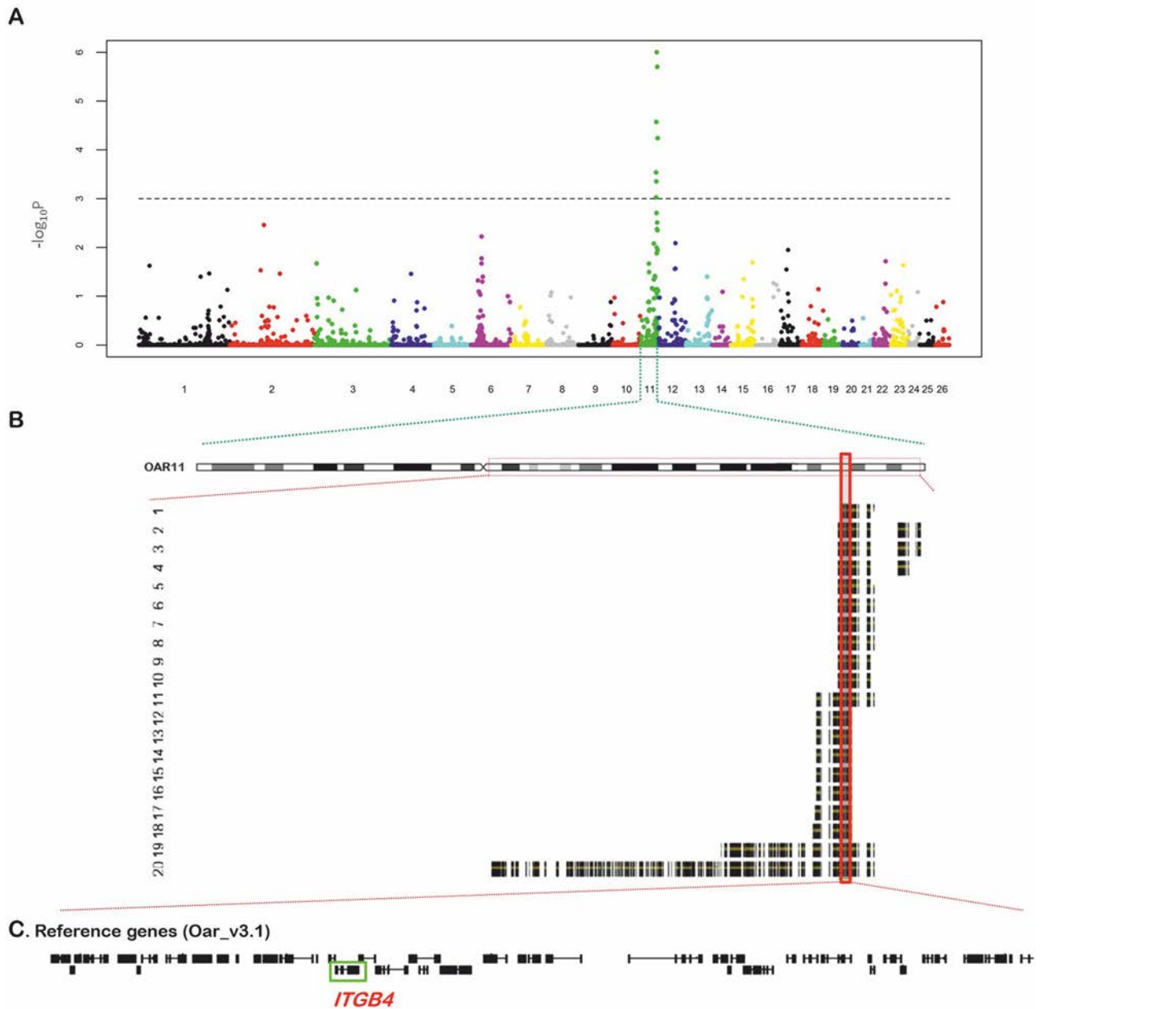
The location of the JEB locus was refined using the identical-by-descent (IBD) mapping method. The IBD analysis narrowed the defective region to an 868-kb homozygous segment, from 54,632,309–55,500,100 bp on OAR11, in all 20 affected lambs (Fig 2B). The consensus region contains 16 consecutive SNP loci. Regarding the chromosomal location of the genes related to JEB in humans (OMIM #226730: *ITGB4*, *ITGA6*), the *ITGB4* gene was identified as a promising positional and functional candidate, as it is located at 54,848,050–54,874,496 on OAR11, which is within the identified block of homozygosity common to all of the cases (Fig 2C). The *ITGB4* gene codes for integrin  $\beta 4$  protein. In stratified epithelia, integrin  $\alpha 6\beta 4$  is located in specialized epidermal cell-basement membrane adhesion structures, called hemidesmosomes [33].

### Ovine *ITGB4* gene sequence and annotation

Because of its association with human JEB [17, 22, 34], the *ITGB4* gene was identified as the most promising candidate. Our first aim was to identify the causal mutation via Sanger sequencing of the *ITGB4* gene in four JEB cases and four unrelated controls. Through the Sanger approach for genomic sequencing, we were able to obtain 84% (29,936 bp) of the complete genomic sequence of the *ITGB4* gene (ENSOARG0000009764), in the four cases and four controls. By comparing the sequences of the unaffected and the JEB-affected lambs, we identified eight single nucleotide polymorphisms (SNPs) that had different genotypes in the cases (*g.54857417T>G*, *g.54857293C>T*, *g.54856642A>G*, *g.54853969C>G*, *g.5485255C>A*, *g.54851357C>T*, *g.54850814C>T* and *g.54848707G>A*). All of the identified SNPs were located in non-coding *ITGB4* regions. A search for the SNPs in the International Sheep Genomic Consortium (ISGC) database (<https://data.csiro.au/dap/landingpage?execution=e1s2>) allowed us to discard these mutations as causal mutations, as all of them were already published on healthy sheep from other breeds.

Because of the difficulties in performing Sanger DNA sequencing of the whole gene using the *ITGB4* reference sequence available in Ensembl (ENSOARG0000009764), we isolated the *ITGB4* sequence from three ovine BAC clones (CH243-344E15, CH243-306J7 and CH243-504K6) [35]. The BAC clone sequencing allowed us to obtain the complete genomic sequence of *ITGB4* (GenBank: KP025765). Pairwise alignment between the sheep *ITGB4* reference obtained via BAC clone sequencing and the *ITGB4* reference available in Ensembl (ENSOARG0000009764) revealed an identity of 83.1% between the two sequences and a gap percentage of 16.6%. Excluding the 5'- and 3'UTR regions that were not available in the Ensembl sequence, 14 gaps, ranging from 2 to 1119 bp, were detected (S1 Fig).

In parallel to the BAC clone sequencing, an RNA-seq analysis was conducted. Sequencing of the cDNA libraries from the RNA skin samples from one JEB case and one control generated a total of 69.31-million paired-end reads, with an average of 34.65-million paired-end reads for



**Fig 2. Workflow chart for the discovery of the mutant locus of inherited JEB in Churra sheep.** (A) Manhattan plot that shows the results of the case-control GWA analysis performed. The X-axis shows the positions of the SNPs analyzed across the 26 ovine autosomes; different colors represent each different chromosome. The Y-axis represents the  $-\log_{10}P$  values obtained after performing 1 million permutations using PLINK. The most significant associated SNP is located at position 54,939,690 bp on OAR11. (B) IBD analysis on OAR11. The analysis of SNPs genotypes from affected lambs shows an extended homozygous region on chromosome 11. All twenty affected lambs shared overlapping homozygous blocks (black and yellow blocks, which represent the genes within the homozygous block of each animal). The consensus haplotype block (indicated by a solid red box) that contains the putative causal mutation expands a region of 868-kb, from 54,632,309–55,500,100 bp on OAR11. (C) Gene content of the 868-kb interval shared by the JEB affected.

doi:10.1371/journal.pone.0126416.g002

each sample. The aligner STAR [36] was used to map the paired-end reads from each sample to the ovine reference genome sequence (Oar\_v3.1). Approximately 83.54% of the reads (85.69% in the case and 81.39% in the control) were mapped to a unique location in the ovine v3.1 genome. The combination of the *ITGB4* sequence obtained from the BAC clone with the RNA-seq data from the control allowed us to predict the complete CDS and the integrin  $\beta 4$

protein sequence. To determine the gene structure of the *ITGB4* sequence, Augustus software was used [37, 38]. In total, of 5360 reads that were overlapping the *ITGB4* region in the control sample were used to generate a hints file for Augustus. A complete CDS of 5256 bp was found. This transcript is distributed into 38 exons. The sheep *ITGB4* CDS sequence encodes a predicted protein of 1751 amino acids. The BLAST tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to compare the predicted amino acid sequence from the ovine *ITGB4* gene with several available sequences, including the cow (DAA18168.1), camel (EQB77252.1) human (aai43743.1) and mouse (NP\_598424.2) sequences. The predicted ovine integrin  $\beta 4$  protein sequence showed 92% homology with the bovine protein, 92% homology with the camel sequence, 89% homology with the human, and 88% homology with the mouse sequence.

### Detection of the genetic variation between the JEB-affected and unaffected skin samples via RNA-Seq

GATK [39] was used for variant detection in the RNA-seq data from the JEB-affected and unaffected samples. GATK can identify single base substitutions in addition to small insertions and deletions. In total, 495921 different variants were detected using the RNA-seq reads in both samples. The VEP tool from Ensembl processed a total of 490,196 variants, among which 5.2% (25247) corresponded to already-described variants and 94.8% (464949) corresponded to novel variants. There were 20,828 variants overlapping with known transcripts. Most of the variants detected were SNPs (91.5%), and 8.4% (40956) were indels.

Among the variants, we selected those that overlapped with the *ITGB4* gene. In total, 54 variants were found in the *ITGB4* region (OAR11: 54,848,050–54,874,496). Following the assumption that JEB in the Churra sheep is caused by a homozygous recessive mutation, the variants were sifted to retain those present at a homozygous stage in the affected case (1/1) and absent in the control animal (0/0). Fourteen variants matched this filter: nine intron variants, three synonymous variants, one missense variant, and one frameshift variant and feature truncation. We discarded the mutations located in introns, which had average DP (coverage) values of 3.2 in the case and 5.6 in the control, and the synonymous variants.

The missense variant detected in the RNA-seq variant analysis is a G>A substitution located at OAR11, which was 54,868,894 bp of the sheep Oar\_v3.1 assembly. Regarding the *ITGB4* sequence (GenBank: KP025765) the mutation is positioned at 9,381 bp of the *ITGB4* genomic sequence. In the predicted CDS, the SNP is located at 1337 bp (*c.1337G>A*), in exon ten of the *ITGB4* gene (Fig 3). The mutation leads to an arginine-to-glutamine residue change at position 446 of the integrin  $\beta 4$  protein (*G>A p.Arg446Gln*). The genotype quality (GQ) value for *c.1337G>A* was 99 for the control and 42 for the case. In general, the determined coverage (DP value) for the *ITGB4* variants was higher in the control lamb than in the case; for this SNP the DP value was 75 and 14, respectively. This non-synonymous SNP has been observed at a homozygous stage in other sheep breeds (ISGC SNP database) and was predicted to be non-deleterious by SIFT [40] and PolyPhen-2 [41]. Therefore, the *c.1337G>A* variant was discarded as a putative causal mutation of JEB.

The frameshift variant detected in the RNA-seq variant analysis is a four-bp deletion located at OAR11:54,849,767–54,849,770 bp of sheep Oar\_v3.1 assembly. In the *ITGB4* sequence (GenBank: KP025765) the mutation is positioned at 30,204–30,207 bp of the *ITGB4* genomic sequence. In the predicted CDS, the deletion is located at 4,412–4,415 bp (*c.4412\_4415del*). The GQ value for the deletion was 99 for the control and 35 for the case, and the DP value was 180 and 11, respectively. This mutation is located on sheep *ITGB4* predicted exon 33 and leads to a shift in the open reading frame of the *ITGB4* coding sequence and a stop codon at position 1,473 (Fig 3). This mutation could lead to a truncated protein of 1,472 amino acids (1,470





$P < 0.001$ ). The mRNA expression analysis suggested that the JEB was likely caused by reduced expression of *ITGB4* in the skin.

Immunohistochemical labeling of integrin  $\beta 4$  showed the absence of this protein in the skin of lambs with gross JEB lesions, whereas it was widely expressed in the basal cell layer in the control animal (Fig 4). The labeling of the two other epidermal basement membrane proteins that were studied, collagen VII and integrin  $\alpha 6$ , was similar between the control and the affected lambs. In the lambs showing gross JEB lesions, collagen VII was expressed primarily at the base of the blister, whereas integrin  $\alpha 6$  was found at the roof. This labeling pattern confirms the localization of the cleavage between the lamina lucida and the lamina densa of the basement membrane (Fig 4).

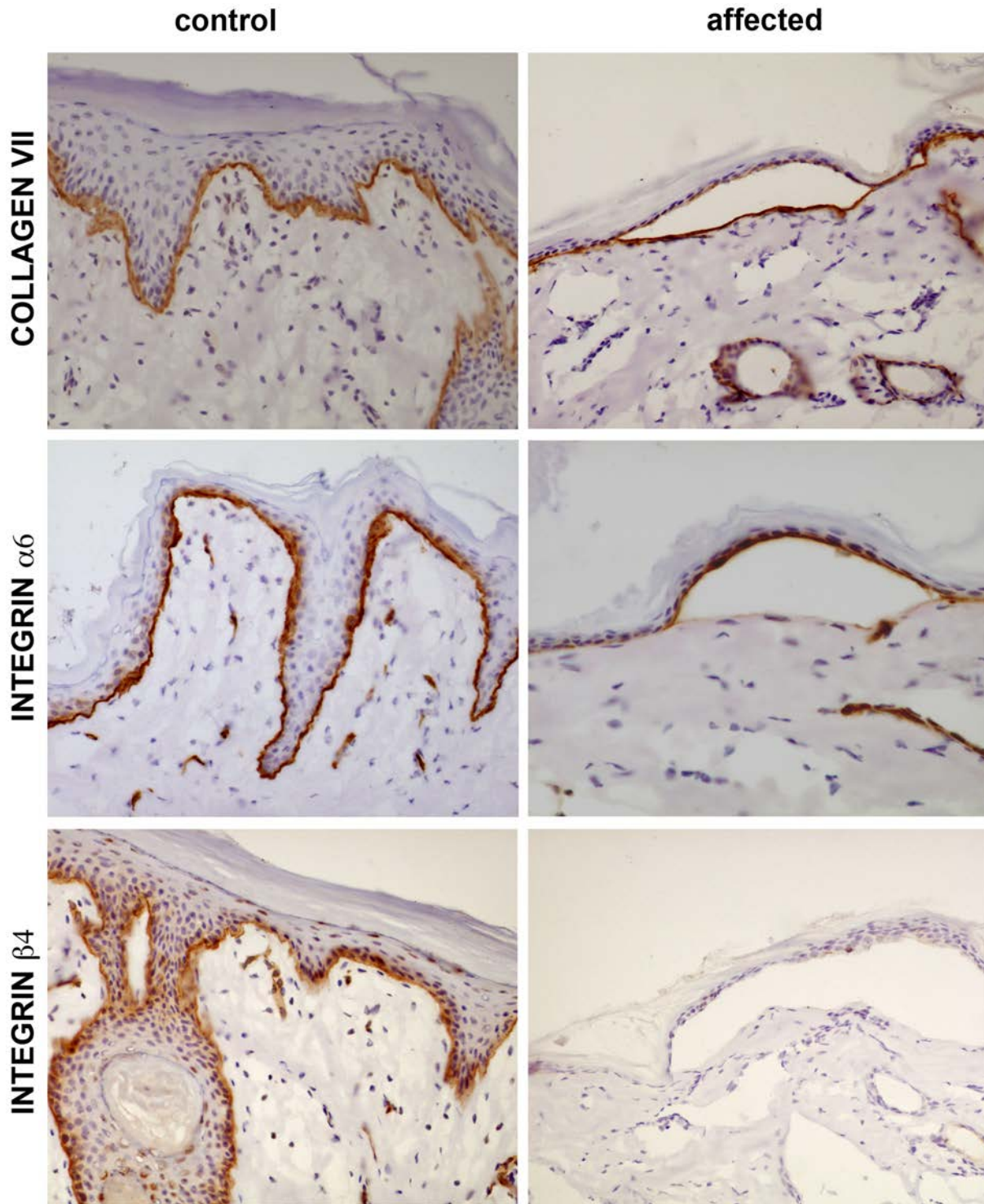
### Validation of c.4412\_4415del at the DNA level

After the putative causal mutation in the RNA-seq variant analysis was identified, the mutation was confirmed via analysis of 46 samples at the genomic DNA level. The association between the c.4412\_4415del allelic variant and the JEB phenotype was confirmed by genotyping the deletion in the gDNA sequences of affected, alleged carrier and control animals. From the DNA samples, the amplicon of the putative exon 33 was obtained via PCR for 27 cases, 10 unrelated controls and sires and dams of affected lambs. The resulting PCR product was sequenced and analyzed using SeqScape v2.5 (Applied Biosystems). All of the samples from the JEB-affected animals were homozygous for the identified 4-bp deletion (c.4412\_4415del/c.4412\_4415del or EB/EB), the sires and dams for later animals were heterozygous (EB/+) and the controls from the different flocks were homozygous (+/+) for the wild-type allele.

### Discussion

We herein demonstrated that a 4-bp deletion within exon 33 of *ITGB4* gene (c.4412\_4415del) is responsible for JEB in Churra sheep, based on medium-resolution mapping and RNA-seq variant analysis. The JEB locus was mapped to *Ovis aries* chromosome 11 via a GWA analysis using cases, related animals and unrelated controls. The region was fine-mapped to an 868-kb homozygous segment using IBD analysis. The *ITGB4* gene, which is located within this region, was the best positional and functional candidate gene. The RNA-seq variant analysis of control and JEB-affected skin RNA samples allowed us to find the causal mutation (c.4412\_4415del). In this study, we found that individuals that are homozygous for this mutation lack the integrin  $\beta 4$  polypeptide. The mutation was confirmed at the genomic level, and three genotypes segregated with the disease status: homozygous mutants (EB/EB) were the affected lambs, heterozygous animals (EB/+) corresponded to the phenotypically normal carriers (ewes and rams of the affected lambs), and the wild-type genotype (+/+) was identified in all of the analyzed control individuals. These genotypes completely agreed with the autosomal recessive mode of Mendelian inheritance inferred for JEB in Churra sheep. The discovery of this mutation has allowed screening of affected flocks to identify phenotypically normal individuals carrying the 4-bp deletion to avoid at-risk matings. Simple and rapid JEB genotyping methods like fragment length analysis in capillary electrophoresis sequencer and restriction fragment length polymorphism analysis are proposed (S2 Fig). The implementation of this gene-assisted selection approach will solve the problem of JEB in flocks of Churra sheep.

De novo BAC clone sequencing allowed us to obtain the complete *ITGB4* gene sequence (GenBank KP025765). The previous *ITGB4* ovine gene sequence (ENSOARG00000009764) was missing 16.6% of the bases when compared with KP025765 (S1 Fig). The RNA-seq approach enables the generation of extensive transcriptome information in the absence of a well-annotated genome. In the present study, the transcriptome information was essential in the



**Fig 4. Immunohistochemical labeling of collagen VII and integrins  $\alpha 6$  and  $\beta 4$ .** The three proteins are labeled (brown signal) in the skin samples of the control lamb, at the basal cell layer while in the samples from the affected lamb integrin  $\beta 4$ . The signal of the other two proteins is normal. Note the localization of collagen VII at the base of the blister in the sample from the affected lamb while the integrin  $\alpha 6$  remains attached to the basal keratinocytes, at the roof of the blister.

doi:10.1371/journal.pone.0126416.g004

annotation of the *ITGB4* gene sequence (GenBank: KP025765). The sheep *ITGB4* CDS obtained from the KP025765 sequence is distributed into 38 exons and encodes a predicted protein of 1751 amino acids. The transcript available in Ensembl (ENSOART00000010637) has 42 exons and encodes a predicted protein of 1835 residues. The percentage of identity of the ovine integrin  $\beta 4$  protein with that of other species, such as cow (DAA18168.1) or human (AAI43743.1), is higher for the integrin  $\beta 4$  protein obtained from KP025765 (92% and 89%, respectively) than for the sequence obtained from the ENSOART00000010637 transcript (86% and 82%, respectively), which suggests that the *ITGB4* sequence obtained in this study is more reliable than the available sequence from the ovine genome assembly (Oar\_v3.1).

Integrin  $\beta 4$  associates with integrin  $\alpha 6$  to form a transmembrane constituent of hemidesmosomes anchoring complex [3]. Mutations in  $\alpha 6\beta 4$  have been associated with JEB-PA in humans (OMIM: #226730). Although several mutations have been described in these two genes, most mutations were found in the *ITGB4* gene (approximately 70), whereas only five cases have been reported in the *ITGA6* gene [42]. In this study we found a causal mutation in the *ITGB4* gene which leads into JEB in sheep. Several forms of EB have been associated with congenital absence of skin, primarily with DEB (Bart type (OMIM: #132000)). Churra lambs lack PA; however, previous reports on humans suggest that PA is not necessarily a feature of integrin-associated JEB [42, 43]. A recent report about JEB in cattle due to a deletion in *ITGB4* also describes the absence of PA [13]. Lack of PA could be due to the particular anatomical characteristics of ruminant's pyloric sphincter, less muscular than in monogastric animals. The resemblance between the integrin  $\beta 4$ -associated EB lesions described in humans [18], cattle [13] and ovine JEB supports the causative nature of the 4-bp deletion in the *ITGB4* gene that was reported here in relation to the JEB phenotype observed in Churra sheep.

The 4-bp deletion on putative exon 33 of the *ITGB4* gene (*c.4412\_4415del*) results in a premature termination codon (PTC) and is predicted to delete the region spanning the last 281 amino acids in the intracellular domain (Fig 3), which has been shown to interact with the 180-kD bullous pemphigoid antigen [44]. This results in the loss of the C-terminal region, including the third and fourth FNIII repeats. The lack of labeling by the monoclonal antibody (clone 450-10D) used against integrin $\beta 4$  is consistent with this prediction, as this antibody binds to the intracellular region of the integrin [45]. The persistency of the extracellular domain could enable the formation of an  $\alpha 6\beta 4$  dimer. Also supporting this possibility, is the wide expression of integrin  $\alpha 6$ , which was observed via immunohistochemistry at its physiological location, the lamina lucida. However, the highly decreased amount of mRNA in the JEB skin samples compared with the control samples (identified through the qRT-PCR analysis) suggests that the truncated mRNA products are degraded by nonsense-mediated mRNA decay (NMD) machinery. Therefore, the absence of both the extracellular and intracellular domains, of integrin  $\beta 4$  cannot be ruled out, as deficiency of the whole protein would also produce a lack of labeling by the monoclonal antibody specific for the intracellular epitope. In humans, mutations related with PTC in both alleles are predominantly associated with lethal forms of the disease, whereas missense mutations are associated with mid-range phenotypes [46]. However, lethal/non-lethal consequences are primarily associated with the existence of a functional integrin  $\beta 4$ , which, is why the immunological labeling of the protein is also important in the diagnostic procedure [21]. Similar to the integrin  $\beta 4$  subunit, when transmission electron microscopy (TEM) is performed in the skin samples of JEB lambs we could see that hemidesmosomes were also absent. Schaapveld et al (1998) [47] indicated that keratinocytes lacking  $\beta 4$  expression are not capable of forming organized hemidesmosome-like structures. No hemidesmosome structures were found in *ITGB4* knock-out mouse [48], it was thus proposed that alternative adhesion molecules, such as integrin  $\alpha 3\beta 1$ , could explain the residual adhesion of the non-detached skin areas.



In conclusion, we have identified the causal mutation for ovine JEB in Churra sheep. This mutation (*c.4412\_4415del*) is located in putative exon 33 of the *ITGB4* gene and leads to the formation of a premature termination codon without protein expression. The *ITGB4* gene can be rapidly applied as a selection marker for identification of carriers to avoid at-risk matings, enabling the eradication of this lethal mutation from the Churra sheep breed. We have also demonstrated the successful combination of medium-density SNP panels with RNA-seq approaches for localization of mutant locus causing a recessive defect, which can serve as a reference for other studies performed in species in which high resolution exome sequencing is not straightforward. Furthermore, given the relatively low cost of maintaining the carriers, sheep with JEB could be used as a valuable large-animal model for the study of this disease and the development of therapeutic approaches.

## Material and Methods

### Ethics statement

The blood and tissue samples were collected from rams, ewes and lambs by qualified veterinarians following standard procedures and conducted under a license issued in accordance with the European Union legislation (European Community Directive, 86/609/E and Directive 2010/63/EU of the European Parliament and of the Council). All of the animals were managed in accordance with the guidelines for the accommodation and care of animals. The DNA samples used in this study were extracted from leucocytes in 3 ml of blood obtained via jugular venipuncture. The tissues were collected immediately after euthanasia (performed by a qualified veterinarian via intravenous injection of T-61 (Intervet)); animals were euthanized as standard of care. As the samples were from two commercial flocks that underwent veterinary examination, we were in a special situation in veterinary medicine and there was no “animal experiment” according to the legal definitions in Spain (Animal care legislation “Ley 32/2007”). According to the Ethics Commission of the University of León, formal ethical approval is not required under these circumstances.

### Animals

Blood and/or tissue samples were collected from a total of 27 JEB-affected lambs from two commercial flocks belonging to the Spanish Churra Sheep Breeders' Association (ANCHE). Additionally, samples from 28 healthy Churra individuals from the same flocks were collected. The samples included the parents, siblings and half-siblings of the JEB-affected offspring. Fifty samples from unaffected Churra sheep from different flocks with unknown relatedness to the affected lambs were collected for use as controls.

### Genome-wide association and homozygosity mapping studies

In total, 96 DNA samples were genotyped using the Illumina OvineSNP50 BeadChip. These samples included 48 unrelated, healthy Churra individuals from different flocks of the Churra Selection Nucleus and 48 animals from the affected flock. Twenty of the animals from the affected flock were JEB lambs, and the rest of the animals were related to them (sires, dams, siblings and half-siblings).

The results were analyzed using PLINK, v1.07 [49]. A quality control procedure was performed to eliminate SNPs with genotyping rates lower than 0.05 and minor allele frequencies lower than 0.01 and to eliminate animals with call rates lower than 0.95. After the QC pruning, 44,785 SNPs were considered in the subsequent analysis. The—assoc option in PLINK was

applied to perform a case-control genome-wide association study (GWAS). To correct empirical p-values, a one-million permutation procedure was performed.

After the GWAS was performed, a run of homozygosity or identity-by-descent (IBD) analysis was performed. The run of homozygosity is performed by filtering the cases to identify the allele-sharing regions. The SNPs flanking the consensus region obtained from this analysis were used to trace this region in the Ovine Genome Assembly Oar\_v3.1 Browser (<http://gbrowse-ext.bioinformatics.csiro.au/fgb2/gbrowse/oarv3/>). The genes identified within the homozygosity block were evaluated as putative positional candidate genes.

## Isolation of the ovine ITGB4 gene

*ITGB4* was isolated from three sheep CHORI-243 bacterial artificial chromosome (BAC) libraries [50]. The selected BAC clones predicted to span the region in which the *ITGB4* gene is located were CH243-344E15, CH243-306J7 and CH243-504K6. The BACMAX DNA Purification Kit (Epicentre) was used for purification of the DNA.

Ion Torrent PGM data were generated from the three BAC libraries constructed from 500 ng of DNA quantitated using the Qubit Fluorometer 2.0 (Life Technologies, Grand Island, NY). The fragment libraries were constructed using the Ion Xpress Plus Fragment Library kit based on Ion Shear chemistry according to the user guide. The barcode and adaptor ligation were performed per the manufacturer's protocol using the Ion Xpress Barcode Adapters 1–16 Kit. The quantitation and size distribution of the fragments were analyzed on an Agilent Bioanalyzer using the High Sensitivity Kit (Agilent). The template preparation and emulsion PCR were performed using the Ion PGM Template OT2 200 Kit (Life Technologies) according to the manufacturer's instructions. The template product then was quantitated using the Ion Sphere Quality Control Kit on at Qubit Fluorometer 2.0. OneTouch ES was used for the enrichment of the Ion Sphere Particles (ISP) template products. Finally, the samples were loaded onto a 316 chip V2 and sequenced using the Ion PGM 200 Sequencing Kit V2 (Life Technologies).

The reads were collected using Ion Torrent Suite v4.0, which also sorts the data according to the barcodes. The Torrent Suite software scores the quality of the reads by assigning Q20 scores according to Ion Torrent's quality scoring computation. MIRA (version 3.4.0) was used for de novo assembly of all of reads from each of the three BACs. The mean coverage was 120X for each BAC. The Torrent Mapping Alignment Program (TMAP) was used for alignment of all of the reads from each of the three BACs against the reference (Chr 19 Cow sequence; chromosome: UMD3.1:19:1:64057457:1). Both, the BAC DNA sequencing and the de novo assembly were performed at AC-gen Reading Life (Valladolid, Spain).

## RNA-seq analysis

For the RNA-seq analysis, skin samples from a control lamb and a JEB-affected lamb were collected for RNA extraction. The skin samples were harvested immediately after euthanasia and preserved in RNA Stabilization Reagent (RNAlater, Ambion). Slices of up to 500 mg of skin tissue were processed using the RNeasy Fibrous Tissue Midi Kit (Qiagen) to extract the total RNA. RNA integrity and concentration were evaluated using an Agilent Technologies 2100 Bioanalyzer. Both samples had RNA Integrity Number (RIN) values higher than 7.3. The TrueSeq RNA Sample Prep Kit v2 (Illumina) was used to prepare paired-end libraries with fragments of 300 bp. Two paired-end cDNA libraries were constructed, one for each of the two samples (JEB-affected and unaffected skin samples). The cDNA libraries were loaded onto the flow cell channels of an Illumina Hi-Seq 2000 platform to perform massive parallel sequencing

at the CNAG (Centro Nacional de Análisis Genómico, Barcelona), generating 75-bp paired-end reads.

The *Ovis aries* genome, the annotation file and the reference variant file were downloaded from the Ensembl database (<http://www.ensembl.org/index.html>). The quality of the reads generated was assessed using the program FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The paired ends for each sample were aligned to Oar v3.1 using STAR [36].

For the *ITGB4* gene prediction, Augustus software was used [37, 38]. Once the alignment to the reference genome was performed, Samtools [51] was used to sort, index and extract the IDs of the reads aligned within the *ITGB4* region from the bam files (OAR11: 54,830,000–54,880,000). Seqtk tool (<https://github.com/lh3/seqtk>) was used with the ID list from the bam file and the initial fastq file to create a new fastq file that contained only the reads that align within the *ITGB4* region. The new fastq files were mapped against the *ITGB4* sequence obtained from the BAC clones using STAR. Cufflinks [52] was used to assemble the bam files and to create an exonpart hints file for Augustus. The putative coding transcript for the *ITGB4* BAC clone sequence was identified with Augustus by providing the *ITGB4* sequence, exon boundaries and an integrin  $\beta 4$  protein profile and allowing the program to predict exactly one gene.

For variant detection, GATK was used [39]. The GATK analysis was performed on the STAR alignments following the RNA-seq workflow recommended in the best-practices section of the webpage for the GATK software [53, 54]. Prior to the variant analysis, the PCR duplicates from the bam files were removed using Picard tools (<http://picard.sourceforge.net/>). Then, an indel realignment and a base quality score recalibration were performed using the *Ovis aries* vcf file downloaded from Ensembl as a reference. These steps were followed by the variant calling analysis performed using the “-T HaplotypeCaller” option of GATK. The biological impact predictions for detected variants were obtained using Ensembl Variant Effect Predictor (VEP: <http://www.ensembl.org/tools.html>). Once the variant calling analysis was performed, the variants located in the *ITGB4* region were selected. These variants were filtered to keep only those located in exons homozygous for the reference sequence in the control and homozygous and distinct from the reference in the case.

### Quantification of the *ITGB4* mRNA via qRT-PCR

Total RNA was isolated from the skin using the RNeasy Fibrous Tissue Midi Kit (Qiagen) according to the manufacturer's instructions. The high-capacity cDNA Reverse Transcription Kit (Applied Biosystems) was used to synthesize the cDNA according to the manufacturer's instructions. The primers were designed using Primer3 [55, 56] resulting in two primer sets (S1 Table) spanning the region containing the 4-bp-deletion. One primer set produced a 334-bp product and the second produced a 182-bp product. The expression profiling of *ITGB4* was performed on a StepOne Real-Time PCR System (Applied Biosystems) with the Power SYBR Green PCR Master Mix according to the manufacturer's instructions. *ITGB4* transcript-specific expression was evaluated via relative quantification against the sheep glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) expression level ( $\Delta$ Ct) using one of the control homozygous replicates samples as a calibrator.

### Immunohistochemical studies

Samples of grossly normal skin (metacarpal region) were taken from three lambs showing gross JEB lesions. Control samples, taken from the same area, were also obtained from one lamb from the same flock that was euthanized because of diarrhoea and not affected by JEB.

The samples were snap-frozen immediately after euthanasia in 2-methylbutane cooled in liquid nitrogen and stored at  $-70^{\circ}\text{C}$  until used. Frozen sections ( $5\ \mu\text{m}$ ) were cut with a cryostat, collected on slides and fixed in acetone for 10 min. The samples underwent immunohistochemical examination with monoclonal antibodies against collagen VII, clone LH7.2, (100 dilution; Neomarkers), integrin  $\alpha 6$ , clone GoH3 (100 dilution; Abcam) and integrin  $\beta 4$ , clone 450-10D (2000 dilution; [45]) was carried out, using Dako EnVision System-HRP (Dako North America Inc.) with 3,3'-diaminobenzidine (Dako North America Inc.) as chromogen substrate.

## Confirmation and genotyping of the causal mutation

DNA was extracted from the blood samples collected from the 27 cases, the 6 mothers and 3 half-brothers of the affected animals, and the 10 unrelated controls. Primers (*ITGB4*ovine\_ex33fw: 5' - CCAGCAGTCAGGGAGGTG and *ITGB4*ovine\_ex33rv: 5' - CTCACCGCCGTT CAGCAG) were designed using Primer3 [55, 56]. The primers *ITGB4*ovine\_ex33fw and *ITGB4*ovine\_ex33rv amplify a 243-bp fragment of the gene sequence that contains the 4-bp deletion that was identified has the putative causal mutation. The products were amplified via PCR in a 30  $\mu\text{L}$  reaction volume containing 100 ng of gDNA, 480  $\mu\text{M}$  dNTPs, 1.5  $\mu\text{M}$   $\text{MgCl}_2$ , buffer 10X diluted 1X (Applied Biosystems), the specific primers at 0.5  $\mu\text{M}$ , 0.25 M betaine and 0.75  $\mu\text{l}$  of AmpliTaq Gold polymerase (5 U/ml) (Applied Biosystems). The PCR conditions were 5 min at  $96^{\circ}\text{C}$ ; 40 cycles of 30 sec at  $96^{\circ}\text{C}$ , 45 sec at  $58^{\circ}\text{C}$ , and 40 sec at  $72^{\circ}\text{C}$ ; and then a final extension of 10 min at  $72^{\circ}\text{C}$ . The amplicons were purified using Illustra Exonuclease I and alkaline phosphatase (Illustra ExoProStar 1-Step, GE Healthcare Life Sciences) treatment and were dideoxy-sequenced in both directions using the Big Dye Terminator Cycle Sequencing Kit, v3.1 (Applied Biosystems) using the same primers used for the fragment amplification. After sequencing on an ABI Prism 3130 automatic sequencer (Applied Biosystems), the sequence data were analyzed using SeqScape v2.5 (Applied Biosystems).

## Supporting Information

**S1 Fig. Ovine *ITGB4* genomic sequence.** Genomic sequences of sheep: ENSOARG0000009764 and KP025765 are aligned using the *EMBOSS Needle* software ([http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)).  
(PDF)

**S2 Fig. Examples of rapid genotyping methods for the detection of JEB carriers.** Examples for fluorescent fragment length analysis and RFLP used in carrier detection are displayed.  
(PDF)

**S1 Table. Primers used in quantification of the *ITGB4* mRNA via qRT-PCR.** The length of the amplified product, the melting temperature ( $T_m$ ) and position in Oar\_v3.1 are also indicated.  
(XLSX)

## Author Contributions

Conceived and designed the experiments: JJA. Performed the experiments: ASV JB VP BGG JJA. Analyzed the data: ASV BGG GTK CK. Contributed reagents/materials/analysis tools: SJK CK. Wrote the paper: ASV BGG JJA.



## References

1. Fine JD, Eady RA, Bauer EA, Bauer JW, Bruckner-Tuderman L, Heagerty A, et al. The classification of inherited epidermolysis bullosa (EB): Report of the third international consensus meeting on diagnosis and classification of EB. *J Am Acad Dermatol*. 2008; 58(6): 931–950. doi: [10.1016/j.jaad.2008.02.004](https://doi.org/10.1016/j.jaad.2008.02.004) PMID: [18374450](https://pubmed.ncbi.nlm.nih.gov/18374450/)
2. Fine JD, Bruckner-Tuderman L, Eady RA, Bauer EA, Bauer JW, Has C, et al. Inherited epidermolysis bullosa: Updated recommendations on diagnosis and classification. *J Am Acad Dermatol*. 2014; 70(6): 1103–1126. doi: [10.1016/j.jaad.2014.01.903](https://doi.org/10.1016/j.jaad.2014.01.903) PMID: [24690439](https://pubmed.ncbi.nlm.nih.gov/24690439/)
3. Borradori L, Sonnenberg A. Structure and function of hemidesmosomes: More than simple adhesion complexes. *J Invest Dermatol*. 1999; 112(4): 411–418. doi: [10.1046/j.1523-1747.1999.00546.x](https://doi.org/10.1046/j.1523-1747.1999.00546.x) PMID: [10201522](https://pubmed.ncbi.nlm.nih.gov/10201522/)
4. Uitto J, Pulkkinen L, McLean WH. Epidermolysis bullosa: A spectrum of clinical phenotypes explained by molecular heterogeneity. *Mol Med Today*. 1997; 3(10): 457–465. PMID: [9358473](https://pubmed.ncbi.nlm.nih.gov/9358473/)
5. Varki R, Sadowski S, Pfendner E, Uitto J. Epidermolysis bullosa. I. molecular genetics of the junctional and hemidesmosomal variants. *J Med Genet*. 2006; 43(8): 641–652. PMID: [16473856](https://pubmed.ncbi.nlm.nih.gov/16473856/)
6. Olivry T, Poujade-Delverdier A, Dunston SM, Fine J, Ortonne J. Absent expression of collagen XVII (BPAG2, BP180) in canine familial localized junctional epidermolysis bullosa. *Vet Dermatol*. 1997; 8(3): 203–212. doi: [10.1046/j.1365-3164.1997.d01-17.x](https://doi.org/10.1046/j.1365-3164.1997.d01-17.x)
7. Spirito F, Charlesworth A, Linder K, Ortonne JP, Baird J, Meneguzzi G. Animal models for skin blistering conditions: Absence of laminin 5 causes hereditary junctional mechanobullous disease in the belgian horse. *J Invest Dermatol*. 2002; 119(3): 684–691. PMID: [12230513](https://pubmed.ncbi.nlm.nih.gov/12230513/)
8. Milenkovic D, Chaffaux S, Taourit S, Guerin G. A mutation in the LAMC2 gene causes the herlitz junctional epidermolysis bullosa (H-JEB) in two french draft horse breeds. *Genet Sel Evol*. 2003; 35(2): 249–256. doi: [10.1051/gse:2003007](https://doi.org/10.1051/gse:2003007) PMID: [12633536](https://pubmed.ncbi.nlm.nih.gov/12633536/)
9. Alhaidari Z, Olivry T, Spadafora A, Thomas RC, Perrin C, Meneguzzi G, et al. Junctional epidermolysis bullosa in two domestic shorthair kittens. *Vet Dermatol*. 2005; 16(1): 69–73. PMID: [15725108](https://pubmed.ncbi.nlm.nih.gov/15725108/)
10. Capt A, Spirito F, Guaguere E, Spadafora A, Ortonne JP, Meneguzzi G. Inherited junctional epidermolysis bullosa in the german pointer: Establishment of a large animal model. *J Invest Dermatol*. 2005; 124(3): 530–535. PMID: [15737193](https://pubmed.ncbi.nlm.nih.gov/15737193/)
11. Graves KT, Henney PJ, Ennis RB. Partial deletion of the LAMA3 gene is responsible for hereditary junctional epidermolysis bullosa in the American saddlebred horse. *Anim Genet*. 2009; 40(1): 35–41. doi: [10.1111/j.1365-2052.2008.01795.x](https://doi.org/10.1111/j.1365-2052.2008.01795.x) PMID: [19016681](https://pubmed.ncbi.nlm.nih.gov/19016681/)
12. Medeiros GX, Riet-Correa F, Armien AG, Dantas AF, de Galiza GJ, Simoes SV. Junctional epidermolysis bullosa in a calf. *J Vet Diagn Invest*. 2012; 24(1): 231–234. doi: [10.1177/1040638711425953](https://doi.org/10.1177/1040638711425953) PMID: [22362960](https://pubmed.ncbi.nlm.nih.gov/22362960/)
13. Peters M, Reber I, Jagannathan V, Raddatz B, Wohlsein P, Drogemuller C. DNA-based diagnosis of rare diseases in veterinary medicine: A 4.4 kb deletion of ITGB4 is associated with epidermolysis bullosa in charolais cattle. *BMC Vet Res*. 2015; 11(1): 366–015–0366–0. doi: [10.1186/s12917-015-0366-0](https://doi.org/10.1186/s12917-015-0366-0)
14. Kerkmann A, Ganter M, Frase R, Ostmeier M, Hewicker-Trautwein M, Distl O. Epidermolysis bullosa in german black headed mutton sheep. *Berl Munch Tierarztl Wochenschr*. 2010; 123(9–10): 413–421. PMID: [21141276](https://pubmed.ncbi.nlm.nih.gov/21141276/)
15. Momke S, Kerkmann A, Wohlke A, Ostmeier M, Hewicker-Trautwein M, Ganter M, et al. A frameshift mutation within LAMC2 is responsible for herlitz type junctional epidermolysis bullosa (HJEB) in black headed mutton sheep. *PLoS One*. 2011; 6(5): e18943. doi: [10.1371/journal.pone.0018943](https://doi.org/10.1371/journal.pone.0018943) PMID: [21573221](https://pubmed.ncbi.nlm.nih.gov/21573221/)
16. Ostmeier M, Kerkmann A, Frase R, Ganter M, Distl O, Hewicker-Trautwein M. Inherited junctional epidermolysis bullosa (herlitz type) in german black-headed mutton sheep. *J Comp Pathol*. 2012; 146(4): 338–347. doi: [10.1016/j.jcpa.2011.08.008](https://doi.org/10.1016/j.jcpa.2011.08.008) PMID: [22000950](https://pubmed.ncbi.nlm.nih.gov/22000950/)
17. Carmi R, Sofer S, Karplus M, Ben-Yakar Y, Mahler D, Zirklin H, et al. Aplasia cutis congenita in two sibs discordant for pyloric atresia. *Am J Med Genet*. 1982; 11(3): 319–328. doi: [10.1002/ajmg.1320110308](https://doi.org/10.1002/ajmg.1320110308) PMID: [6177243](https://pubmed.ncbi.nlm.nih.gov/6177243/)
18. Vidal F, Aberdam D, Miquel C, Christiano AM, Pulkkinen L, Uitto J, et al. Integrin beta 4 mutations associated with junctional epidermolysis bullosa with pyloric atresia. *Nat Genet*. 1995; 10(2): 229–234. doi: [10.1038/ng0695-229](https://doi.org/10.1038/ng0695-229) PMID: [7545057](https://pubmed.ncbi.nlm.nih.gov/7545057/)
19. Ruzzi L, Gagnoux-Palacios L, Pinola M, Belli S, Meneguzzi G, D'Alessio M, et al. A homozygous mutation in the integrin alpha6 gene in junctional epidermolysis bullosa with pyloric atresia. *J Clin Invest*. 1997; 99(12): 2826–2831. doi: [10.1172/JCI119474](https://doi.org/10.1172/JCI119474) PMID: [9185503](https://pubmed.ncbi.nlm.nih.gov/9185503/)
20. Inoue M, Tamai K, Shimizu H, Owaribe K, Nakama T, Hashimoto T, et al. A homozygous missense mutation in the cytoplasmic tail of beta4 integrin, G931D, that disrupts hemidesmosome assembly and

- underlies non-herlitz junctional epidermolysis bullosa without pyloric atresia? *J Invest Dermatol.* 2000; 114(5): 1061–1064. PMID: [10792571](#)
21. Nakano A, Pulkkinen L, Murrell D, Rico J, Lucky AW, Garzon M, et al. Epidermolysis bullosa with congenital pyloric atresia: Novel mutations in the beta 4 integrin gene (ITGB4) and genotype/phenotype correlations. *Pediatr Res.* 2001; 49(5): 618–626. doi: [10.1203/00006450-200105000-00003](#) PMID: [11328943](#)
  22. Birnbaum RY, Landau D, Elbedour K, Ofir R, Birk OS, Carmi R. Deletion of the first pair of fibronectin type III repeats of the integrin beta-4 gene is associated with epidermolysis bullosa, pyloric atresia and aplasia cutis congenita in the original carmi syndrome patients. *Am J Med Genet A.* 2008; 146A(8): 1063–1066. doi: [10.1002/ajmg.a.31903](#) PMID: [18348258](#)
  23. Shariflou MR, Wade CM, Kijas J, McCulloch R, Windsor PA, Tammen I, et al. Brachygnathia, cardiomegaly and renal hypoplasia syndrome (BCRHS) in merino sheep maps to a 1.1-megabase region on ovine chromosome OAR2. *Anim Genet.* 2013; 44(2): 231–233. doi: [10.1111/j.1365-2052.2012.02392.x](#) PMID: [22762779](#)
  24. Suarez-Vega A, Gutierrez-Gil B, Cuchillo-Ibanez I, Saez-Valero J, Perez V, Garcia-Gamez E, et al. Identification of a 31-bp deletion in the RELN gene causing lissencephaly with cerebellar hypoplasia in sheep. *PLoS One.* 2013; 8(11): e81072. doi: [10.1371/journal.pone.0081072](#) PMID: [24260534](#)
  25. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* 2008; 92(5): 255–264. doi: [10.1016/j.ygeno.2008.07.001](#) PMID: [18703132](#)
  26. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods.* 2008; 5(7): 621–628. doi: [10.1038/nmeth.1226](#) PMID: [18516045](#)
  27. Wang Z, Gerstein M, Snyder M. RNA-seq: A revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009; 10(1): 57–63. doi: [10.1038/nrg2484](#) PMID: [19015660](#)
  28. Canovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano JF. SNP discovery in the bovine milk transcriptome using RNA-seq technology. *Mamm Genome.* 2010; 21(11–12): 592–598. doi: [10.1007/s00335-010-9297-z](#) PMID: [20886216](#)
  29. Cox LA, Glenn JP, Spradling KD, Nijland MJ, Garcia R, Nathanielsz PW, et al. A genome resource to address mechanisms of developmental programming: Determination of the fetal sheep heart transcriptome. *J Physiol.* 2012; 590(Pt12): 2873–2884. doi: [10.1113/jphysiol.2011.222398](#) PMID: [22508961](#)
  30. Djari A, Esquerre D, Weiss B, Martins F, Meersseman C, Boussaha M, et al. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics.* 2013; 14: 307–2164–14–307. doi: [10.1186/1471-2164-14-307](#) PMID: [24074255](#)
  31. Poulter JA, El-Sayed W, Shore RC, Kirkham J, Inglehearn CF, Mighell AJ. Whole-exome sequencing, without prior linkage, identifies a mutation in LAMB3 as a cause of dominant hypoplastic amelogenesis imperfecta. *Eur J Hum Genet.* 2014; 22(1): 132–135. doi: [10.1038/ejhg.2013.76](#) PMID: [23632796](#)
  32. Takeichi T, Liu L, Fong K, Ozoemena L, McMillan JR, Salam A, et al. Whole-exome sequencing improves mutation detection in a diagnostic epidermolysis bullosa laboratory. *Br J Dermatol.* 2015; 172(1): 94–100. doi: [10.1111/bjd.13190](#) PMID: [24947307](#)
  33. Stepp MA, Spurr-Michaud S, Tisdale A, Elwell J, Gipson IK. Alpha 6 beta 4 integrin heterodimer is a component of hemidesmosomes. *Proc Natl Acad Sci USA.* 1990; 87(22): 8970–8974. PMID: [2247472](#)
  34. Chavanas S, Gache Y, Vailly J, Kanitakis J, Pulkkinen L, Utito J, et al. Splicing modulation of integrin beta4 pre-mRNA carrying a branch point mutation underlies epidermolysis bullosa with pyloric atresia undergoing spontaneous amelioration with ageing. *Hum Mol Genet.* 1999; 8(11): 2097–2105. PMID: [10484780](#)
  35. Dalrymple BP, Kirkness EF, Nefedov M, McWilliam S, Ratnakumar A, Barris W, et al. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biol.* 2007; 8(7): R152. doi: [10.1186/gb-2007-8-7-r152](#) PMID: [17663790](#)
  36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29(1): 15–21. doi: [10.1093/bioinformatics/bts635](#) PMID: [23104886](#)
  37. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 2008; 24(5): 637–644. doi: [10.1093/bioinformatics/btn013](#) PMID: [18218656](#)
  38. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics.* 2011; 27(6): 757–763. doi: [10.1093/bioinformatics/btr010](#) PMID: [21216780](#)
  39. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20(9): 1297–1303. doi: [10.1101/gr.107524.110](#) PMID: [20644199](#)

40. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009; 4(7): 1073–1081. doi: [10.1038/nprot.2009.86](https://doi.org/10.1038/nprot.2009.86) PMID: [19561590](https://pubmed.ncbi.nlm.nih.gov/19561590/)
41. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7(4): 248–249. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248) PMID: [20354512](https://pubmed.ncbi.nlm.nih.gov/20354512/)
42. Schumann H, Kiritsi D, Pigors M, Hausser I, Kohlhasse J, Peters J, et al. Phenotypic spectrum of epidermolysis bullosa associated with alpha6beta4 integrin mutations. *Br J Dermatol.* 2013; 169(1): 115–124. doi: [10.1111/bjd.12317](https://doi.org/10.1111/bjd.12317) PMID: [23496044](https://pubmed.ncbi.nlm.nih.gov/23496044/)
43. Lee M, Chen Q, Wang H, Zhang J, Lin Z, Yang Y. ITGB4-associated junctional epidermolysis bullosa without pylori atresia but profound genito-urinary involvement. *Acta Derm Venereol.* 2014; 95(1): 112–3. doi: [10.2340/00015555-1888](https://doi.org/10.2340/00015555-1888)
44. Aho S, Uitto J. Direct interaction between the intracellular domains of bullous pemphigoid antigen 2 (BP180) and beta 4 integrin, hemidesmosomal components of basal keratinocytes. *Biochem Biophys Res Commun.* 1998; 243(3): 694–699. PMID: [9500991](https://pubmed.ncbi.nlm.nih.gov/9500991/)
45. Kennel SJ, Epler RG, Lankford TK, Foote LJ, Dickas V, Canamucio M, et al. Second generation monoclonal antibodies to the human integrin alpha 6 beta 4. *Hybridoma.* 1990; 9(3): 243–255. PMID: [2365382](https://pubmed.ncbi.nlm.nih.gov/2365382/)
46. Pulkkinen L, Rouan F, Bruckner-Tuderman L, Wallerstein R, Garzon M, Brown T, et al. Novel ITGB4 mutations in lethal and nonlethal variants of epidermolysis bullosa with pyloric atresia: Missense versus nonsense. *Am J Hum Genet.* 1998; 63(5): 1376–1387. PMID: [9792864](https://pubmed.ncbi.nlm.nih.gov/9792864/)
47. Schaapveld RQ, Borradori L, Geerts D, van Leusden MR, Kuikman I, Nievers MG, et al. Hemidesmosome formation is initiated by the beta4 integrin subunit, requires complex formation of beta4 and HD1/ plectin, and involves a direct interaction between beta4 and the bullous pemphigoid antigen 180. *J Cell Biol.* 1998; 142(1): 271–284. PMID: [9660880](https://pubmed.ncbi.nlm.nih.gov/9660880/)
48. Van der Neut R, Krimpenfort P, Calafat J, Niessen CM, Sonnenberg A. Epithelial detachment due to absence of hemidesmosomes in integrin beta 4 null mice. *Nat Genet.* 1996; 13(3): 366–369. doi: [10.1038/ng0796-366](https://doi.org/10.1038/ng0796-366) PMID: [8673140](https://pubmed.ncbi.nlm.nih.gov/8673140/)
49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3): 559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795) PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
50. Ratnakumar A, Kirkness EF, Dalrymple BP. Quality control of the sheep bacterial artificial chromosome library, CHORI-243. *BMC Res Notes.* 2010; 3: 334–0500–3–334. doi: [10.1186/1756-0500-3-334](https://doi.org/10.1186/1756-0500-3-334) PMID: [21144002](https://pubmed.ncbi.nlm.nih.gov/21144002/)
51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009; 25(16): 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
52. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics.* 2011; 27(17): 2325–2329. doi: [10.1093/bioinformatics/btr355](https://doi.org/10.1093/bioinformatics/btr355) PMID: [21697122](https://pubmed.ncbi.nlm.nih.gov/21697122/)
53. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43(5): 491–498. doi: [10.1038/ng.806](https://doi.org/10.1038/ng.806) PMID: [21478889](https://pubmed.ncbi.nlm.nih.gov/21478889/)
54. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastq data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013; 25: 1754–1760. doi: [10.1002/0471250953.bi1110s43](https://doi.org/10.1002/0471250953.bi1110s43)
55. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 2012; 40(15): e115. PMID: [22730293](https://pubmed.ncbi.nlm.nih.gov/22730293/)
56. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics.* 2007; 23(10): 1289–1291. PMID: [17379693](https://pubmed.ncbi.nlm.nih.gov/17379693/)



**Procedimiento y Kit de Diagnóstico de Epidermólisis Bullosa Juntural en  
Ganado Ovino.**

Aroa Suárez-Vega, Beatriz Gutiérrez-Gil y Juan José Arranz.

Dpto. Producción Animal, Universidad de León, 24071, León, Spain.

Solicitud de Patente Nacional en España Nº 201431635 (2015).



# **PROCEDIMIENTO Y KIT DE DIAGNÓSTICO DE EPIDERMÓLISIS BULLOSA JUNTURAL EN GANADO OVINO**

## **DESCRIPCIÓN**

### **CAMPO DE LA INVENCIÓN**

El campo de la invención es el diagnóstico de enfermedades genéticas de herencia mendeliana simple (monogénica) en animales domésticos. En particular, la presente invención se refiere a la detección de la mutación responsable de la enfermedad genética denominada epidermólisis bullosa juntural o de la unión en el ganado ovino y al método de determinación del estatus genético de un individuo (sano, portador, enfermo) en relación con esta enfermedad.

### **ANTECEDENTES DE LA INVENCIÓN**

El término epidermólisis bullosa hereditaria incluye una serie de procesos de base genética cuyo rasgo común es una marcada fragilidad de la piel y mucosas que desencadena la facilidad para desarrollar ampollas y úlceras por trauma o roce.

Existen diversos tipos de epidermólisis bullosa siendo la clasificación más aceptada la que distingue tres tipos principales: (i) epidermólisis bullosa simple, donde se producen ampollas intraepidérmicas que curan sin cicatriz, (ii) epidermólisis bullosa de la unión o juntural en la que las ampollas se deben a una formación anormal de los componentes de los hemidesmosomas, lo que da lugar a roturas a nivel de la lámina lúcida, y (iii) epidermólisis bullosa distrófica, que se produce a nivel de la lámina basal epidérmica, y en la que las ampollas subepidérmicas se reparan con cicatrices distróficas.

En el ganado ovino de raza *Black Headed Mutton* (BHM) se ha descrito recientemente una mutación en el gen *LAMC2*, que codifica para la laminina subunidad gamma-2 (Mömke S. et al., 2011. A frameshift mutation within *LAMC2* is responsible for Herlitz type junctional epidermolysis bullosa (HJEB) in black headed mutton sheep. *PLoS One* 6(5):e18943). Sin embargo, existen otros genes que pueden causar la EBJ como es el caso del *LAMA3*, *LAMB3*, *COL17A1*, *ITGB4* e *ITGA6* (Siañez-González C. et al., 2009. Epidermólisis ampollosa congénita: revisión del tema. *Actas Dermosifiliográficas* 100(10):842-56).

El procedimiento diagnóstico de la EBJ descrito en la presente invención se basa en la identificación de una mutación asociada a la enfermedad EBJ en un fragmento del gen *ITGB4*, que codifica para la proteína Integrina-beta 4. La Integrina-beta 4 es una proteína expresada principalmente en la membrana basal epitelial formando parte de la placa externa de los hemidesmosomas. Esta proteína contribuye a la estabilidad de la unión de la epidermis con la membrana basal subyacente.

Según la base de datos *Online Mendelian Inheritance in Men* (OMIM), las mutaciones en el gen *ITGB4* se relacionan principalmente con tres fenotipos: Epidermolisis bullosa de manos y pies (OMIM: 131800), epidermolisis bullosa juntural, tipo non-Herlitz (OMIM:226650) y epidermolisis bullosa juntural con atresia pilórica (OMIM:226730).

El gen *ITGB4* está situado en el cromosoma ovino 11 (OAR11), en la región 54.848.050-54.874.496 pb según la versión Oarv3.1 del genoma ovino ([http://www.ensembl.org/Ovis\\_aries/Info/Index](http://www.ensembl.org/Ovis_aries/Info/Index)), y está organizado en un total de 38 exones (GenBank Acc. no.: KP025765).

#### DESCRIPCIÓN DE LA INVENCION

Una realización es un procedimiento de diagnóstico de epidermolisis bullosa juntural en ganado ovino que comprende la detección de la presencia o ausencia de un fragmento de ADN del gen *ITGB4* ovino, identificado por la secuencia de nucleótidos ccac, en una muestra biológica obtenida de ganado ovino, en adelante procedimiento de la invención.

El fragmento de ADN identificado por la secuencia de nucleótidos ccac corresponde a una delección de 4 pb entre las posiciones 4412 y 4415 de la secuencia codificante del ARNm del gen *ITGB4* ovino disponible en GenBank (Acc. no. KP025765) correspondiente a las posiciones 54849767 y 54849770 pb en la secuencia de ADN del genoma ovino (Oarv3.1), en el exón 33 (posiciones 30204 a 30207), según la anotación del gen disponible en GenBank (Acc. no. KP025765). Esta delección se ha identificado por primera vez en la presente invención como responsable de la epidermolisis bullosa juntural en oveja, mediante un estudio comparado de la secuencia codificante de este gen en animales enfermos y sanos.

El procedimiento de la invención permite detectar individuos sanos portadores de la enfermedad. Dichos portadores serán heterocigotos para la delección de 4 pb localizada



en el exón 33 del gen, es decir tendrán una copia del gen completa y otra copia portadora de la delección.

5 Otra realización es el procedimiento de la invención, donde dicha detección se realiza con un biosensor o microarray.

Otra realización es el procedimiento de la invención, que comprende:

- 10 (a) amplificar por PCR en dicha muestra biológica un fragmento de ADN situado entre los cebadores con una identidad de al menos 80% respecto a las secuencias SEQ ID NO: 1 y 2,  
(b) determinar el número y tamaño de los fragmentos de ADN resultantes de dicha amplificación y  
(c) realizar un diagnóstico de epidermólisis bullosa juntural a partir de la información obtenida en el paso (b).

15 En la presente solicitud, dicho porcentaje de identidad en una secuencia determinada se calcula teniendo en cuenta que un 80% de identidad significa que un 80% de residuos de la secuencia completa de los cebadores identificados por las secuencias SEQ ID NO: 1 y 2 son idénticos a los residuos de la secuencia determinada.

20 Otra realización es el procedimiento de la invención, donde en la etapa (b) el número y tamaño de los fragmentos de ADN se determina por electroforesis.

Otra realización es el procedimiento de la invención, en el que antes de la etapa (a) se realiza una extracción de ADN de dicha muestra biológica.

25 La extracción del ADN a partir de estas muestras se puede realizar por métodos como el fenol-cloroformo descrito en Sambrook, 1989 (Sambrook, J. et. al., 1989, Molecular Cloning: A Laboratory Manual. 2nd edition. New York: Cold Spring Harbor Laboratory Press); por el método "Salting out" descrito en Miller, 1988 (MILLER, S.A. et al, 1988. A simple salting out  
30 procedure for extracting DNA from human nucleated cells. *Nucleic acids research*, 16(3), pp. 1215) en el caso del semen o la utilización de Chelex-100 (BioRad Laboratories, Hercules, CA) para la extracción a partir de muestras de pelo. La extracción del ADN también se puede realizar utilizando Kits comerciales de extracción de ADN, sobre todo para muestras de saliva o de muestras de flujo nasal en las que se parte de una escasa cantidad de células  
35 nucleadas.

Otra realización es el procedimiento de la invención, donde dicha identidad de secuencia es de al menos 90%.

5 Otra realización es el procedimiento de la invención, donde dichos cebadores están identificados por las secuencias SEQ ID NO: 2 y SEQ ID NO: 3.

10 Otra realización es el procedimiento de la invención, donde al menos uno de dichos cebadores está unido a una molécula marcadora. En una realización particular, dicha molécula marcadora está seleccionada del grupo compuesto por 6-carboxi-fluoresceína, hexacloro-fluoresceína y tetracloro-fluoresceína. Estas moléculas marcadores podrían ser reemplazadas por otras moléculas marcadoras.

15 Otra realización es el procedimiento de la invención, donde dicha muestra biológica está seleccionada del grupo compuesto por pelo, tejido, célula o fluido biológico.

Otra realización es el procedimiento de la invención, donde dicho fluido biológico está seleccionado del grupo compuesto por saliva, leche, suero, plasma sanguíneo, sangre, semen y fluido nasal.

20 Otra realización es un kit para realizar el procedimiento de la invención, que comprende los cebadores identificados por las secuencias SEQ ID NO: 1 y SEQ ID NO: 2.

25 El procedimiento de la invención constituye un método sencillo, rápido y económico de detección de animales portadores, evitando que animales heterocigotos puedan seleccionarse para reposición en el rebaño y ser incluidos en el programa de mejora correspondiente, hechos que favorecerían la diseminación de la enfermedad.

30 El procedimiento de la invención puede utilizarse también para la detección de animales portadores en el caso de cruzamientos programados para obtener animales que sirvan como modelo animal en el estudio de la enfermedad.

35 El procedimiento de la invención puede aplicarse a corderos que presenten lesiones cutáneas compatibles con la enfermedad con el fin de descartarla sin necesidad de realizar una necropsia.

TEXTO LIBRE DE LA LISTA DE SECUENCIAS

A continuación se aporta una traducción del texto libre en inglés que aparece en la lista de secuencias.

- 5 SEQ ID NO: 1. Cebador sentido para PCR.  
SEQ ID NO: 2. Cebador antisentido para PCR.  
SEQ ID NO: 3. Fragmento del gen salvaje ovino *ITGB4*.  
89..238 Exón 33  
152..155 Fragmento de delección
- 10 SEQ ID NO: 4. Fragmento del gen ovino *ITGB4* resultante de la delección.  
89..234 Exón 33

#### BREVE DESCRIPCIÓN DE LAS FIGURAS

15 Figura 1. Se muestran los resultados para 4 muestras procesadas mediante una electroforesis en gel de poliacrilamida en un secuenciador semiautomático y analizadas posteriormente con un programa de análisis de fragmentos. De arriba abajo se muestran los resultados del análisis para cuatro muestras.

20 En la primera de las muestras se obtuvo un único pico de 239 pb, que indica genotipo homocigoto para la delección de 4 pb (del/del), únicamente identificable en individuos afectados por la enfermedad.

En la segunda y cuarta muestra se identificaron dos picos de 243 y 239 pb, que indican  
25 genotipo heterocigoto para la delección (+/del), correspondiente a animales sanos pero portadores.

En la tercera muestra se observó un único pico con un tamaño estimado de 243 pb, que indica genotipo homocigoto para el alelo salvaje (+/+), y por tanto animal sano y no portador.

30

#### MODOS DE REALIZACIÓN PREFERENTE

Ejemplo 1. Diagnóstico genético de animales afectados por epidermólisis bullosa juntural y animales portadores de dicha enfermedad

35

En varias explotaciones ganaderas nacieron corderos con lesiones cutáneas, erosiones y úlceras en piel así como en las membranas mucosas. Los corderos afectados fueron eutanasiados debido al deterioro de su condición. La microscopía electrónica mostró que las separaciones de la unión dermoepidérmica se localizaron en la lámina lúcida de la membrana basal.

Se realizaron análisis anatomopatológicos correspondientes y se descartó la posible afección por otro tipo de enfermedades vesiculosas y ampollosas de la piel tales como los otros subtipos de epidermólisis bullosa (simple y distrófica), infecciones cutáneas por estafilococos, pénfigo bulloso etc. A continuación se realizó un análisis del pedigrí de los animales afectados mediante marcadores microsatélite siguiendo el protocolo descrito en GLOWATZKI-MULLIS, M.L. et al, 2007. Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. *Animal Genetics*, 38(1), pp. 86-88. Tanto los análisis anatomopatológicos como los genéticos concluyeron que los animales sufrían epidermólisis bullosa juntural.

#### Extracción del ADN a partir de la muestra biológica obtenida del animal en estudio.

En el presente ejemplo se utilizaron 4 de los corderos enfermos, las madres de los mismos (clasificadas como portadoras) y 30 machos de los que no se conocía el estatus genético. La extracción de DNA se realizó a partir de sangre por el método de "Salting out" descrito en Miller, 1988.

#### Diseño de los cebadores

Los cebadores se diseñaron específicamente con el programa *Primer3* (<http://frodo.wi.mit.edu/primer3>), a partir de la secuencia del genoma ovino v3.1. Así se partió de una región de 243 pb que incluía el Exon 33, 88 bp de la región intrónica anterior y 5 pb de la región intrónica posterior al mismo.

Se seleccionó una pareja de cebadores, cebador sentido, identificado con la secuencia SEQ ID NO: 1 y cebador antisentido, identificado con la secuencia SEQ ID NO: 2.

El tamaño del fragmento amplificado con esta pareja de cebadores fue 243 bp para la secuencia salvaje o no mutada, dicho fragmento incluye la secuencia completa del exón 33 (150 pb), 88 bp de la región intrónica anterior y 5 pb de la región intrónica

posterior al mismo. En el caso de la presencia de la delección de 4 pb el tamaño del amplicón fue de 239 pb.

5 Uno de los cebadores, en concreto el cebador antisentido SEQ ID NO: 2 se marcó con fluorocromo 6-fam; la presencia de un cebador marcado nos permitirá determinar la longitud de los fragmentos para así identificar los distintos genotipos (homocigoto normal, homocigoto para la delección y heterocigoto) mediante fluorescencia en el gel de poliacrilamida.

## 10 Amplificación del ADN

Se realizó la amplificación del fragmento de ADN mediante la reacción en cadena de la polimerasa (PCR). En este estudio, las reacciones de PCR estuvieron constituidas por un 80% de lo que denominamos “mezcla”, mientras que otro 20% fue de ADN total (10-30  
15 ng/ml). La “mezcla” estuvo constituida por un 1X de tampón 10X (*Gold Applied Biosystem, AB*), MgCl<sub>2</sub> (de 1,5 µM a 2,5 µM), dNTPs (480 µM), los cebadores diseñados específicamente para la amplificación del fragmento de ADN de interés (0,5 µM), betaína (0,25 M), dH<sub>2</sub>O y la encima polimerasa (*AmpliTaq Gold™ de AB*) (5 U/ml).

20 La amplificación se llevó a cabo en un termociclador automático (GeneAmp® PCR System 9700, AB). Los ciclos de temperatura siempre se vieron precedidos de un paso de desnaturalización a 96°C durante 5 minutos. Posteriormente se realizaron 40 ciclos, estando cada uno compuesto de 3 pasos:

- 25
- Desnaturalización del ADN: 96°C, 30s.
  - Hibridación de los cebadores con el ADN: 58°C, 45 s.
  - Síntesis de ADN: 72°C, 40 s.

El último ciclo finalizó en todos los casos con una extensión final de 10 minutos a 72°C.

30

Separación y visualización del fragmento o fragmentos amplificados por electroforesis en gel de poliacrilamida en un secuenciador semiautomático.

Una vez que se amplificaron los fragmentos mediante la técnica de PCR las muestras fueron  
35 preparadas para ser introducidas en el secuenciador. Para ello se cogieron 1.5 µl del producto de la PCR y este fue diluido en 20 µl de H<sub>2</sub>O mQ. Una vez realizada la dilución se

añaden 1.5  $\mu$ l de la misma a una mezcla de 10  $\mu$ l de formamida y 0.75  $\mu$ l del marcador estándar de tamaño.

5 Para la detección de los fragmentos, 12  $\mu$ l de la mezcla anterior se cargaron en un secuenciador semiautomático ABI PRISM 3130 (Applied BioSystems). Posteriormente los fragmentos fueron analizados mediante un programa de análisis de fragmentos.

Los resultados obtenidos permitieron agrupar a las muestras en los siguientes grupos:

- 10 - Un único pico de 243 pb, que indica genotipo homocigoto para el alelo salvaje (+/+), y por tanto animales sanos y no portadores.
- Un único pico con un tamaño estimado de 239 pb, que indica genotipo homocigoto para la delección de 4 pb (del/del), únicamente identificable en individuos afectados por la enfermedad.
- 15 - Dos picos de 243 y 239 pb, que indican genotipo heterocigoto para la delección (+/del), correspondiente a animales sanos pero portadores.

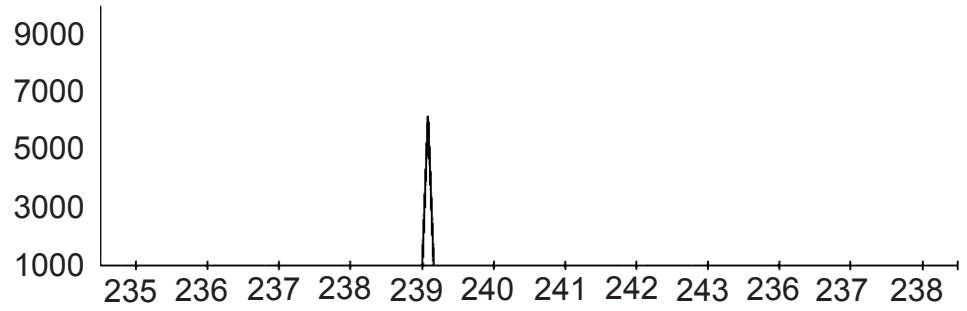
## REIVINDICACIONES

1. Procedimiento de diagnóstico de epidermólisis bullosa juntural en ganado ovino que comprende la detección de la presencia o ausencia de un fragmento de ADN del gen *ITGB4* ovino, identificado por la secuencia de nucleótidos *ccac*, en una muestra biológica obtenida de ganado ovino.
2. Procedimiento según la reivindicación 1, caracterizado por que comprende:
  - (a) amplificar por PCR en dicha muestra biológica un fragmento de ADN situado entre los cebadores con una identidad de al menos 80% respecto a las secuencias SEQ ID NO: 1 y 2,
  - (b) determinar el número y tamaño de los fragmentos de ADN resultantes de dicha amplificación y
  - (c) realizar un diagnóstico de epidermólisis bullosa juntural a partir de la información obtenida en el paso (b).
3. Procedimiento según la reivindicación 2, caracterizado por que antes de la etapa (a) se realiza una extracción de ADN de dicha muestra biológica.
4. Procedimiento según la reivindicación 2, caracterizado por que dicha identidad de secuencia es de al menos 90%.
5. Procedimiento según la reivindicación 2, caracterizado por que dichos cebadores están identificados por las secuencias SEQ ID NO: 1 y SEQ ID NO: 2.
6. Procedimiento según la reivindicación 2, caracterizado por que al menos uno de dichos cebadores está unido a una molécula marcadora.
7. Procedimiento según la reivindicación 6, caracterizado por que dicha molécula marcadora está seleccionada del grupo compuesto por 6-carboxi-fluoresceína, hexacloro-fluoresceína y tetracloro-fluoresceína.
8. Procedimiento según la reivindicación 1, caracterizado por que dicha detección se realiza con un biosensor o microarray.
9. Procedimiento según la reivindicación 2, caracterizado por que en la etapa (b) el número y tamaño de los fragmentos de ADN se determina por electroforesis.
10. Procedimiento según cualquiera de las reivindicaciones 1 a 9, caracterizado por que dicha muestra biológica está seleccionada del grupo compuesto por pelo, tejido, célula o fluido biológico.
11. Procedimiento según la reivindicación 10, caracterizado por que dicho fluido biológico está seleccionado del grupo compuesto por saliva, leche, suero, plasma sanguíneo, sangre, semen y fluido nasal.

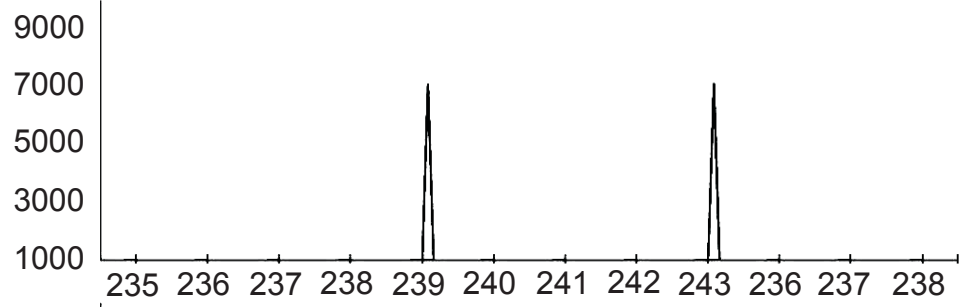
12. Kit para realizar el procedimiento según la reivindicación 2, caracterizado por que comprende los cebadores identificados por las secuencias SEQ ID NO: 1 y SEQ ID NO: 2.



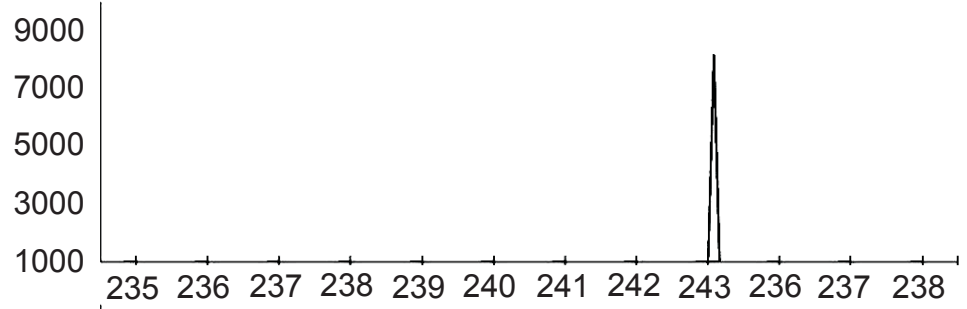
del/del



+/del



+/+



+/del

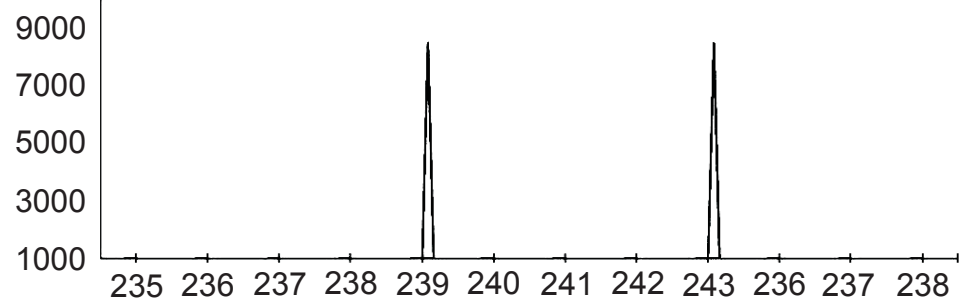


Fig. 1

## **RESUMEN**

Procedimiento y kit de diagnóstico de epidermólisis bullosa juntural en ganado ovino. Procedimiento de diagnóstico de epidermólisis bullosa juntural en ganado ovino que comprende la detección de la presencia o ausencia de un fragmento de ADN del gen ITGB4 ovino, identificado por la secuencia de nucleótidos ccac, en una muestra biológica obtenida de ganado ovino. Kit para realizar este procedimiento, que comprende los cebadores identificados por las secuencias SEQ ID NO: 1 y SEQ ID NO: 2.

2014-4727 seq listing  
SEQUENCE LISTING

<110> UNIVERSIDAD DE LEÓN

<120> PROCEDIMIENTO Y KIT DE DIAGNÓSTICO DE EPIDERMÓLISIS BULLOSA JUNTURAL EN GANADO OVINO

<160> 4

<170> BISSAP 1.2

<210> 1

<211> 18

<212> DNA

<213> Artificial sequence

<220>

<221> source

<222> 1..18

<223> /mol\_type="unassigned DNA"  
/note="Sense PCR primer"  
/organism="Artificial sequence"

<400> 1

ccagcagtca gggaggtg

18

<210> 2

<211> 18

<212> DNA

<213> Artificial sequence

<220>

<221> source

<222> 1..18

<223> /mol\_type="unassigned DNA"  
/note="Antisense PCR primer"  
/organism="Artificial sequence"

<400> 2

ctcaccgccg ttcagcag

18

<210> 3

<211> 243

<212> DNA

<213> Artificial sequence

<220>

<221> source

<222> 1..243

<223> /mol\_type="unassigned DNA"  
/note="Wild type ovine ITGB4 gene fragment"  
/organism="Artificial sequence"

<220>

<221> exon

<222> 89..238

<223> /note="Exon 33"

<220>

<221> misc\_feature

<222> 152..155

<223> /note="Deletion fragment"

<400> 3

ccagcagtca gggaggtgcg gctgagaacc aggactcct gccagacac ccccgctggg

60

2014-4727 seq listing

ggctgactgt ggcctcccc atctgcaggc tcccgctgg ctgcgggcgt gcccaacacg 120  
cccacccgcc tgggtttctc cgccctcgga cccacgtctc tgaaagtgag ctggcaggag 180  
ccgcagtgtg agcgtgcgct gcagggctac agtgtggagt accagctgct gaacggcggt 240  
gag 243

<210> 4  
<211> 239  
<212> DNA  
<213> Artificial Sequence

<220>  
<221> source  
<222> 1..239  
<223> /mol\_type="unassigned DNA"  
/note="Ovine ITGB4 gene fragment resulting from deletion"  
/organism="Artificial Sequence"

<220>  
<221> exon  
<222> 89..234  
<223> /note="Exon 33"

<400> 4  
ccagcagtca gggaggtgcg gctgagaacc agggactcct gccagacac ccccgctggg 60  
ggctgactgt ggcctcccc atctgcaggc tcccgctgg ctgcgggcgt gcccaacacg 120  
cccacccgcc tgggtttctc cgccctcgga cgtctctgaa agtgagctgg caggagccgc 180  
agtgtagcg tgcgctgcag ggctacagtg tggagtacca gctgctgaac ggcggtgag 239

**Initial RNA-seq analysis on the milk transcriptional profiling of two sheep breeds.**

Aroa Suárez-Vega<sup>1</sup>, Beatriz Gutiérrez-Gil<sup>1</sup>, Cristina Esteban<sup>2</sup> and Juan José Arranz<sup>1</sup>.

<sup>1</sup>Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain; <sup>2</sup>Fundación Centro de Supercomputación de Castilla y León, Edif. Crai-Tic, Campus de Vegazana, s/n, León 24071 , Spain.

Proceedings, 10th World Congress on Genetics Applied to Livestock Production (WCGALP). Vancouver, Canada. 17-22, Agosto, 2014.



## Initial RNA-seq analysis on the milk transcriptional profiling of two sheep breeds

A. Suárez-Vega<sup>1</sup>, B. Gutiérrez-Gil<sup>1</sup>, C. Esteban<sup>2</sup> and J.J. Arranz<sup>1</sup>.

<sup>1</sup>University of León, <sup>2</sup>Supercomputing Center of Castile and León, León, Spain.

**ABSTRACT:** In this study, massively parallel sequencing of transcripts (RNA-Seq) was used to make a preliminary characterization of the milk transcriptome of two dairy sheep breeds, Churra and Assaf, at day 120 of lactation. A total of 230 million paired-end reads were obtained from RNA sequencing. The 91% of these reads were mapped to a unique location in the ovine genome. Genes encoding caseins and whey protein showed the highest expression in this stage of lactation. Eight genes were identified as differentially expressed between the two dairy sheep breeds under study. A relationship with milk composition and lactation maintenance was established for some of the differentially expressed genes found.

Keywords:

Sheep

Milk production

RNA-seq

### Introduction

Massively parallel DNA sequencing, widely referred as "next-generation sequencing" (NGS), is changing the biological research view by enabling the comprehensive and relatively inexpensive analysis of genomes and transcriptomes (Metzker (2010)). RNA-seq technologies have emerged proving a unique opportunity to quantify transcripts and identify differential regulation in a single experiment (Wang et al. (2009)).

Sheep production can lead many potential outputs (milk, meat, skin, fiber). Sheep's milk is commonly used to produce many cultured dairy products. Dairy sheep industry is mainly concentrated in Europe and the countries on or near the Mediterranean Sea. The milk transcriptome profile has been analyzed in different mammalian species (Shu et al. (2012), Wickramasinghe et al. (2012), Lemay et al. (2013)) in order to better characterize the genes involved in lactation, although an specific study on the ovine milk transcriptome has yet not been reported.

The aim of this research was performing a preliminary analysis of the milk somatic cell transcriptome profile in two dairy sheep breeds (Spanish Churra and Assaf) at day 120 of lactation. Churra is an autochthonous dairy breed reared in the northwest of Spain and Assaf is a specialized dairy breed, with a higher milk production (De la Fuente et al. (2006)). The stage of lactation analyzed in this study is interesting from the point of view of the different lactation length in these two dairy sheep breeds, normalized to 120 and 150 lactation days in Churra and Assaf, respectively. For Assaf ewes, day 120 of lactation corresponds to a transition stage from peak to final lactation while that time-point is related to nearly final lactation in Spanish Churra sheep.

### Materials and Methods

**Sample collection, RNA isolation and sequencing.** The milk samples were obtained from three Churra sheep and three Assaf sheep at the 120 lactation day. For each animal we collected 50 ml of fresh milk, from both mammary glands. Somatic cells were separated as described by Wickramasinghe et al. (2012) with some modifications. Cells were pelleted by centrifugation at 1800 rpm in 50 mL sterile tubes for 10 min. at 4°C in the presence of a final concentration of 0.5 mM EDTA. The cell pellet was washed in PBS twice, pH 7.2 with 0.5 mM EDTA and suspended in 500 µl of Trizol. RNA extraction continued following the Trizol protocol (Invitrogen, Carlsbad, CA, USA).

The integrity of the RNA was assessed using Agilent 2100 Bioanalyzer device (Agilent technologies, Santa Clara, CA, USA). The RNA integrity value (RIN) of the samples ranged between 8.2 and 8.9. Paired-end libraries with fragments of 300 bp were prepared using the True-Seq RNA-Seq sample preparation Kit v2 (Illumina, San Diego, CA, USA). The fragments were sequenced on an Illumina Hi-Seq 2000 sequencer (Fasteris SA, Plan-les-Ouates, Switzerland).

**Mapping and assembly.** The quality of the raw RNA-seq reads was assessed by using FastQC ([www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc)). The quality control of the studied samples was later performed using the option HEADCROP of Trimmomatic ([www.usadellab.org/cms/?page=trimmomatic](http://www.usadellab.org/cms/?page=trimmomatic)). In this step we aimed to remove the first 10 bps of the paired-end reads in order to eliminate the bias induced by the Illumina hexamer priming in the nucleotide composition at the beginning of the reads. The last bases of the reads which a 'per base sequence quality' lower than 20 were also trimmed.

After the quality control, the paired-end reads were mapped against the ovine reference genome (OARv3.1) using TopHat v.2.0.10 (Trapnell et al. (2009)) with Bowtie2 (v2.2.0) applying the default settings. In addition, we used the -G option followed by the ENSEMBL transcriptome annotation, which allows the program to do a first alignment of the reads to the virtual sheep transcriptome. We also set the distance between both pairs to 50 bp (inner-mean distance) and the standard deviation to 150 bp. SAMtools (Li et al. (2009)) was used to sort the mapped reads and to filter them in order to maintain only those that map to a unique genomic position (MAPQ = 50). The Cufflinks v1.3.0 8 Reference Annotation Based Transcript (RABT) assembly method (Trapnell et al. (2012)) was used to construct, identify and estimate the expression of both

known and novel transcript fragments (transfrags) from TopHat alignment results with the `-g` option to supply a reference annotation (GFF) to guide the RABT assembly.

**Expression quantification and differential expression analysis.** The estimation of the number of reads mapping to each transcript and the differential expression analysis was done with the Cufflinks package (Cufflinks, Cuffmerge, Cuffcompare, Cuffdiff and CummeRbund) following the analysis pipeline described by Trapnell et al. (2012). The parameter FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) was used to quantify transcripts expression. FPKM was calculated based on the mapped transcript fragments, transcript length and sequencing depth. Differential expression analysis of the two groups of milk samples was performed using Cuffdiff. Cuffdiff was run using the multi-read correction (`-u` option) and the `-b` option which improves the accuracy of transcript abundance estimates by running a new bias detection. The CummeRbund Bioconductor R package was employed to analyze Cuffdiff outputs.

## Results and Discussion

**Global analysis of gene expression in day 120 of lactation.** A total of 230 million paired-end reads were obtained from RNA sequencing. Three replicates were analyzed for each breed with reads ranging from 34 to 42 million paired-end reads per sample. Approximately, the 91% of the sequences (95 million paired reads in Assaf and 87 million paired-end reads in Churra) were mapped to a unique location in the ovine genome.

From the results obtained with cuffcompare (Table 1), it can be observed that more transcripts were identified for Churra than for Assaf (52,690 versus 47,510 reads). The distribution of the percentage across the different cuffcompare classes for the two breeds is very similar; the highest difference can be noticed for the unknown/intergenic transcripts which are slightly increased for the Churra breed (7% higher).

Figure 1 shows the six most highly expressed genes in the milk samples of Churra and Assaf breeds. Genes with the highest value of FPKM at day 120 of lactation are *LALBA* ( $\alpha$ -lactoalbumin), *ENSOARG0000005099* (*LGB*,  $\beta$ -lactoglobulin), *CSNIS1* ( $\alpha$ -S1-casein), *CSNIS2* ( $\alpha$ -S2-casein), *CSN3* ( $\kappa$ -casein), *CSN2* ( $\beta$ -casein). These highly expressed genes encode caseins and whey proteins, principal components of milk. From the Figure 1 we can observe that the caseins, with the exception of *CSNIS1* are slightly more expressed in Churra than in Assaf, while the whey proteins ( $\alpha$ -lactoglobulin and  $\beta$ -lactoglobulin) are lightly more expressed in Assaf than in Churra. These results agree with the differences observed in cheese yield between the two breeds, as the Churra breed is reported to have more protein and fat contents in milk compared to Assaf sheep (De la Fuente et al. (2006)).

The most highly expressed genes from this study were compared with a study performed in bovine milk in three stages of lactation (transition, peak, and late lactation) (Wickramasinghe et al. (2012)). Interestingly, the profile of

the most highly expressed genes in sheep at day 120 of lactation is more similar to the gene expression profile in cows at the peak and transition stages of the lactation than at late lactation, contrary to what might be expected.

**Analysis of differentially expressed genes.** The differential expression analysis of annotated genes performed with Cuffdiff reveal a total of eight differentially expressed genes (DE) between the two breeds, Churra and Assaf (Figure 2). Five of these genes showed the highest expression in Assaf sheep (*GPRC5C* (G-protein coupled receptor family C group 5 member C), *FGL2* (Fibrinogen-Like 2), *LIPG* (Endothelial Lipase), *TGM3* (Protein-glutamine gamma-glutamyltransferase E), *ENSOARG00000016020*) and three showed the highest expression in Churra sheep (*MGP* (Matrix Gla Protein), *PDK4* (Pyruvate Dehydrogenase Kinase, Isozyme 4), *SLCO2B1* (solute carrier organic anion transporter family, member 2B1) (Figure 2). As can be observed in the heatmap plot the most highly expressed DE gene is *MGP* in Churra (1378 FPKM). The fold change of these DE genes ranged from 2.47 to 3.65. The gene with the highest fold change between two conditions is *FGL2*.

A deeper analysis of the eight DE genes was performed. For some of these DE genes, a relationship with lactation maintenance and milk composition was found. This relation is discussed below.

*FGL2* is a gene related to innate immunity; a study of the effects of increasing milking frequency in bovine mammary gland (Connor et al. (2008)) showed that under stress conditions, like increasing milking frequency, the *FGL2* is down-regulated.

The *GPRC5C* gene encodes a membrane protein receptor. This gene product is included among the milk fat globule proteins where appears to be involved in signal transduction and lipid droplet transport (D'Alessandro et al. (2011)).

Other of the genes identified as DE in this work, *LIPG*, is involved in the lipid metabolism of the mammary gland. The *LIPG* protein influences milk fatty acid composition, showing a large negative correlation with CLA10c12, CLAc11t9 and other trans-fatty acids, and large positive correlations with C16, SCFA and SFA (Mach et al. (2013)). The DE observed for *LIPG* between these two breeds could suggest differences in the milk fatty acid composition between Churra and Assaf, which have not been studied up to date.

The *MGP* gene encodes the Matrix-Gla protein, which acts as an inhibitor of calcification (Luo et al. (1997)). The absence of this protein makes the expression of osteopontin, another protein implicated in the modulation/inhibition of calcification, to be upregulated (Speer et al. (2002)). Osteopontin is a protein present in tissues elaborating fluids, such as lactating mammary gland, where it has been shown to be a regulator of lactation (Sheehy et al. (2009)). Levels of osteopontin are high during the initiation of lactation and the decrease, but remain high during involution (Rittling and Novick (1997)). Due to the similar functions of these two proteins, Matrix-Gla could act in a similar way to osteopontin during lactation.



## Conclusion

The undertaken RNA-seq experiment has allowed presenting herein a preliminary study on the gene expression profiling in sheep's milk somatic cells. RNA-seq has shown to be a powerful tool to analyze the ovine milk transcriptome.

Genes encoding caseins and whey proteins show the highest expression at day 120 of lactation in sheep's milk. In general, from these six genes, whey proteins were slightly more abundant in Assaf and caseins in Churra.

The differential expression analysis revealed a total of eight DE genes between the two dairy breeds, Churra and Assaf. Some of these genes display known functions related to lactation and milk composition.

Further analysis on functional enrichment and covering other stages of lactation should be carried out in order to obtain a global expression profiling of the milk somatic cells in sheep.

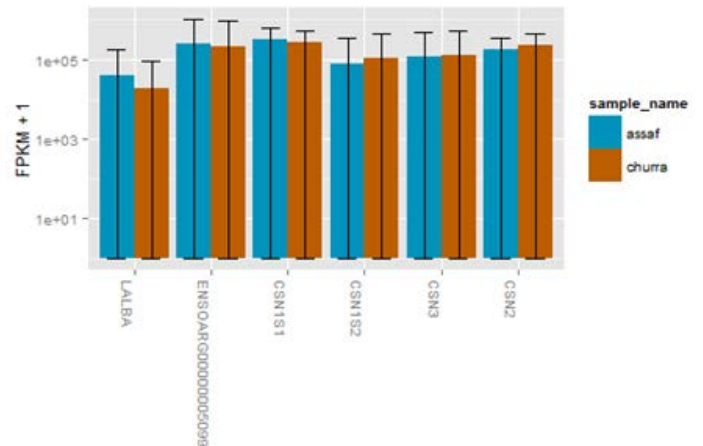
## Literature Cited

- Connor E.E., Siferd S., Elsasser T.H., et al. (2008) BMC Genomics, 9: 362-2164-9-362.
- D'Alessandro A., Zolla L., Scaloni A. (2011). Mol. Biosyst., 7: 579-597.
- De la Fuente, L.F., Gabiña D., Carolino N., et al. (2006). E.A.A.P. 57 Annual Meeting, Paper S14.2.
- Lemay D.G., Ballard O.A., Hughes M.A., et al. (2013) PLoS One, 8: e67531.
- Li H., Handsaker B., Wysoker A., et al. (2009). Bioinformatics, 25: 2078-2079.
- Luo G., Ducey P., McKee M.D., et al. (1997). Nature, 386: 78-81.
- Mach N., Goselink R.M.A., van Baal J., et al. (2013). Livestock Sci., 151: 92-96.
- Metzker M.L. (2010). Nat. Rev. Genet., 11: 31-46.
- Rittling S.R., Novick K.E. (1997) Cell Growth Differ. , 8: 1061-1069.
- Sheehy P.A., Riley L.G., Raadsma H.W., et al. (2009). A Anim. Genet., 40: 492-498.
- Shu D.P., Chen B.L., Hong J., et al. (2012). OMICS, 16: 123-137.
- Speer M.Y., McKee M.D., Guldberg R.E., et al. (2002). J. Exp. Med., 196: 1047-1055.
- Trapnell C., Pachter L., Salzberg S.L. (2009). Bioinformatics, 25: 1105-1111.
- Trapnell C., Roberts A., Goff L., et al. (2012). Nat. Protoc., 7: 562-578.
- Wang Z., Gerstein M., Snyder M. (2009). Nat. Rev. Genet. 10: 57-63.
- Wickramasinghe S., Rincon G., Islas-Trejo A., Medrano J.F. (2012). BMC Genomics, 13: 45-2164-13-45.

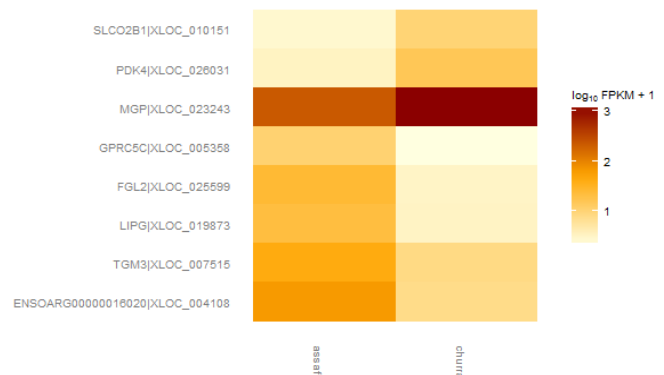
**Table 1: Classification of the transcripts identified in the milk samples in relation to Ensembl annotated genes in sheep genome.**

| Cuffcompare class                                | N° transcripts assaf | % assaf | N° transcripts churra | % churra |
|--|----------------------|---------|-----------------------|----------|
| Complete match of intron chain                   | 25024                | 52,67   | 25046                 | 47,53    |
| Multiple classifications                         | 0                    | 0       | 0                     | 0        |
| Contained in the reference                       | 0                    | 0       | 0                     | 0        |
| Possible pre-mRNA fragment                       | 0                    | 0       | 0                     | 0        |
| Transcript falling within a reference intron     | 14                   | 0,03    | 22                    | 0,04     |
| Potentially novel isoforms                       | 15294                | 32,19   | 16510                 | 31,33    |
| Generic overlap with a reference transcript      | 717                  | 1,51    | 783                   | 1,49     |
| Possible polymerase run-on fragment              | 0                    | 0       | 0                     | 0        |
| Intergenic transcript, unknown                   | 6233                 | 13,12   | 10054                 | 19,08    |
| Exonic overlap on opposite strand                | 228                  | 0,48    | 273                   | 0,52     |
| Repeat   | 0                    | 0       | 0                     | 0        |
| Overlapping intron transfrag in the other strand | 0                    | 0       | 2                     | 0,01     |

**Figure 1: Graphical representation of the six most highly expressed genes in the milk samples of two sheep breeds (Churra and Assaf) at day 120 of lactation.**



**Figure 2: Heatmap plot of the differentially expressed (DE) genes between the two sheep breeds under study.**





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

Aroa Suárez-Vega<sup>1</sup>, Beatriz Gutiérrez-Gil<sup>1</sup>, Christophe Klopp<sup>5</sup>, Christèle Robert-Granie<sup>2,3,4</sup>, Gwenola Tosser-Klopp<sup>2,3,4</sup> and Juan José Arranz<sup>1</sup>

<sup>1</sup>Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain; <sup>2</sup>INRA, UMR1388 GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31326 Castanet-Tolosan, France; <sup>3</sup>Université de Toulouse, INP, ENSAT, GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31326 Castanet-Tolosan, France; <sup>4</sup>Université de Toulouse, INP, ENVT, GenPhySE (Génétique, Physiologie et Systèmes d'Élevage), F-31076 Toulouse, France ; <sup>5</sup>INRA, Sigénae, UR875 Biométrie et Intelligence Artificielle, BP 52627, 31326 Castanet-Tolosan Cedex, France



## **Abstract**

This study presents a dynamic characterization of the sheep milk transcriptome aiming at achieving a better understanding of the sheep lactating mammary gland. Transcriptome sequencing (RNA-seq) was performed on total RNA extracted from milk somatic cells from ewes on days 10, 50, 120 and 150 after lambing. The experiment was performed in Spanish Churra and Assaf breeds, which differ in their milk production traits. Nearly 67% of the annotated genes in the reference genome (Oar\_v3.1) were expressed in ovine milk somatic cells. For the two breeds and across the four lactation stages studied, the most highly expressed genes encoded caseins and whey proteins. We detected differentially expressed genes (DEGs) across lactation points, with the largest differences being found, for both breeds, between day 10 and day 150. Upregulated GO terms at late lactation stages were linked mainly to developmental processes and organ morphogenesis. A total of 256 annotated DEGs were detected in the Assaf and Churra comparison. The genes that were selectively upregulated in the Churra breed, grouped in endopeptidase and channel activity GO terms, could be related to the higher cheese yield of this breed. Overall, this study provides the first integrated overview on sheep milk gene expression.

**Keywords:** sheep, lactation, RNA-seq

## **Introduction**

Lactation is a specific mammalian function essential for newborn feeding. The milk transcriptome has been characterized in various mammalian species<sup>1-4</sup>. The knowledge of gene expression involved in lactation informs the basic study of mammary gland biology, morphogenesis and metabolic activity as well as enhances our understanding of milk composition in relation to the production of milk-based newborn formulas, milk and milk derivatives.

Livestock species have been used for centuries to provide milk and dairy products for humans. Sheep milk is currently an important source of revenue, ranking fourth in terms of global milk production in 2013 (<http://faostat.fao.org/>). The dairy sheep industry is primarily concentrated in Europe and the countries near the Mediterranean Sea <sup>5</sup>. Sheep milk is commonly used to produce many cultured dairy products. Spain was the world's seventh largest producer of sheep milk in 2012 (<http://faostat.fao.org/>), and the 70% of that milk production is concentrated in the community of Castile and Leon (366.537 million liters in 2012) (<http://www.magrama.gob.es/es/>). Assaf and Churra are the most important dairy breeds in this region, with 77,896 and 35,094 recorded ewes, respectively, in 2014 (<http://www.magrama.gob.es/es/>). Churra is a Spanish autochthonous breed, characterized by its rusticity; it is well adapted to the predominantly harsh environments of its production areas. Assaf is a more specialized dairy sheep created as a crossbreed between Awassi (5/8) and Milschchaf (3/8) breeds. Assaf was introduced in Spain in 1977, and its population has increased primarily because of its high production potential and its efficient adaptation to local conditions <sup>6</sup>. Lactation is normalized to 120 days in Churra and 150 days in Assaf. The Assaf milk yield (400 kg) is more than double of the milk yield in Churra (117 kg), although Assaf milk has lower fat (6.65 vs. 7.01) and protein content (5.40 vs. 5.79) (<http://www.magrama.gob.es/es/>).

Transcriptome sequencing (RNA-seq) technologies provide a unique opportunity to characterize cell transcripts (including alternative splicing and the discovery of new genes and single nucleotide polymorphism in coding sequences), to quantify transcripts and to identify differential regulation in a single experiment <sup>7</sup>. In recent years, RNA-seq technology has been applied to the study of lactating mammary gland in dairy cattle <sup>3, 8-10</sup>.

Various RNA sources have been used to study the mammary gland transcriptome during lactation in mammals, including mammary gland biopsies <sup>4, 11</sup>, milk fat globules <sup>1</sup> and milk somatic cells (MSCs) <sup>3</sup>. For this study, RNA was extracted from sheep MSCs. The analysis of MSCs is a more accessible method compared to invasive approaches, especially when dynamic

studies with several sampling time points are required for the same animal<sup>8,12</sup>. Recent analyses suggest that MSCs are representative sources of RNA in mammary gland tissue, and MSCs isolation is an effective and simple method to study the mammary gland transcriptome through RNA-Seq<sup>13</sup>.

The aim of this research was to gain a better understanding of the sheep lactating mammary gland and to compare the mammary gland transcriptome of two sheep breeds with different dairy production characteristics, Spanish Churra and Assaf. To that end, we performed RNA-seq analysis of MSCs in these two dairy sheep breeds at different lactation time points. Here, the successful isolation and generation of global gene expression data from ovine MSCs is reported for the first time. The characterization of milk sheep profile and differences between the milk transcriptomes of Assaf and Churra breeds are described. The results herein provide a significant advance in our knowledge of sheep lactating mammary gland gene expression and valuable information for future studies.

## **Material and Methods**

### **Animals and sampling**

A total of eight healthy sheep were selected to be included in the experiment, four Assaf and four Churra ewes. The animals belong to the commercial farm of the University of León. These sheep were kept in free stall housing, fed with the same rations and had no water restrictions. Animals were milked twice a day: at 8 a.m. and 6 p.m. The lambing by these sheep was between the November 11, 2012, and December 11, 2012. Milk samples were collected on days 10 (D10), 50 (D50), 120 (D120) and 150 (D150) after lambing. For each sampled animal and lactation point, at least four samples of 50 ml of milk were collected; two of them were obtained on the exact sampling day whereas two additional samples were collected the previous or the following day to ensure RNA source material for each desired sampling.

With the aim of maximizing the number of somatic cells present in milk, the sample collection was made one hour after the 8 a.m. routine milking and ten minutes after the injection of 5 IU of Oxitocine Facilpart (Syva, León, Spain), as indicated by <sup>14</sup>. All protocols involving animals were approved by the Animal Welfare Committee of the University of Leon, Spain, following proceedings described in Spanish and EU legislations (Law 32/2007, R.D. 1201/2005, and Council Directive 2010/63/EU). All animals used in this study were handled in strict accordance with good clinical practices and all efforts were made to minimize suffering.

To ensure RNA purification of high yield and quality, we used the following protocol during the sampling process. Before sampling, the collecting milk containers were cleaned with RNaseZap (Ambion, Austin, TX, USA) and autoclaved. In the farm, udder cleaning was performed with special care: first, the udders were cleaned with water and soap; then, they were disinfected with povidone iodine; and finally the nipples were cleaned with RNaseZap (Ambion, Austin, TX, USA). Sterile gauze was used to cover the collecting container during milk collection. The milk was passed from the collecting container to RNase-free 50 ml tubes after collection. Samples were maintained at 4°C during their transport from the farm to the laboratory.

### **RNA extraction**

Approximately 50 ml of milk was used for RNA extraction. The pellet of MSCs was obtained as described by <sup>3</sup> with some modifications. The cells were pelleted by centrifugation at 540 *g* in 50 ml RNase free sterile tubes for 10 minutes at 4 °C in the presence of a final concentration of 0.5 mM of EDTA. After centrifugation, the supernatant was discarded. During this step, a fatty layer was usually placed in the first part of the tube. To remove it a sterile pipette tip was introduced to separate this fatty layer from the tube walls. Then, the cell pellet was washed in PBS (pH 7.2) with 0.5 mM EDTA and centrifuged at 540 *g* in 15 ml RNase free sterile tubes for 10 minutes at 4°C. The last step was repeated until the fatty layer was minimized. Once the pellet was clean, it was resuspended in 500 µl of Trizol (Invitrogen, Carlsbad, CA, USA) and



homogenized by a vortexer. RNA extraction continued following a standard Trizol protocol (Invitrogen, Carlsbad, CA, USA).

### **RNA sequencing**

The Agilent 2100 Bioanalyzer device (Agilent Technologies, Santa Clara, CA, USA) was used to assess the integrity of the RNA. The RNA integrity value (RIN) of the samples ranged between 7.1 and 9. Paired-end libraries with fragments of 300 bp were prepared using the True-Seq RNA-Seq sample preparation Kit v2 (Illumina, San Diego, CA, USA). The fragments were sequenced on an Illumina Hi-Seq 2000 sequencer (Fasteris SA, Plan-les-Ouates, Switzerland), according to the manufacturer's instructions at CNAG (Centro Nacional de Análisis Genómico, Barcelona, Spain), generating paired-end reads of 75 bp.

### **Quality Control, Mapping and Quantification**

FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to assess the quality of raw sequencing data. Reads were mapped against the ovine genome assembly v.3.1. (Oar\_v3.1) using STAR aligner (v. 2.3.1y)<sup>15</sup>. The alignments were sorted with Samtools<sup>16</sup>. The data were also tested for contamination using BWA<sup>17</sup> on the *Escherichia coli* genome.

To compare expression between genes within the same sample, gene expression was estimated using Cuffquant and Cuffnorm packages from Cufflinks<sup>18</sup>. Gene abundances were normalized by library and gene length by calculating Fragments Per Kilobase Of Exon Per Million Fragments Mapped (FPKM) using the Ensembl annotated genes (Oar\_v3.1) as a reference.

### **Assembly and differential expression analysis**

The aim of the assembly was producing a new annotation reference including novel genes and transcripts, to be used in the downstream analysis. The Samtools package was used to remove the duplicated reads from the alignments with the “rmdup” option and these deduplicated alignments were merged with the “merge” option. The Cufflinks and Cuffmerge tools from the Cufflinks package<sup>18-20</sup> were used to create a “transcripts.gtf” file to be used as reference in our

assembly. The Cufflinks option “-g” followed by the available gtf file from the Oar\_v3.1 reference sequence was used to guide the assembly but not excluding new genes. Cuffmerge was used to filter genes with low or no expression from our reference gtf file. Cuffcompare was used to compare the reference gtf file obtained with Cuffmerge with the reference annotation file downloaded from Ensembl ([http://www.ensembl.org/Ovis\\_aries/Info/Index](http://www.ensembl.org/Ovis_aries/Info/Index)) to estimate the abundance of transcripts classified in different Cufflinks class codes, such as “complete match intron chain”, “novel isoforms” or “intergenic transcripts”.

To compare the expression levels of genes across samples, raw counts for the genes and transcripts were obtained using SigCufflinks (available at <http://www.sigenae.org>). SigCufflinks is a modified version of the cufflinks code that provides raw read counts per gene and transcript, by using the sorted bam file from the alignment and the reference gtf file created in the assembly. The “-G” option of sigcufflinks was used to guide the alignment but excluding new genes.

To evaluate the differentially expressed genes (DEGs) from the RNA-seq data, two different R packages, DESeq2<sup>21</sup> and edgeR<sup>22</sup>, were used. The DESeq2 package performs independent filtering. In general, genes that have very low counts are not likely to see significant differences due to high dispersion. The edgeR package has no function for independent filtering. To choose a cut-off to eliminate the lowest expressed genes, we plotted a histogram on R with the  $\log_{10}(\text{effmax}+1)$ , thereby maintaining those genes for which the distribution becomes normal and eliminating the genes with less than 15 counts.

RNA-Seq read counts were modeled by a generalized linear model considering the experimental design, with two breed groups (Churra and Assaf) and four lactation time points per group (D10, D50, D120 and D150). The model for both programs includes *breed* and *day* factors and the interaction *day x breed*. After discarding a possible interaction between these two factors, we performed the following analyses with edgeR and DESeq2. First, with the aim of providing a dynamic expression profile of the mammary gland across lactation for each breed, differential

expression analyses were performed for all the possible time point pairs. Second, a differential expression analysis between Churra and Assaf samples was performed.

DESeq2 and edgeR perform pairwise comparisons between two or more groups using parametric tests where read-counts follow a negative binomial distribution with a gene-specific dispersion parameter. These packages mainly differ in the estimation of the dispersion parameter and the type of normalization they follow. DESeq2 bases the estimation of the dispersion on calculated mean–variance relationships in the data set provided, while edgeR assumes a common dispersion for all the genes. These programs normalize the read count per gene based on total gene depth per individual; DeSeq2 computes a scaling factor as the median of the ratio for each gene of its read count over its geometric mean across the lanes<sup>21</sup>. edgeR follows the Trimmed Mean of M-values method (TMM normalization) in which the TMM factor is computed for each lane as the weighted mean of log ratios between the test and the reference<sup>22</sup>. Our choice of these two methods was based on the literature evidence for their robustness.

For DESeq2, the DEGs were defined as those genes that had an absolute log<sub>2</sub>-fold change > 1.5 and an adjusted p-value ( $p_{adj}$ -value) < 0.05, whereas edgeR DEGs were those that had a log<sub>2</sub>-fold change > 1.5 and an adjusted FDR (false discovery rate) < 0.05. To identify the core DEGs, genes identified as DEGs by the two packages were defined as core DEGs and were subjected to subsequent functional analyses.

### **Gene functional classification, network and canonical pathways**

The web-based Gene Set Analysis Toolkit (WebGestalt<sup>23</sup>) was used to perform a Gene-Ontology (GO) enrichment analysis. Enriched terms were considered statistically significant when  $p_{adj}$ -value was <0.05 and a minimum of six genes were grouped for each significant term. The GO terms were categorized in three major functional groups: biological process, molecular function and cellular component.

## **Results and discussion**

### Summary statistics for the RNA-seq data

Based on the quality scores of the extracted RNA samples for each breed, samples from four animals were sequenced for time points D10, D50 and D150, and three biological replicates were sequenced for D120. A total of 1,116 million paired-end reads were obtained from the transcriptome sequencing of the 30 milk samples analyzed. Alignment of the reads to the *Ovis aries* genome Oar\_v3.1 genome build yielded a mean of 985.05 million reads (88.10%) that aligned to unique locations in the ovine genome per RNA-seq sample; a mean of 1.47 million reads (4.01%) per sample that aligned to multiple locations in the genome; and a mean of 2.84 million reads (7.65%) per sample that did not align to any genome location. No contamination was found in the alignment against the *Escherichia coli* genome. The average percentage of uniquely mapped reads for each sample was substantially higher than that obtained in the analysis of the bovine milk transcriptome, where approximately 65% of the total reads uniquely mapped to the Btau 4.0 reference genome<sup>3</sup>.

The distribution of uniquely mapped reads to annotated genes in the Oar\_v3.1 reference is described in Table 1. According to the FPKM value, mapped genes were divided into a low expression group (<10 FPKM), a moderate expression group ( $\geq 10$  FPKM to 500 FPKM) and a high expression group ( $\geq 500$  FPKM). Using a threshold of > 0.01 FPKM to define potentially significant gene expression<sup>24</sup>, we found an average of 16,757.25 and 16,897.00 unique expressed genes in Assaf and Churra, respectively. These figures represent 66.51% and 67.06%, respectively, of the total annotated genes in the Oar\_v3.1 assembly. For both breeds, the lactation time point with the highest number of expressed genes was D50 (17,386 genes in Assaf and 17,192 genes in Churra). From a global view, the transcriptome results revealed that the majority of the genes had low expression (FPKM < 10) across lactation for both breeds (Table 1). The average gene expression levels reported here for sheep MSCs are in agreement with previous reports on the mammary gland transcriptome of other ruminant species<sup>3, 25</sup>.

A total of 107,877 transcripts were assembled by Cufflinks using the 30 analyzed milk samples. According to the Cuffcompare results (Table 2), 23.95% of the transcripts matched exactly with annotated exons, and a total of 12,057 (11.18%) potentially novel isoforms were predicted. A high percentage, 61.87% (66,739) of transcripts, were annotated as intergenic transcripts, which is in agreement with the results reported from the RNA-Seq analysis of the ovine muscle transcriptome in which the proportion of intergenic transcripts ranged between 55.44% (Small-tailed Han sheep) and 66.40% (Dorper sheep) (DP) (using the ovine genome Oarv2.0)<sup>26</sup>. These results unfortunately underline the incompleteness of the annotation of the sheep transcriptome, and the need for further studies to decipher its complexity. The Functional Annotation of Animal Genomes (FAANG) project (<http://www.faang.org/>) aims to identify all functional elements in animal genomes, and sheep is one of the target species.

#### **Analysis of the ten most expressed genes in each stage of lactation**

Gene abundances were normalized by library size and gene length (FPKM) to determine the list of the most expressed genes at each lactation time point. This normalization approach facilitates the comparison of genes within a sample<sup>27</sup>. Table 1 shows the number of genes with the highest FPKM ( $\geq 500$ ) for each breed at each time point analyzed. The results show that in sheep milk, the top-10 genes have very high expression values. For both breeds, Churra and Assaf, the ten highest expressed genes at each time point accumulate at approximately 70% of the total gene FPKM reads, which means that a small number of genes contribute to a large fraction of the total RNA extracted from MSCs. A total of 13 genes, ranging between means of 4,459 to 219,181 FPKM, are encompassed in the top-10 highest expressed genes (Fig. 1). In general, with exceptions for very specific changes in the profile expression, the same genes are highly expressed at all of the different time points assessed during sheep lactation. In contrast, in dairy cattle late lactation, the caseins are not included in the top-expressed category and are instead replaced by other genes involved in proteolysis, anti-apoptotic activity and immune functions<sup>3</sup>.

The genes with the highest FPKM values for both breeds and at the four studied lactation time points are *CSN2* ( $\beta$ -casein), *CSN3* ( $\kappa$ -casein), *ENSOARG0000005099* (*LGB*,  $\beta$ -lactoglobulin), *CSNIS2* (casein- $\alpha$ -S2), *CSNIS1* ( $\alpha$ -S1-casein) and *LALBA* ( $\alpha$ -lactoalbumin). These highly expressed genes encode four caseins and two whey proteins, principal components of milk. Caseins and whey proteins encompass the 5.5% of total milk composition in sheep, a higher percentage than in bovine milk (3.2%)<sup>28</sup>. Caseins and whey protein expression remained constant for Churra along lactation. As shown in Fig. 1, these six genes were slightly more abundant in Churra sheep at the first two sampling time points (D10 and D50) than in Assaf sheep. In contrast, the expression of caseins and whey proteins in the Assaf breed underwent an increase at the last two time points studied (D120 and D150). The *LALBA* gene, which encodes for the  $\alpha$ -lactoalbumin, had higher expression in Assaf at D120 and D150. A SNP described in this gene has been associated with a higher milk protein and fat content (milk concentration) in Churra sheep<sup>29</sup>. The  $\alpha$ -lactoalbumin is a major milk whey protein involved in the synthesis of lactose, which is responsible for drawing water into the milk. The higher milk yields of Assaf as compared to Churra could be due, among other factors, to higher expression levels of the *LALBA* gene in this breed.

Apart from caseins and whey proteins, there are two other highly expressed genes in the two breeds and at the four stages of lactation studied: *GLYCAM-1* and *B2M*. *GLYCAM-1* encodes the glycosylation-dependent cell adhesion molecule 1, a member of the glycoprotein mucin family, which is a component of the milk fat globule membrane<sup>30</sup> and appears to be hormonally regulated in the sheep mammary gland<sup>31</sup>. *B2M* encodes for the beta-2-microglobulin protein, an integral component of the Fc receptor heterodimer, which is involved in transferring Immunoglobulin G (IgG) from serum into milk across mammary epithelial cells<sup>32</sup>. Some *B2M* haplotypes have been reported to be related to higher concentrations of IgG1 in bovine milk<sup>33</sup>. Increasing IgG levels in milk could become important as IgG enhanced dairy products are in demand by consumers to obtain protective immunity<sup>34</sup>.

Three other genes included among the top-10 highly expressed genes were found in the two studied breeds but not across all the time points. The *5.8S rRNA* (5.8S ribosomal RNA) gene is a non-coding RNA component of the large subunit of the eukaryotic ribosome. It is part of the ten most expressed genes in MSCs at D10, D50 and D150 for Churra and Assaf breeds. Expression of this gene in the Churra breed remains constant, while in the Assaf breed, it decreases as lactation proceeds (Fig. 1). In the rabbit mammary gland, expression of rRNA is related to high levels of protein synthesis, correlations between prolactin binding to mammary epithelial cells and high levels of rRNA synthesis<sup>35</sup>. The *SERP1* gene, which encodes the stress-associated endoplasmic reticulum protein 1, is among the ten most highly expressed genes at D10 in both breeds, at D120 in Assaf and at D150 in Churra. This protein is enhanced by stress causing the accumulation of unfolded proteins in endoplasmic reticulum, suppressing the aggregation and/or degradation of newly synthesized integral membrane proteins, and facilitating their glycosylation when the stress is removed<sup>36</sup>. The endoplasmic reticulum stress gene network has been analyzed in the mammary glands of ruminants and non-ruminants to study their role in maintaining protein cellular homeostasis as well as the potential implications in lipid homeostasis<sup>37,38</sup>. The *OST* (also known as *SPP1*) gene encodes for osteopontin, a major phosphoglycoprotein highly expressed in the mammary gland. This gene is one of the ten most expressed genes at D50, D120 and D150 of lactation in the Assaf breed but only at D50 in the Churra breed. The role of osteopontin in the mammary gland is not completely clear. Several studies have shown an association between the expression of the *OST* gene and milk yield by enhancing the expression of caseins *CSN2* and *CSN3*<sup>39</sup>, and *OST* has also been related to mammary gland morphogenesis<sup>40</sup> and new born immunity<sup>41,42</sup>.

In addition, the top-10 list of genes at D120 in Churra sheep contained two genes that were not included in this category in the Assaf breed at any time point. One of them, *CLU*, encodes for the clusterin protein. In mice, the *CLU* gene is highly expressed during mammary gland prepartum development and also at the involution stage and is highly associated with tubuloalveolar morphogenesis and alveolar epithelial cell differentiation<sup>43</sup>. Here, in agreement

with the observations described in mice, *CLU* is highly expressed at the end of lactation. In Assaf sheep, where the lactation is longer than in Churra sheep, although this gene was not found within the top-10 list of any time point, it was found within the group of highly expressed genes at D120 and D150. *CLU* has also been found to be associated to apoptotic events and environmental stress<sup>43</sup>. In normalized lactation, D120 corresponds to the last stage lactation in Churra sheep; hence, the intense remodeling of mammary epithelial cells due to the milking on this last stage could stimulate the *CLU* expression. The second gene exclusively identified in Churra sheep in the top-10 list at D120 is *SAA*, which encodes for a member of the serum amyloid family of apolipoproteins. High levels of *SAA* have been related to states of inflammation in the mammary gland<sup>44</sup>. However, studies in pigs in healthy lactating mammary glands showed an increase of this gene's mRNA in the second half of lactation, providing evidence that this peptide could play a physiological role in the normal mammary gland, stimulating neonatal gut immune responses<sup>45</sup>.

### **Differentially expressed genes across lactation**

For the four comparisons for which DEGs genes were detected by at least one of the packages (all except D50 vs. D120 and D120 vs. D150), a higher number of genes were detected with DESeq2 than with edgeR (Fig. 2). In total, the number of unique DEGs identified was 217 with edgeR and 503 with DESeq2. For the two breeds, the D10 vs. D150 comparison had the largest number of DEGs. Because this comparison for the Churra breed showed the highest number of DEGs (173 with edgeR and 386 with DESeq2), we used this specific contrast to compare the significant results from the two packages (Fig. 3a). While the highly conservative correction method of edgeR was identified as one of the main factors explaining the discrepancies, when considering the genes commonly identified by the two software solutions (Fig. 3b), we observed a strong linear correlation between the logFC values obtained by the two packages. Based on this, the only DEGs commonly identified with the two packages whose number ranged between 0 (for the Assaf D10 vs. D50 contrast) and 152 (for the Churra D10 vs. D150 contrast) (Fig. 2) were subjected to a functional GO term enrichment analysis. Due to the large number of results



generated in this analysis for all the comparisons (See Supplementary File S1), we only provide a detailed discussion of the results obtained for the D10 vs. D150 comparisons. For the two breeds, this was the comparison-pair with the largest number of DEGs: 152 in the Churra breed and 47 in the Assaf breed.

For the 152 DEGs identified in the Churra breed, 89 were annotated genes. Forty-four of these genes had a significantly higher expression at D10 than at D150. A total of 40 terms in the GO biological process database were found to be significantly enriched for the upregulated genes in the Churra breed at D10. Among the significant GO terms detected, we found terms related to response to stimulus ( $p_{\text{adj}} = 0.0272$ ), response to stress ( $p_{\text{adj}} = 0.0223$ ), signaling genes ( $p_{\text{adj}} = 0.0223$ ) implicated in positive regulation of signal transduction ( $p_{\text{adj}} = 0.0233$ ), cell communication ( $p_{\text{adj}} = 0.0252$ ), cell proliferation ( $p_{\text{adj}} = 0.0252$ ), regulation of cell differentiation ( $p_{\text{adj}} = 0.0252$ ) and positive regulation of protein phosphorylation ( $p_{\text{adj}} = 0.0223$ ). As expected for the D10 time point, all of these terms involve biological regulation and metabolic protein synthesis processes clearly related to the extreme metabolic stress associated with high levels of milk production. Four GO terms were significant for the GO cellular component cluster and were all related to the extracellular region (extracellular matrix ( $p_{\text{adj}} = 0.0017$ ), extracellular region ( $p_{\text{adj}} = 0.0468$ ), extracellular region part ( $p_{\text{adj}} = 0.0017$ ) and extracellular space ( $p_{\text{adj}} = 0.0017$ )), while no terms were significantly enriched for the GO molecular function cluster.

Forty-five of the DEGs were upregulated at D150 in the Churra breed. From these, a total of 32 GO terms from the GO biological process database were significantly enriched. Among them, we found terms related to developmental process ( $p_{\text{adj}} = 0.0208$ ), organ morphogenesis ( $p_{\text{adj}} = 0.0222$ ), cell motility ( $p_{\text{adj}} = 0.0156$ ), cell migration ( $p_{\text{adj}} = 0.0123$ ), chemotaxis ( $p_{\text{adj}} = 0.0067$ ) and nervous system development ( $p_{\text{adj}} = 0.0083$ ). The nervous system has a role in mammary gland milk ejection, and pathways related to the nervous system have also been found in the bovine mammary transcriptome<sup>11</sup>. Here, GO terms related to apoptotic processes were not found, but significant enrichment terms related to mammary gland remodeling, which is a

characteristic process of the lactation decline, were observed. In cattle, changes in the mammary gland during lactation decline and involution reflects a change in secretory stage rather than extensive tissue degeneration<sup>46</sup>. No significant GO molecular function terms were found for upregulated genes at D150, whereas twelve enriched terms were detected for GO cellular components. Among them, we found plasma membrane ( $p_{adj} = 0.0268$ ), extracellular matrix ( $p_{adj} = 0.0096$ ), extracellular region ( $p_{adj} = 0.0435$ ) and endoplasmic reticulum ( $p_{adj} = 0.0268$ ). We highlight the significant GO terms related to endoplasmic reticulum as this organelle is linked to the lipid secretory mechanism of the mammary epithelial cells<sup>47</sup>. Initial analyses on the expression of genes related to mammary gland fat metabolism suggest that genes implicated in the *de novo* synthesis of sheep milk fat have increased expression at late lactation<sup>48</sup>.

For the Assaf breed, among the 47 DEGs genes detected in the D10 vs. D150 comparison, 38 were annotated. Of them, 17 and 21 genes were upregulated at D10 and D150, respectively. Twenty-two genes were in common with the DEGs found in the Churra breed. No significant enriched GO terms were found for upregulated genes at D10. For the genes upregulated at D150, 13 terms were enriched in the GO biological process database; as in the Churra breed at D150, these genes are primarily involved in developmental processes ( $p_{adj} = 0.0329$ ). The GO biological function term related to macromolecule biosynthetic process ( $p_{adj} = 0.0415$ ) was also enriched. No significant GO molecular function terms were found, and within the GO cellular component database, the single significantly enriched GO term was extracellular region ( $p_{adj} = 0.0342$ ).

The results reported here show that sheep gene expression is very similar across lactation time points. The major differences in gene expression levels and DEGs were found between extreme stages of lactation in the two breeds analyzed. Fig. 4 shows a heatmap with the unique 191 common DEGs detected across lactation with the two analysis software solutions used here. The genes are clustered in two major groups, one corresponding to D10 and D50 time points (related to the initial stages of lactation) for both breeds and the other corresponding to D120 and D150 time points (associated with the late stages of lactation). These observations differ with those

reported in cattle, where the analysis of the MSC transcriptome showed that the highest number of DEGs was found for the D15 vs. D90 comparison (6,243 genes;  $p < 0.05$ ,  $FDR < 0.3$ ), representing the comparison between the transition and the peak lactation stages<sup>3</sup>. To interpret these differences, we have to take into account that dairy cattle have a longer lactation (normalized to 305 days in Holstein) than sheep (normalized to 150 days in Assaf and 120 days in Churra). In addition, dairy sheep breeds, such as those analyzed here, are normally used with a double production objective: providing milk for cheese and lambs for meat. The standard management system involves one lambing per year in the Assaf breed (dairy specialized breed) and three lambings in two years for the Churra breed. Our results show that the profile of highly expressed genes is very similar between the two breeds and that within a breed, there are few differences among the highly expressed genes across different lactation stages. Thus, sheep lactation could physiologically continue longer, but this option is not currently exploited by breeders because of the economic benefits that lamb sales provide to the general farm economy. Larger differences were found between the extremes of lactation for the autochthonous Churra breed than for the specialized dairy breed Assaf. The Churra breed has shorter lactation than the Assaf breed, although here, we have elongated the milking period by 30 days such that larger differences at late lactation were expected.

### **Differentially expressed genes between Churra and Assaf**

In the comparison between the Churra and Assaf breeds, edgeR identified a total of 1,039 DEGs, whereas 774 genes were identified as DEGs with the DESeq2 package. A total of 630 genes, of which 256 were annotated, were found in common as DEGs from both packages. A total of 172 annotated genes were upregulated in the Churra breed, and 84 annotated genes were upregulated in the Assaf breed. The GO terms with a significant  $p_{adj}$  ( $<0.05$ ) for the Churra and Assaf breeds' upregulated genes are shown in Supplementary File S2.

A total of 17 GO terms related to cell adhesion ( $p_{adj} = 0.0259$ ), cellular homeostasis ( $p_{adj} = 0.0259$ ) and developmental process ( $p_{adj} = 0.0259$ ) were enriched for the upregulated DEGs

identified in Churra sheep in the GO biological processes database. It has been demonstrated that interactions between mammary epithelial cells, and thus the genes involved in cell adhesion, are essential for cell viability and lactation<sup>49-51</sup>. In the Churra breed, it is remarkable that 10 terms related to the GO molecular functions endopeptidase activity ( $p_{\text{adj}} = 0.0016$ ) and channel activity ( $p_{\text{adj}} = 0.0325$ ) were found. Churra milk has better characteristics for cheese production than Assaf milk, due to its higher protein and fat contents. Endopeptidases have effects on the physicochemical characteristics and quality of dairy products<sup>52</sup>. It has been suggested that ion channel activity could be related to the regulation of milk secretion<sup>53</sup>. Milk volume is determined by osmotic-coupled water flow. The secretion of  $K^+$ ,  $Na^+$  and  $Cl^-$  in dairy cattle determines approximately 40% of the driving force, with the rest being determined by lactose<sup>54</sup>. Mineral concentrations also affect milk physicochemical properties. Renneting properties are influenced by pH, large casein micelle, more calcium per casein weight and other mineral concentrations in milk<sup>55</sup>. Additionally, in the Churra breed, among the 12 significant terms identified related to GO Cellular components, there were eight enriched terms related to cell components such as plasma membrane part ( $p_{\text{adj}} = 0.0017$ ), cell projection ( $p_{\text{adj}} = 0.0112$ ) and secretory granule ( $p_{\text{adj}} = 0.0211$ ) and four terms related to the extracellular region ( $p_{\text{adj}} = 0.0001$ ), extracellular region part ( $p_{\text{adj}} = 0.0142$ ), extracellular matrix ( $p_{\text{adj}} = 0.0142$ ) and proteinaceous extracellular matrix ( $p_{\text{adj}} = 0.0062$ ) terms. Components of the mammary gland extracellular matrix (ECM) are essential for supporting the basal lamina structure and providing and anchor for mammary epithelial cells<sup>56</sup>. Additionally, it has been shown that an increase in mammary gland matrix components can contribute to the regulation of mammary epithelial cell division<sup>57</sup>.

For the 84 DEGs with higher expression levels in Assaf as compared to Churra sheep, 3 terms were enriched in the GO biological process database that are related to the immune system process ( $p_{\text{adj}} = 0.003$ ), immune response ( $p_{\text{adj}} = 0.0001$ ) and response to stress ( $p_{\text{adj}} = 0.0385$ ) terms. There is an enriched GO molecular function term related to receptor binding ( $p_{\text{adj}} = 0.001$ ). For the GO cellular component, three significant terms related to the extracellular region

were found (extracellular region ( $p_{\text{adj}} = 0.0000375$ ), extracellular region part ( $p_{\text{adj}} = 0.0004$ ), extracellular space ( $p_{\text{adj}} = 0.0003$ )). In summary, from the breed comparison, a higher number of upregulated genes was found during lactation in Churra sheep. Churra DEGs grouped in GO molecular functions such as endopeptidase and channel activity, which could be related to the better cheese yield of this breed, whereas gene expression in Assaf sheep was dominated by the genes involved in immunity and stress responses.

Overall, this study represents the first integrated overview on the dynamic expression profile of the milk sheep transcriptome. In addition, this study allowed us to compare the milk transcriptome of two dairy sheep breeds, Churra and Assaf, showing that the differences in the gene expression profiles, although small, could serve to identify candidate genes explaining the known differences in production characteristics (milk yield, and milk composition) that exist between these two breeds. This work may provide fundamental information for future studies on specific pathways involved in lactation as well as the functional annotation of novel genes detected.

## References

1. Maningat, P. D. *et al.* Gene expression in the human mammary epithelium during lactation: the milk fat globule transcriptome. *Physiol. Genomics* **37**, 12-22 (2009).
2. Shu, D. P. *et al.* Global transcriptional profiling in porcine mammary glands from late pregnancy to peak lactation. *OMICS* **16**, 123-137 (2012).
3. Wickramasinghe, S., Rincon, G., Islas-Trejo, A. & Medrano, J. F. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* **13**, 45-2164-13-45 (2012).
4. Lemay, D. G. *et al.* Sequencing the transcriptome of milk production: milk trumps mammary tissue. *BMC Genomics* **14**, 872-2164-14-872 (2013).
5. Haenlein, G. F. Past, present, and future perspectives of small ruminant dairy research. *J. Dairy Sci.* **84**, 2097-2115 (2001).
6. De la Fuente, L., Gabina, D., Carolino, N. & Ugarte, E. The Awassi and Assaf breeds in Spain and Portugal. Proc. 57th Annual Meeting of the EAAP. Antalya, Turkey, Paper S14.2 (2006).

7. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57-63 (2009).
8. Medrano, J., Rincon, G. & Islas-Trejo, A. Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. 9th World congress on genetics applied to livestock production, Leipzig, Germany, No. 0852 (2010).
9. Canovas, A., Rincon, G., Islas-Trejo, A., Wickramasinghe, S. & Medrano, J. F. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm. Genome* **21**, 592-598 (2010).
10. Cui, X. *et al.* Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing. *BMC Genomics* **15**, 226-2164-15-226 (2014).
11. Bionaz, M. *et al.* Old and new stories: revelations from functional analysis of the bovine mammary transcriptome during the lactation cycle. *PLoS One* **7**, e33268 (2012).
12. Boutinaud, M. & Jammes, H. Potential uses of milk epithelial cells: a review. *Reprod. Nutr. Dev.* **42**, 133-147 (2002).
13. Canovas, A. *et al.* Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. *Sci. Rep.* **4**, 5297 (2014).
14. Gonzalo, C., Carriedo, J. A., Gomez, J. D., Gomez, L. D. & San Primitivo, F. Diurnal variation in the somatic cell count of ewe milk. *J. Dairy Sci.* **77**, 1856-1859 (1994).
15. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
16. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
17. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
18. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511-515 (2010).
19. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325-2329 (2011).
20. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562-578 (2012).
21. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

22. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
23. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77-83 (2013).
24. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621-628 (2008).
25. Lin, J. *et al.* Transcriptome analysis of the mammary gland from GH transgenic goats during involution. *Gene* **565**, 228-234 (2015).
26. Zhang, C. *et al.* Characterization and comparative analyses of muscle transcriptomes in Dorper and small-tailed Han sheep using RNA-Seq technique. *PLoS One* **8**, e72686 (2013).
27. Dillies, M. A. *et al.* A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**, 671-683 (2013).
28. Pulina, G. & Nudda, A. Milk Production in *Dairy Sheep Feeding and Nutrition* (ed. Poulina, G.) Ch. 1, 1-12 (CABI Publishing, 2004).
29. Garcia-Gamez, E. *et al.* GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PLoS One* **7**, e47782 (2012).
30. Le Provost, F., Cassy, S., Hayes, H. & Martin, P. Structure and expression of goat GLYCAM1 gene: lactogenic-dependent expression in ruminant mammary gland and interspecies conservation of the proximal promoter. *Gene* **313**, 83-89 (2003).
31. Kann, G., Delobelle-Deroide, A., Belair, L., Gertler, A. & Djiane, J. Demonstration of in vivo mammogenic and lactogenic effects of recombinant ovine placental lactogen and mammogenic effect of recombinant ovine GH in ewes during artificial induction of lactation. *J. Endocrinol.* **160**, 365-377 (1999).
32. Anderson, C. L. *et al.* Perspective-- FcRn transports albumin: relevance to immunology and medicine. *Trends Immunol.* **27**, 343-348 (2006).
33. Zhao, S. *et al.* Association of polymorphisms of beta-2-microglobulin gene ( $\beta 2m$ ) with milk IgG1 content in Chinese Holstein dairy cows. *Livestock Science* **143**, 289-292 (2012).
34. Hurley, W. L. & Theil, P. K. Perspectives on immunoglobulins in colostrum and milk. *Nutrients* **3**, 442-474 (2011).
35. Teysstot, B. & Houdebine, L. M. Role of progesterone and glucocorticoids in the transcription of the beta-casein and 28-S ribosomal genes in the rabbit mammary gland. *Eur. J. Biochem.* **114**, 597-608 (1981).

36. Yamaguchi, A. *et al.* Stress-associated endoplasmic reticulum protein 1 (SERP1)/Ribosome-associated membrane protein 4 (RAMP4) stabilizes membrane proteins during stress and facilitates subsequent glycosylation. *J. Cell Biol.* **147**, 1195-1204 (1999).
37. Bobrovnikova-Marjon, E. *et al.* PERK-dependent regulation of lipogenesis during mouse mammary gland development and adipocyte differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16314-16319 (2008).
38. Invernizzi, G., Naeem, A. & Loor, J. J. Short communication: Endoplasmic reticulum stress gene network expression in bovine mammary tissue during the lactation cycle. *J. Dairy Sci.* **95**, 2562-2566 (2012).
39. Sheehy, P. A., Riley, L. G., Raadsma, H. W., Williamson, P. & Wynn, P. C. A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Anim. Genet.* **40**, 492-498 (2009).
40. Nemir, M. *et al.* Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency. *J. Biol. Chem.* **275**, 969-976 (2000).
41. Alain, K. *et al.* Osteopontin: an early innate immune marker of Escherichia coli mastitis harbors genetic polymorphisms with possible links with resistance to mastitis. *BMC Genomics* **10**, 444-2164-10-444 (2009).
42. Schack, L. *et al.* Osteopontin enhances phagocytosis through a novel osteopontin receptor, the alphaXbeta2 integrin. *J. Immunol.* **182**, 6943-6950 (2009).
43. French, L. E., Soriano, J. V., Montesano, R. & Pepper, M. S. Modulation of clusterin gene expression in the rat mammary gland during pregnancy, lactation, and involution. *Biol. Reprod.* **55**, 1213-1220 (1996).
44. Pyorala, S. Indicators of inflammation in the diagnosis of mastitis. *Vet. Res.* **34**, 565-578 (2003).
45. de Jesus Rodriguez, B. *et al.* Identification in milk of a serum amyloid A peptide chemoattractant for B lymphoblasts. *BMC Immunol.* **10**, 4-2172-10-4 (2009).
46. Holst, B. D., Hurley, W. L. & Nelson, D. R. Involution of the bovine mammary gland: histological and ultrastructural changes. *J. Dairy Sci.* **70**, 935-944 (1987).
47. Wu, C. C., Howell, K. E., Neville, M. C., Yates, J. R., 3rd & McManaman, J. L. Proteomics reveal a link between the endoplasmic reticulum and lipid secretory mechanisms in mammary epithelial cells. *Electrophoresis* **21**, 3470-3482 (2000).
48. Suarez-Vega, A., Gutierrez-Gil, B. & Arranz, J. J. Estudio del perfil transcripcional de genes implicados en el metabolismo de los ácidos grasos de la leche en dos razas ovinas in XVI *Jornadas sobre Producción Animal*, Tomo II (ed. Asociación Interprofesional para el Desarrollo Agrario) 462-464 (INO Reproducciones S.A., 2015)



49. Boudreau, N., Simpson, C. J., Werb, Z. & Bissell, M. J. Suppression of ICE and apoptosis in mammary epithelial cells by extracellular matrix. *Science* **267**, 891-893 (1995).
50. Pullan, S. *et al.* Requirement of basement membrane for the suppression of programmed cell death in mammary epithelium. *J. Cell. Sci.* **109 ( Pt 3)**, 631-642 (1996).
51. Merto, G. R., Cella, N. & Hynes, N. E. Apoptosis is accompanied by changes in Bcl-2 and Bax expression, induced by loss of attachment, and inhibited by specific extracellular matrix proteins in mammary epithelial cells. *Cell Growth Differ.* **8**, 251-260 (1997).
52. Marino, R., Considine, T., Sevi, A., McSweeney, P. & Kelly, A. Contribution of proteolytic activity associated with somatic cells in milk to cheese ripening. *Int. Dairy J.* **15**, 1026-1033 (2005).
53. Silanikove, N., Shamay, A., Shinder, D. & Moran, A. Stress down regulates milk yield in cows by plasmin induced beta-casein product that blocks K<sup>+</sup> channels on the apical membranes. *Life Sci.* **67**, 2201-2212 (2000).
54. Shennan, D. B. & Peaker, M. Transport of milk constituents by the mammary gland. *Physiol. Rev.* **80**, 925-951 (2000).
55. Park, Y., Juárez, M., Ramos, M. & Haenlein, G. Physico-chemical characteristics of goat and sheep milk. *Small ruminant research* **68**, 88-113 (2007).
56. Maller, O., Martinson, H. & Schedin, P. Extracellular matrix composition reveals complex and dynamic stromal-epithelial interactions in the mammary gland. *J. Mammary Gland Biol. Neoplasia* **15**, 301-318 (2010).
57. Houdebine, L. M. Lactation: Comparative molecular physiology and molecular biology in *1st World Congress of Dairy products in human health and nutrition* (eds. Serrano-Ríos, M. et al) Ch. 1, 3-12 (CRC Press,1994).

## **Acknowledgements**

This work is included in the framework of the project AGL2012-34437 funded by the Spanish Ministry of Economy and Competitiveness (MINECO). BGG is funded through the Spanish “Ramón y Cajal” Programme (RYC-2012-10230) from the MINECO. ASV is funded by an FPU contract from Spanish Ministry of Education.

## **Author Contributions**

Conceived and designed the experiments: JJA. Performed the experiments: ASV BGG JJA.

Analyzed the data: ASV, GTK, CK, CRG. Wrote the paper: ASV, BGG, JJA.

## **Competing financial interests**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Tables

**Table 1.** RNA-Seq gene expression distribution for the two breeds and the four lactation time points studied in the present work.

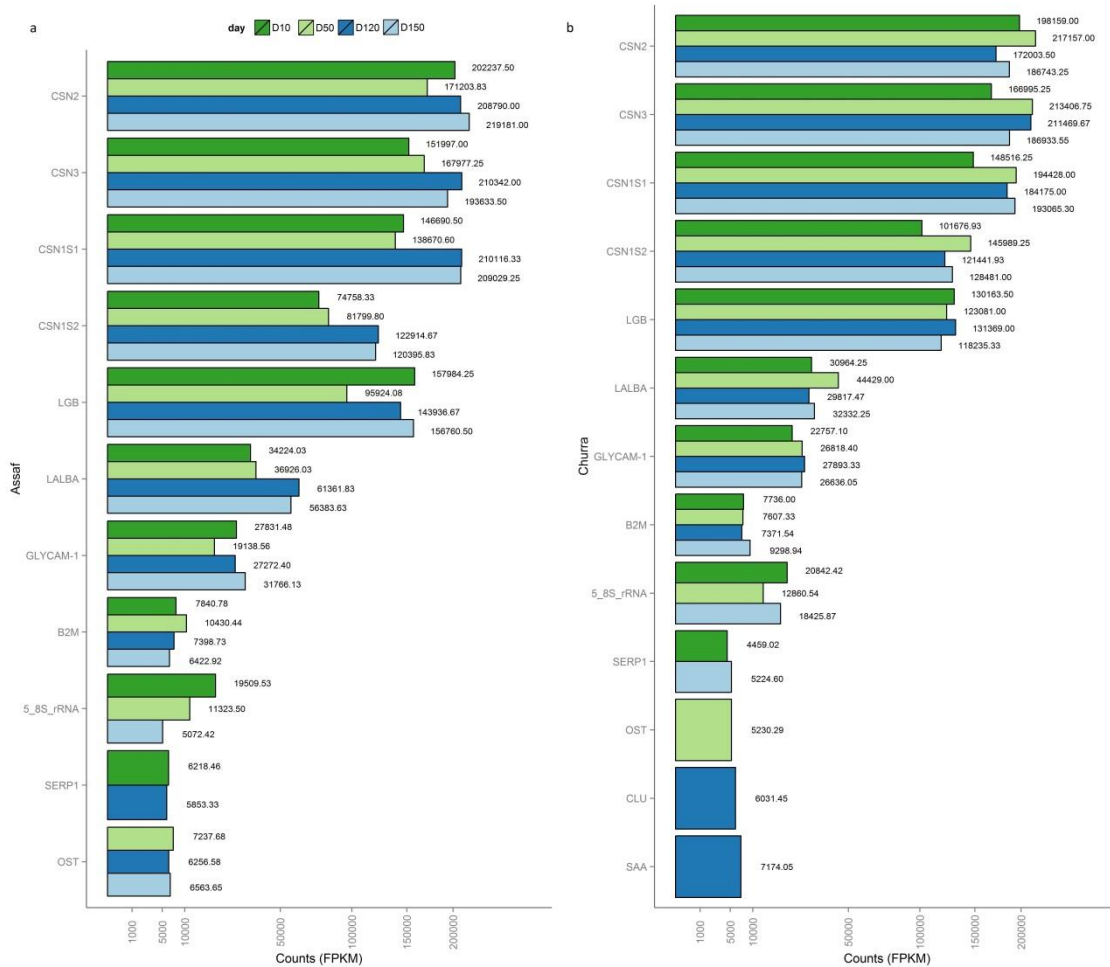
|   | <b>Churra</b> |            |             |             | <b>Assaf</b> |            |             |             |
|---|---------------|------------|-------------|-------------|--------------|------------|-------------|-------------|
|   | <b>D10</b>    | <b>D50</b> | <b>D120</b> | <b>D150</b> | <b>D10</b>   | <b>D50</b> | <b>D120</b> | <b>D150</b> |
| <b>Highly expressed genes<br/>(≥ 500 FPKM)</b>            | 129           | 109        | 116         | 111         | 129          | 141        | 107         | 100         |
| <b>Medium expressed genes<br/>(≥ 10 FPKM to 500 FPKM)</b> | 5982          | 4747       | 5238        | 4986        | 5977         | 5916       | 4664        | 4397        |
| <b>Lowly expressed genes<br/>(&lt; 10 FPKM)</b>           | 10688         | 12336      | 11393       | 11753       | 10914        | 11329      | 11448       | 11907       |
| <b>Total expressed genes</b>                              | 16799         | 17192      | 16747       | 16850       | 17020        | 17386      | 16219       | 16404       |
| <b>Non expressed genes</b>                                | 8397          | 8004       | 8449        | 8346        | 8176         | 7810       | 8977        | 8792        |

**Table 2.** Classification of the transcripts identified in the mammary gland samples in relation to the Ensembl annotated sheep genes (Oar\_v3.1) based on the Cuffcompare tool of Cufflinks.

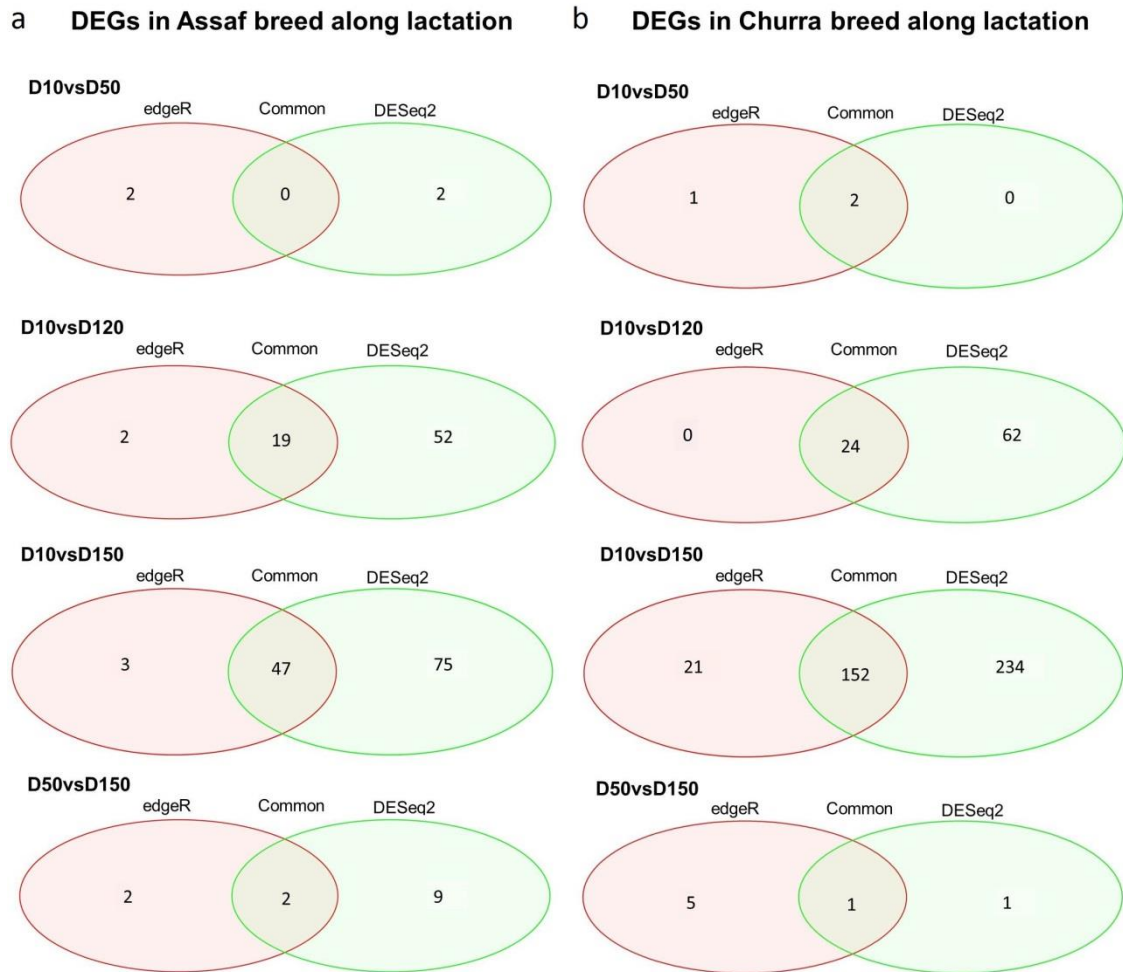
| <b>Cuffcompare_class</b>                         | <b>Number_transcripts</b> | <b>Percentage</b> |
|--|---------------------------|-------------------|
| Complete match of intro chain                    | 25833                     | 23.95             |
| Multiple classifications                         | 0                         | 0                 |
| Contained in the reference                       | 0                         | 0                 |
| Possible pre-mRNA fragment                       | 0                         | 0                 |
| Transcript falling within a reference intron     | 0                         | 0                 |
| Potentially novel isoforms                       | 12057                     | 11.18             |
| Generic overlap with a reference transcript      | 1021                      | 0.95              |
| Possible polymerase run-on fragment              | 0                         | 0                 |
| Intergenic transcript                            | 66739                     | 61.87             |
| Exonic overlap on opposite strand                | 2166                      | 2.01              |
| Repeat   | 0                         | 0                 |
| Overlapping intron transfrag in the other strand | 61                        | 0.06              |

## Figures

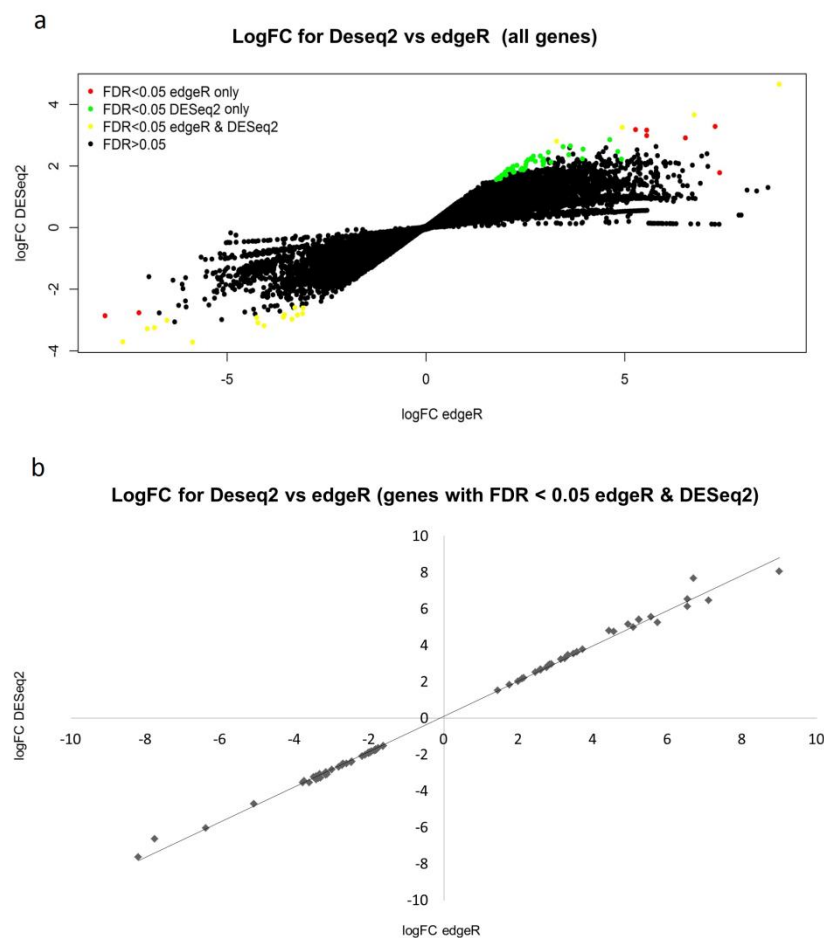
**Figure 1: Bar graph with the highly expressed genes in milk somatic cells at 10, 50, 120 and 150 days of lactation. FPKM values are represented in the X-axis, whereas the gene names are indicated in Y-axis. A color code is used to represent the four time points studied. (a) Top-10 highly expressed genes in Assaf. (b) Top-10 highly expressed genes in Churra.**



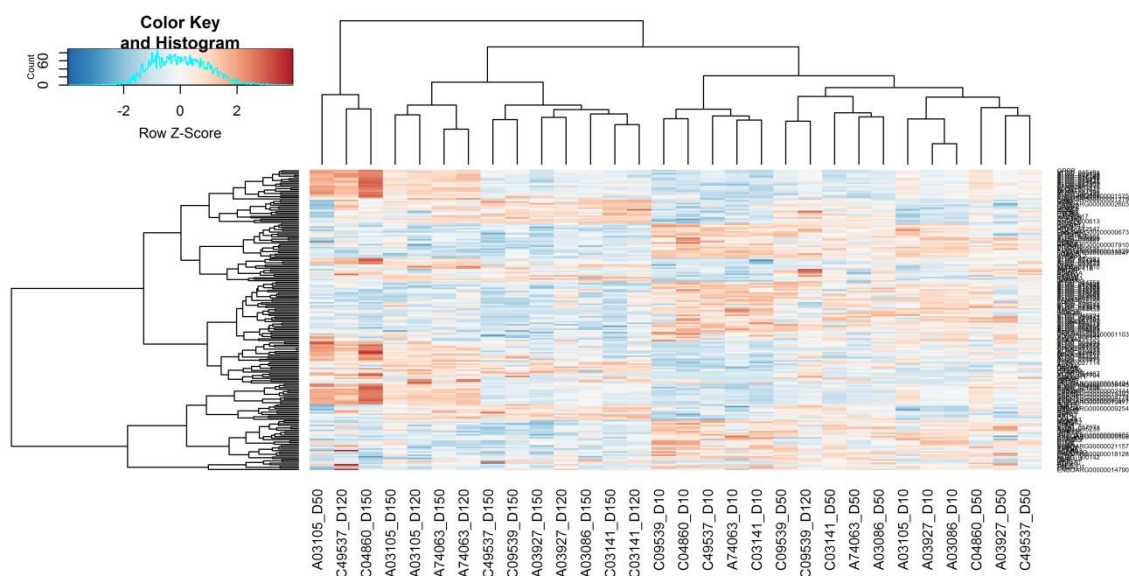
**Figure 2: Venn diagrams showing the number of differential expressed genes (DEG) across lactation for Churra and Assaf sheep.** For the two breeds (a, Assaf; b, Churra) and all the time point pair comparisons considered, the number of DEG identified by the edgeR analysis (red ellipse), the DESeq2 analysis (green ellipse) or both programs (intersection) is indicated.



**Figure 3: Comparison of differentially expressed genes (DEGs) detected with DESeq2 and edgeR in the Churra D10 vs. D150 contrast.** (a) Comparative graphical representation of the logFC values resulting from the differential expression analysis performed for the D10 vs. D150 contrast with the edgeR package (X-axis) and the DESeq2 package (Y-axis). To help the comparison, a color code is used to distinguish the following categories: logFC values of the genes exclusively detected as DEGs by edgeR (red); logFC values of the genes exclusively detected as DEGs by DESeq2 (green); logFC values of the genes detected as DEGs by both packages (yellow); logFC values of the genes not identified as DEG by any software (black). (b) Comparative graphical representation of the logFC values obtained for the genes commonly identified as DEGs by edgeR (X-axis) and DESeq2 (Y-axis) packages. Here it is shown that for these genes (colored in yellow in Fig. 3a) the expression levels (logFC values) measured by the two packages show a high and linear correlation.



**Figure 4: Heatmap and hierarchical clustering of differentially expressed genes (DEGs) across lactation for Assaf and Churra sheep breeds.** Heatmap display of supervised hierarchical clustering of the unique 191 DEGs identified across four time points of sheep lactation (D10, D50, D120, D150). The 191 genes are displayed in rows and the normalized counts per sample are displayed in columns. Each column represents a sample; sample names indicate the corresponding breed (A = Assaf; C = Churra) and day of sampling (D10, D50, D120, D150). A colour code indicates up-regulated (orange) and down-regulated (blue) expression levels,.







**Estudio del perfil transcripcional de genes implicados en el metabolismo  
de los ácidos grasos de la leche en dos razas ovinas.**

Aroa Suárez-Vega, Beatriz Gutiérrez-Gil y Juan José Arranz.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León,  
Campus de Vegazana s/n, León 24071, Spain.

XVI Jornadas sobre Producción Animal 2015, Tomo II (ed. Asociación Interprofesional  
para el Desarrollo Agrario) 462-464 (INO Reproducciones S.A.).



## ESTUDIO DEL PERFIL TRANSCRIPCIONAL DE GENES IMPLICADOS EN EL METABOLISMO DE LOS ÁCIDOS GRASOS DE LA LECHE EN DOS RAZAS OVINAS.

Suárez-Vega, A., Gutiérrez-Gil, B. y Arranz, J.J.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. E-mail del autor responsable [asuav@unileon.es](mailto:asuav@unileon.es)

### INTRODUCCIÓN

La producción mundial de leche de oveja es de unos 10 millones de toneladas anuales siendo la cuarta en importancia. España se sitúa como el séptimo productor mundial de leche de oveja (FAOSTAT, 2012), siendo la comunidad de Castilla y León la que genera un 70% de esta producción láctea (352.501.000 litros en 2011) (MAGRAMA, 2015). En general, la leche de oveja es procesada para la elaboración de productos lácteos, principalmente queso. Desde este punto de vista, el porcentaje de sólidos totales en leche o extracto quesero (grasa y proteína) adquiere una importancia relevante. La síntesis de lípidos en leche tiene especial interés debido a su influencia en el procesado y en las propiedades organolépticas que confieren al queso. Por ello, el conocimiento de los genes que desempeñan un papel relevante en la síntesis de grasa de la leche y de las variaciones en cuantitativas y cualitativas de los mismos a lo largo de la lactación en el ganado ovino, presentan un especial interés en la mejora de la producción lechera de dicha especie. Recientemente, las técnicas de secuenciación masiva paralela de RNA (RNA-seq) nos han permitido comparar los niveles de expresión génica entre distintos grupos de individuos y tejidos (Mortazavi et al., 2008; Wang et al., 2009). Para este estudio se han obtenido datos de RNA-seq de 8 animales pertenecientes a dos razas ovinas lecheras, Assaf y Churra, de las cuales se han tomado muestras de leche a lo largo de la lactación, en concreto los días 10, 50, 120 y 150 después del parto. Estas razas difieren en el porcentaje total de grasa en leche, un 6,65 % en Assaf y un 7,01 % en Churra (MAGRAMA, 2015). El metabolismo lipídico en la glándula mamaria puede subdividirse en: (i) captación de ácidos grasos del torrente circulatorio, (ii) síntesis *de novo* y desaturación de ácidos grasos y (iii) esterificación y secreción de la grasa mamaria. El número de genes que intervienen en estos procesos es elevado y sus interacciones complejas, para este trabajo inicial hemos analizado el perfil de expresión de 7 genes (*VLDLR*, *LPL*, *ACACA*, *FASN*, *SCD1*, *BTN1A1* y *XDH*) con influencia (Bionaz y Looor, 2008) en cada uno de los distintos procesos del metabolismo lipídico de la glándula mamaria a lo largo de la lactación en dos razas ovinas (Churra y Assaf).

### MATERIAL Y MÉTODOS

Para este estudio se han utilizado muestras procedentes de 8 ovejas libres de mamitis y sometidas a la misma dieta, 4 ovejas de raza Assaf y 4 ovejas de raza Churra. De cada animal se recogieron 50 ml de leche los días 10, 50, 120 y 150 de lactación. Las muestras fueron recogidas una hora después del ordeño de las 8 de la mañana, coincidiendo con el punto de máxima concentración de células somáticas en leche, según lo expuesto por Gonzalo et al. (1994). El RNA se extrajo a partir de 50 ml de leche siguiendo el protocolo descrito por Wickramasinghe et al. (2012), con ciertas modificaciones. La integridad del RNA (valor RIN) se analizó utilizando el *Bioanalyzer Agilent 2100* (Agilent technologies, Santa Clara, CA, USA). Los valores de RIN de las muestras de RNA oscilaron entre 7 y 9. Las genotecas de RNA de extremos pareados (*paired-end libraries*) con fragmentos de 300 pb se crearon con el kit *True-Seq RNA-Seq simple preparation v2* (Illumina, San Diego, CA, USA). Los fragmentos fueron secuenciados en el CNAG en un secuenciador *Illumina Hi-Seq 2000*, generando lecturas *paired-end* de 75 pb.

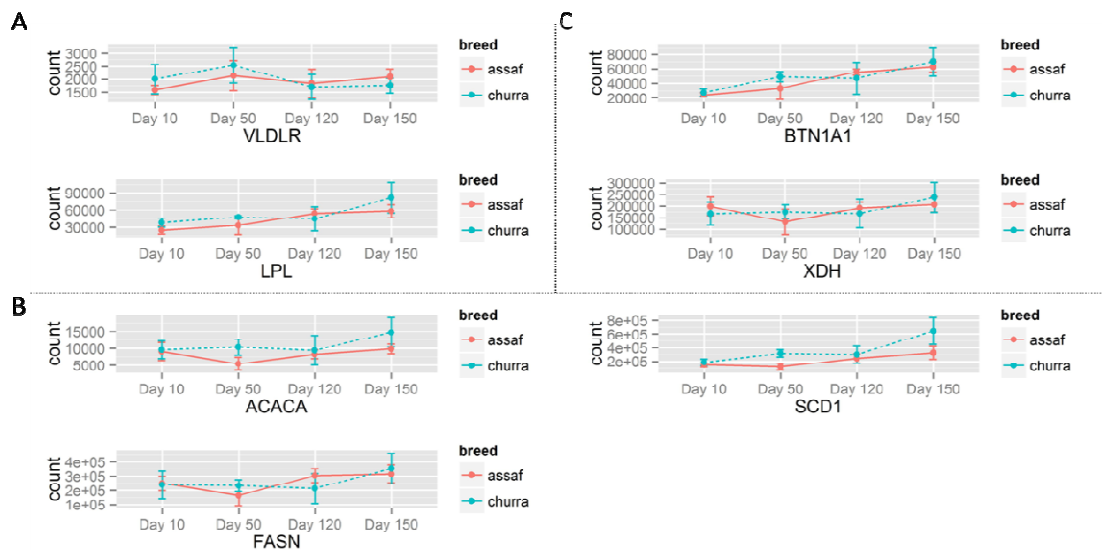
El control de calidad de los datos brutos de secuenciación se realizó utilizando el programa *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Las lecturas fueron mapeadas contra el genoma ovino. (OAR v.3.1) utilizando el programa *STAR\_2.3.0e* (Dobin et al., 2013). El paquete bioinformático *Cufflinks* (Trapnell et al., 2010) se utilizó para ensamblar las lecturas mapeadas y crear un archivo de anotación de referencia. La cuantificación del número bruto de lecturas por gen se llevó a cabo utilizando el programa *SigCufflinks* (<http://www.siggenae.org>).

El paquete de R *DESeq2* (Love et al., 2014), fue utilizado en el análisis de expresión diferencial. Nuestro experimento, está compuesto por 2 razas de animales (Churra y Assaf) y cuatro días diferentes de muestreo para cada grupo (Día 10, Día 50, Día 120 y Día 150).

Para el análisis de la expresión diferencial se utilizó el siguiente modelo:  $Y = raza + día + (raza \times día) + error$ . Una vez realizado el análisis con la función DESeq, se extrajeron los recuentos normalizados para los genes de interés con la función *plotCounts*. Para cada grupo de réplicas biológicas, se calculó la media, desviación estándar y error estándar para cada uno de los genes analizados.

## RESULTADOS Y DISCUSIÓN

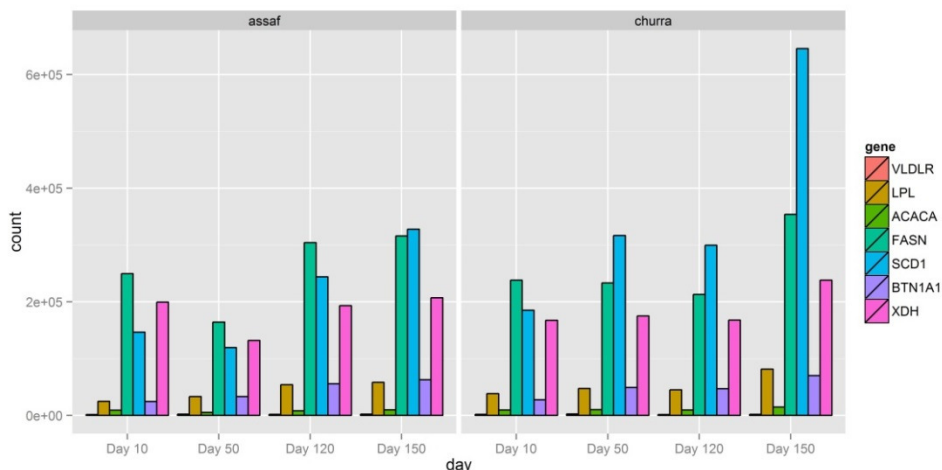
El nivel de expresión de los genes analizados es muy variable a lo largo de los diferentes momentos de la lactación. Ninguno de los genes seleccionados muestra diferencias significativas al comparar las distintas razas o los distintos puntos de lactación. Cuando nos fijamos en el perfil transcripcional de los genes analizados (Figura 1), podemos ver que, en general, la expresión en la raza Churra es ligeramente superior que en Assaf. Aunque cabe resaltar que la variabilidad entre los animales de raza Churra es también superior que entre las réplicas biológicas de la raza Assaf. La lipoproteína-lipasa (*LPL*) y el receptor de lipoproteína de muy baja densidad (*VLDLR*) están implicados en la captación de los ácidos grasos por la glándula mamaria. El *LPL* se expresa más que el *VLDLR* e incrementa su expresión a lo largo de la lactación. El gen *ACACA*, *FASN* y *SCD1* están relacionados con la síntesis *de novo* y la desaturación de ácidos grasos. Al igual que ocurre en el ganado bovino (Wickramasinghe et al., 2012), el gen *ACACA* tiene una expresión muy baja en relación con el gen *FASN*, sin embargo, ambos tienen un perfil de expresión paralelo. La expresión del *SCD1* aumenta progresivamente a lo largo de la lactación. La butirofilina (*BTN1A1*) y la Xantina deshidrogenasa (*XDH*) son proteínas implicadas en la formación del glóbulo de grasa (Bionaz y Loo, 2008). En el caso del ganado bovino la expresión de estos genes es muy elevada al inicio de la lactación y luego decrece (Wickramasinghe et al., 2012). En la oveja, en el caso de la *BTN1A1* se produce un incremento progresivo a lo largo de la lactación, la *XDH* se mantiene más o menos constante incrementándose ligeramente al final.



**Figura 1.** Perfiles de expresión de los genes implicados en el metabolismo lipídico de la glándula mamaria durante la lactación.

En general, la mayoría de los genes analizados incrementan su expresión a lo largo de la lactación, al contrario de lo descrito en el ganado bovino en la que la mayoría de los genes muestran su mayor expresión al inicio de la misma (Wickramasinghe et al., 2012). Como se puede observar en la figura 2, los tres genes con mayor expresión son *FASN*, *SCD1* y *XDH*. *FASN*, que codifica para la ácido graso sintasa, está implicado en la producción de ácidos grasos *de novo* y es el gen más expresado al comienzo de la lactación en las dos razas. La esteroil-CoA desaturasa 1 (*SCD1*), es la principal encima implicada en la biosíntesis de ácidos grasos mono-insaturados en rumiantes (Ntambi y Miyazaki, 2004). El *SCD1*, es el gen más expresado al final de la lactación en ambas razas, siendo también el gen más expresado en Churra los días 50 y 120 de lactación. En la glándula mamaria de rumiantes,

este gen se encarga de la producción de cerca del 80% de la forma más común de ácido linoleico conjugado (CLA, isómero cis-9, trans-11, C18:2) (Corl et al., 2001). Está demostrado que el consumo de ácidos linoleicos conjugados es beneficioso para la salud humana por sus propiedades anti-carcinogénicas, anti-lipogénicas e inmunomoduladoras (Pariza et al., 1999). La elevada expresión de la estearoil-CoA desaturasa en la glándula mamaria de las ovejas de raza Churra podría estar relacionada con una mayor concentración de CLA y, por lo tanto, con una leche potencialmente más saludable.



**Figura 2.** Recuentos normalizados de los genes implicados en el metabolismo lipídico de la glándula mamaria durante la lactación.

#### REFERENCIAS BIBLIOGRÁFICAS

- Bionaz, M. & Looor J.J, 2008. BMC Genomics. 9:366-2164-9-366.
- Corl, B.A. et al., 2001. J Nutr Biochem. 12(11):622-630.
- Dobin, A. et al., 2013. Bioinformatics. 29(1):15-21.
- Gonzalo, C. et al., 1994. J Dairy Sci. 77(7):1856-1859.
- Love, M.I. et al., 2014. Genome Biol. 15(12):550.
- Mortazavi, A. et al., 2008. Nat. Methods. 5(7):621-628.
- Ntambi, J.M. & Miyazaki, M., 2004. Prog. Lipid. Res. 43(2):91-104.
- Pariza, M.W. et al., 1999. Toxicol. Sci. 52(2 Suppl):107-110.
- Trapnell, C. et al., 2010. Nat. Biotechnol. 28(5):511-515.
- Wang, Z. et al., 2009. Nat. Rev. Genet. 10(1):57-63.
- Wickramasinghe, S. et al., 2012. BMC Genomics. 13:45-2164-13-45.

**Agradecimientos:** Este trabajo ha sido financiado por el proyecto AGL2012-34437 del Ministerio de Economía y Competitividad. B. Gutierrez-Gil es investigadora contratada a través del programa “Ramón y Cajal” del Ministerio de Economía y Competitividad.

#### STUDY OF THE TRANSCRIPTIONAL PROFILE OF GENES REALTED TO THE MAMMARY GLAND FAT METABOLISM IN TWO OVINE BREEDS.

**ABSTRACT:** Sheep milk ranks the fourth position in terms of global milk production from different species. Most of the world’s sheep milk is processed into dairy products, mostly cheese. Milk lipid synthesis as well as fatty acid esterification and milk fat secretion have a special interest due to their influence in manufacturing and organoleptic properties in dairy products. In this study we performed a RNA-seq analysis in two different dairy sheep breeds in order to evaluate the transcriptional profile of some genes implicated in the mammary gland fat metabolism along lactation. The selected genes are: *LPL* and *VLDL*, implicated in the fatty acid uptake from blood; *ACACA*, *FASN* and *SCD*, implicated in de novo synthesis and fatty acid desaturation; *BTN1A1* and *XDN* implicated in lipid droplet formation. None of the genes evaluated is significantly differentially expressed between the two breeds nor along lactation, but we can see changes in the expression profile. Mostly all of the genes analyzed increased their expression along lactation. *SCD1*, *ACACA* y *XDH* were the most abundant key genes measured, appearing to be key genes in milk fat metabolism.

**Keywords:** sheep, RNAseq, milk, fat metabolism.





## **5. Resumen de Resultados y Discusión General**





El estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino se ha visto favorecido por la rápida evolución de la genómica en los últimos años. Las técnicas de secuenciación de segunda generación han permitido la secuenciación del genoma de distintas especies domésticas (vaca, pollo, cerdo, oveja...) y la identificación de gran variabilidad a lo largo de estos genomas. Este hecho ha posibilitado el desarrollo de herramientas como las plataformas de genotipado (chips de SNPs) de media y de alta densidad. Así mismo, la disminución de los costes de secuenciación experimentados por las técnicas de secuenciación de segunda generación ha permitido su utilización en ámbitos más amplios y por grupos de investigación más modestos. Estos dos hitos en la evolución de la genómica de las especies domésticas han marcado las actividades realizadas en la presente Tesis Doctoral. Para su realización nos hemos servido fundamentalmente de dos herramientas genómicas, el chip ovino de media densidad (*Ovine SNP50 BeadChip*) y técnicas de secuenciación de segunda generación. Estas herramientas genómicas han sido utilizadas en el estudio de dos tipos de caracteres: caracteres de herencia mendeliana (enfermedades monogénicas), en los que se han utilizado el chip de media densidad ovino y técnicas NGS; y caracteres complejos, relacionados con la producción láctea, en los que se ha utilizado la secuenciación masiva paralela del RNA (RNA-Seq).

En esta apartado se presenta un resumen de los principales resultados obtenidos en la presente Tesis Doctoral y la discusión de los mismos. Para ello hemos dividido la sección en dos subapartados generales, el primero referido a la utilidad de las herramientas genómicas sobre el estudio de enfermedades monogénicas y el segundo referido a la utilidad de la técnica de RNA-Seq sobre el estudio de la glándula mamaria de la oveja durante la lactación.

## **5.1. Utilización de herramientas genómicas en el estudio de la base genética de enfermedades monogénicas**

### *5.1.1. Estudio de la base genética de la lisencefalia con hipoplasia cerebelar en el ganado ovino de raza Churra*

El primer trabajo presentado en esta Tesis Doctoral describe los resultados del estudio de la base genética de la lisencefalia con hipoplasia cerebelar (*Lisencephaly with*

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

*Cerebellar Hypoplasia* (LCH)) en el ganado ovino de raza Churra. La lisencefalia, cuyo significado literal es “cerebro liso”, engloba un grupo de malformaciones severas del cerebro causadas por una migración anormal de las neuronas post-mitóticas al córtex en desarrollo (Barkovich *et al.*, 2005). Macroscópicamente se observa una falta del desarrollo de las circunvoluciones cerebrales. En el caso de la LCH estas malformaciones a nivel del córtex cerebral vienen acompañadas de una disminución acusada del tamaño del cerebelo. Los corderos enfermos cursan con incoordinación de movimientos e incapacidad para mantenerse en la estación, lo que les dificulta la toma de alimento y mueren a las pocas horas de nacer.

El estudio del pedigrí de los animales afectados nos permitió determinar que nos encontrábamos ante una enfermedad genética compatible con una herencia autosómica recesiva. Además, permitió identificar el antecesor común de los corderos afectados objeto de estudio (HKLD8633). Este macho había nacido en 1986 y la enfermedad se diagnosticó por primera vez en el año 2004, aproximadamente, cuatro generaciones más tarde.

Para caracterizar la base genética de la LCH en la raza Churra se realizó un genotipado con el chip de media densidad ovino (*Ovine SNP50 BeadChip*) en un total de 40 animales, 20 controles sanos no relacionados y 20 animales pertenecientes al rebaño afectado, siete de ellos enfermos (casos). Mediante un estudio de asociación a nivel genómico (GWAS) realizado con el software *PLINK* (Purcell *et al.*, 2007) con los datos de genotipado obtenidos con el chip, se identificó una fuerte asociación en el cromosoma cuatro de la oveja (OAR4). El SNP que mostraba la asociación más significativa ( $p\text{-value } 2,4 \times 10^{-4}$ ) fue OAR4\_45088426, localizado en la posición 42.810.217 pb del cromosoma 4 en el genoma de referencia ovino Oar\_v3.1. El análisis de identidad por descendencia (*homozygosity mapping*) permitió acotar la región de estudio a un intervalo de 4,8 Mpb en el OAR4 (42,369-47,251 Mpb). El análisis de la funcionalidad de los genes localizados en el intervalo de homocigosis nos permitió identificar al gen *RELN* como el mejor candidato posicional y funcional. El gen *RELN* juega un papel crítico en la migración neural y la formación de las capas cerebrales

(Frotscher, 2010), relacionándose con un determinado tipo de lisencefalia en humanos, la lisencefalia tipo 2 o síndrome *Norman-Roberts* (Hong *et al.*, 2000).

La técnica de secuenciación *Sanger* aplicada sobre el mRNA en un control sano nos permitió obtener la secuencia de cDNA completa del gen *RELN* en el ganado ovino, que ha sido depositada en *GenBank* (nº acc. *KC590614*). La secuenciación del cDNA de este gen en un cordero enfermo y su comparación con el control sano reveló una delección de 31 pb en el exón 36 del gen *RELN* del cordero afectado (*c.5410\_5440del*). Esta mutación se encontraba en homocigosis en el animal enfermo pero no en el control. La delección (*c.5410\_5440del*) fue confirmada mediante el genotipado de DNA procedente de seis corderos enfermos de LCH, seis controles no relacionados y seis madres de corderos afectados de LCH. En todos los casos los animales afectados resultaron ser homocigotos para la delección; las madres fueron heterocigotas para la mutación; y los controles homocigotos para el alelo salvaje. En el gen *RELN* se han descrito varias mutaciones relacionadas con la LCH en humanos (Hong *et al.*, 2000) y roedores (Takahara *et al.*, 1996; Royaux *et al.*, 1997; D'Arcangelo and Curran, 1998; Yokoi *et al.*, 2003). La mutación detectada (*c.5410\_5440del*) originaría un cambio en el marco de lectura de la secuencia codificante del gen *RELN* provocando un codón de terminación prematuro en la posición 1.817 de la proteína, frente a la proteína normal que constaría de 3.460 aminoácidos. Sin embargo, las pruebas funcionales realizadas (*qRT-PCR* y *Western blot*) en animales enfermos revelaron que la mutación provocaba una disminución en la cantidad de RNA y la ausencia de la proteína en el tejido cerebral.

La caracterización de la mutación causal de la LCH en el ganado ovino de raza Churra permitió el desarrollo y registro como patente de un método de diagnóstico rápido para la identificación de portadores sanos. Esta herramienta de diagnóstico es útil en la implementación de la selección asistida por genes en los rebaños de raza Churra y por tanto en el control de la enfermedad. Dicha identificación se realiza amplificando por PCR un fragmento de DNA que contiene la mutación; para ello se han diseñado dos cebadores específicos que se utilizan para el diagnóstico (*Sense PCR primer: ttgccttctccggtttaatg/ Antisense PCR primer: agggattgtgatgctggac*). Una vez realizada la amplificación, la migración del amplicón en un gel de agarosa al 2% durante 45 minutos nos permite discernir la longitud de los fragmentos amplificados e identificar

### **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

animales homocigotos para la delección (*del/del*), heterocigotos (*del/+*) u homocigotos para el alelo salvaje (*+/+*).

#### *5.1.2. Estudio de la base genética de la epidermólisis bullosa juntural en el ganado ovino de raza Churra*

En el estudio de la base genética de la epidermólisis bullosa se combinó la utilización de dos herramientas genómicas, el genotipado con el chip de media densidad ovino junto con técnicas de secuenciación de segunda generación. Ambas aproximaciones han permitido identificar una delección de 4 pb en el exón 33 del gen *ITGB4* como la mutación causal responsable de la epidermólisis bullosa juntural (*Junctional Epidermolysis Bullosa* (JEB)) el ganado ovino de raza Churra.

La epidermólisis bullosa es una genodermatosis autosómica recesiva que cursa con la formación de ampollas y la fragilidad mecánica de la piel y membranas mucosas, manifestándose al nacimiento. Las lesiones ampollosas en la JEB se producen a nivel de la unión dermo-epidérmica en concreto a nivel de la lámina lúcida, con alteración severa de los hemidesmosomas (Fine *et al.*, 2008). En el núcleo de selección de ANCHE se detectaron dos rebaños en los que nacían corderos que presentaban erosiones y úlceras en la piel y membranas mucosas y en los que se diagnosticó JEB (Benavides *et al.*, 2015). El estudio del pedigrí de los animales afectados nos permitió determinar el modo de herencia, compatible con una enfermedad autosómica recesiva, así como identificar un antecesor común entre los dos rebaños afectados (VS94.11) (Benavides *et al.*, 2015). La JEB ha sido descrita en varias especies de mamíferos (Peters *et al.*, 2015; Olivry *et al.*, 1997; Spirito *et al.*, 2002; Milenkovic *et al.*, 2003; Alhaidari *et al.*, 2005; Capt *et al.*, 2005; Graves *et al.*, 2009; Medeiros *et al.*, 2012). En la oveja, la JEB (subtipo Herlitz) ha sido previamente descrita en un rebaño de German Black-headed Mutton; en este caso la enfermedad se producía por una delección de 2 pb en el exón 18 del gen *LAMC2* (Momke *et al.*, 2011).

El procedimiento llevado a cabo para determinar la región genómica portadora de la mutación causal fue el mismo que para la LCH. El chip ovino de media densidad (*Ovine SNP50 BeadChip*) se utilizó para genotipar un total de 96 animales, 48 controles sanos

no relacionados y 48 animales de los rebaños afectados, de los cuales 20 eran corderos enfermos. Al igual que en la LCH, los análisis GWAS y de identidad por descendencia realizados con los datos de genotipado nos permitieron acotar la localización de la mutación causal a un intervalo de homocigosis de 868 Kpb (54.632.309-55.500.100pb) en el OAR11. El gen *ITGB4*, localizado dentro del intervalo de homocigosis, se seleccionó como mejor candidato posicional y funcional, debido a las referencias bibliográficas que asocian mutaciones en el gen *ITGB4* con diversas formas de JEB en humanos (Carmi *et al.*, 1982; Chavanas *et al.*, 1999; Birnbaum *et al.*, 2008). El gen *ITGB4* codifica para la proteína integrina  $\beta 4$ . En el epitelio estratificado esta proteína se localiza en los hemidesmosomas, estructuras de adhesión especializadas de la unión dermo-epidérmica (Stepp *et al.*, 1990).

La mala calidad de la secuencia de referencia y de la anotación del gen *ITGB4* (ENSOARG0000009764) en el genoma ovino disponible (Oar\_v3.1), hizo que recurriésemos a la secuenciación *de novo* del gen *ITGB4* a partir de tres clones BAC ovinos (CH243-344E15, CH243-306J7 y CH243-504K6) (Ratnakumar *et al.*, 2010).

La técnica de secuenciación masiva paralela del RNA (RNA-Seq) se utilizó para obtener el transcriptoma de dos muestras de piel, la de un control sano y la de un cordero enfermo. La técnica de RNA-Seq permite obtener información del transcriptoma de un tejido celular en ausencia de una buena anotación genómica. La combinación de la secuencia genómica del *ITGB4* (obtenida mediante la secuenciación de los clones BAC) y de la información generada mediante la técnica de RNA-Seq en el control sano permitió obtener la secuencia anotada del gen *ITGB4* ovino, que ha sido depositada en *GenBank* (nº acc. KP025765)

Una de las posibles aplicaciones de la técnica de RNA-Seq es la detección de variantes en regiones codificantes del genoma (Wang *et al.*, 2009). En este estudio, nos servimos de esta utilidad de la técnica para detectar la mutación causal responsable de la JEB en el ganado ovino, una delección de 4 pb en el exón 33 del gen *ITGB4* (*c.4412\_4415 del*). Esta mutación se confirmó a nivel genómico, identificando tres genotipos que segregaban claramente con los tres posibles estatus de la enfermedad: los animales homocigotos para la delección (animales enfermos), los animales

### ***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

heterocigotos portadores (madres y padres de corderos afectados) y los homocigotos para el alelo salvaje (identificados en los controles sanos analizados).

Las consecuencias de la mutación se evaluaron mediante pruebas funcionales (*qRT-PCR* e inmunohistoquímica), en ellas se comprobó una disminución en los niveles de RNA y la ausencia de integrina  $\beta 4$  en las muestras de piel de los corderos afectados. Al realizar microscopia electrónica de transmisión se pudo comprobar que el tejido de animales enfermos también carecía de hemidesmosomas (Benavides *et al.*, 2015). Estudios realizados en queratinocitos en los que no se expresa la integrina  $\beta 4$  y en ratones *knock-out* para esta proteína confirman que la presencia de integrina  $\beta 4$  es esencial para la formación de los hemidesmosomas (Schaapveld *et al.*, 1998; Van der Neut *et al.*, 1996).

Al igual que en el caso de la LCH, se desarrolló y patentó un método de diagnóstico rápido para la detección de portadores sanos y control de la enfermedad en los rebaños de raza Churra. El genotipado de la mutación se realiza mediante la amplificación por PCR de un fragmento del gen que contiene la mutación, para ello se utilizan dos cebadores (*Sense PCR Primer: ccagcagtcagggagtg/ Antisense PCR Primer: ctcaccgccgttcagcag*), uno de ellos marcado con un fluoróforo. Posteriormente, se realiza un análisis de longitud de fragmentos en un secuenciador de electroforesis capilar, en el que se pueden identificar, en función del tamaño de los fragmentos amplificados, los tres genotipos: homocigotos para la delección (*del/del*), heterocigotos (*del/+*) u homocigotos para el alelo salvaje (*+/+*).

#### ***5.1.3. Utilidad de las herramientas genómicas en la caracterización de enfermedades monogénicas***

Los resultados de los dos trabajos aquí presentados en los que se identifican las mutaciones responsables de la lisencefalia con hipoplasia cerebelar y la epidermólisis bullosa juntural en la raza Churra apoyan resultados previos en diferentes especies ganaderas en las que el uso de metodologías genómicas ha tenido importantes consecuencias prácticas permitiendo el rápido control de defectos recesivos

emergentes (Charlier *et al.*, 2008; Nicholas, 2010), lo que deriva en mejoras tanto a nivel productivo como a nivel de bienestar animal.

Estudios previos en nuestro grupo de investigación evaluaban la utilización del chip de SNPs de media densidad como herramienta para el mapeo fino de regiones con influencia sobre caracteres productivos complejos (García-Gamez *et al.*, 2012a). En este estudio se concluía que sería necesario duplicar el número de marcadores para el mapeo de caracteres cuantitativos de interés (García-Gamez *et al.*, 2012b). Sin embargo, al contrario que ocurre con los caracteres complejos y tal y como se refleja en la presente Tesis Doctoral, las plataformas de genotipado de densidad media son altamente eficaces en el mapeo de regiones implicadas en caracteres de herencia monogénica, siendo innecesaria la utilización de chips de mayor densidad que aumentarían los costes de genotipado de los animales.

Un aspecto importante a señalar es que, en nuestra experiencia, el intervalo de confianza obtenido mediante la aproximación de *homozygosity mapping* o mapeo por “identidad por descendencia” depende del número de casos que se analizan. De esta forma, en el mapeo de la LCH dónde únicamente se han genotipado siete animales afectados se ha obtenido intervalo de homocigosis de 4,8 Mpb. Por otro lado, en el caso de la JEB, dónde se han analizado 20 casos, el intervalo de homocigosis obtenido se redujo a 850 Kpb. En este caso se puede observar que el análisis de cuatro veces más animales hace que el intervalo de confianza se reduzca unas 5,5 veces. Estas observaciones, nos indican que aunque la aproximación GWAS es muy eficiente en la detección de la región que muestra desequilibrio de ligamiento con la mutación causal en la población, el tamaño de la región de homocigosis dónde se deben buscar los genes candidatos responsables depende del número de sucesos recombinantes (meiosis) que se analizan desde el antecesor común en el que se produjo la mutación.

Otro punto que hay que tener en cuenta es la calidad de la anotación de la región de búsqueda y el conocimiento sobre la fisiología y funcionalidad de los genes que se localizan en esa región. En las dos enfermedades analizadas en la presente Tesis Doctoral, en el primer análisis de la funcionalidad de los candidatos posicionales se identificó un gen candidato funcional que poseía un papel predominante en relación

con la enfermedad debido a estudios previos realizados en otras especies, principalmente en la especie humana. En otras ocasiones puede ocurrir que, por una deficiente anotación o por ser una enfermedad que no se ha estudiado previamente en profundidad se deban analizar varios de los genes candidatos posicionales. En este caso y en función de la longitud de la región de homocigosis, una estrategia que implique la secuenciación genómica de un caso y un control puede ayudar a identificar la mutación causal (Taylor *et al.*, 2014; Sartelet *et al.*, 2015).

Las técnicas de secuenciación de segunda generación han demostrado ser herramientas útiles en la secuenciación genómica de alto rendimiento, permitiendo la identificación de variantes genéticas causales (Ng *et al.*, 2010). La secuenciación masiva paralela de exomas es la aproximación más extendida en la identificación de la base genética de enfermedades monogénicas, presentándose como una alternativa económica a la secuenciación de genomas completos (Daetwyler *et al.*, 2014). En seres humanos, estudios recientes destacan las ventajas de la utilización de la secuenciación masiva paralela de exomas en el diagnóstico genético de las distintas clases de epidermolísis bullosa (Poulter *et al.*, 2014; Takeichi *et al.*, 2014). De la misma forma, en base a los resultados descritos en la presente Tesis Doctoral, la secuenciación masiva paralela de transcriptomas sería una alternativa eficaz a la secuenciación de genomas completos en aquellas especies con un genoma de referencia en desarrollo, permitiéndonos además conocer el grado de expresión génica.

En los dos estudios de enfermedades monogénicas presentados en esta Tesis Doctoral se han seguido dos aproximaciones distintas, una más clásica (mapeo con el chip de SNPs de media densidad y secuenciación *Sanger*) y otra más innovadora en la que se combinan el genotipado con el chip de SNPs con técnicas NGS. El genotipado con el chip de SNPs permite mapear la región con influencia sobre el carácter mendeliano, la secuenciación con la técnica de RNA-Seq permite la detección de variantes en la región mapeada, pero además, en zonas en las que el genoma de referencia tiene limitaciones, la secuenciación *de novo* con clones BAC para obtener una buena referencia se hace fundamental. Estas aproximaciones mixtas son altamente eficaces en genomas en los que tanto la secuencia como la anotación del genoma de referencia



no están muy avanzadas. La combinación de ambas técnicas supone un ahorro fundamentalmente en tiempo laboratorial. Además, el coste económico comparado de las dos aproximaciones sería similar.

El fin último de la detección de mutaciones causales responsables de enfermedades monogénicas es el control de la aparición de enfermedades genéticas emergentes. La aplicación práctica de la detección de las mutaciones causales en las dos enfermedades caracterizadas en la presente Tesis Doctoral (LCH y JEB) se ve reflejada en las herramientas de diagnóstico desarrolladas y que han sido objeto de patente. Estas herramientas permiten el diagnóstico en animales con síntomas compatibles y facilitan la aplicación de la selección asistida por genes en el núcleo de selección de ANCHE. La detección de mutantes heterocigotos permite la eliminación de estos animales del programa de selección, ayudando a los criadores a evitar apareamientos de riesgo en sus rebaños. Así mismo, los métodos de diagnóstico desarrollados podrían utilizarse en la detección de portadores en el caso de cruzamientos programados para obtener animales que sirvan como modelo en el estudio de ambas enfermedades.

## **5.2. Utilización de la técnica RNA-Seq en el estudio de la lactación en el ganado ovino**

Uno de los objetivos de la presente Tesis Doctoral ha sido el análisis global del transcriptoma mamario en el ganado ovino mediante la técnica de RNA-Seq. En este estudio se han comparado dos razas de ovejas lecheras con distintas características productivas, Churra y Assaf. El planteamiento del trabajo es doble, por un lado ver las diferencias de expresión a lo largo de la lactación en el ganado ovino y en segundo lugar analizar las diferencias de expresión entre dos razas con distintas características productivas y lactación de distinta duración.

En el estudio del perfil de expresión génica de la glándula mamaria de la oveja en lactación se obtuvieron datos de RNA-Seq de ocho animales, cuatro de cada raza, de los cuales se tomaron muestras de leche a lo largo de la lactación, en concreto los días 10, 50, 120 y 150 postparto. El RNA se extrajo a partir de las células somáticas de la leche (CSL). En relación a la expresión génica, las CSL han demostrado ser representativas del tejido mamario en lactación (Canovas *et al.*, 2014). Tratándose,

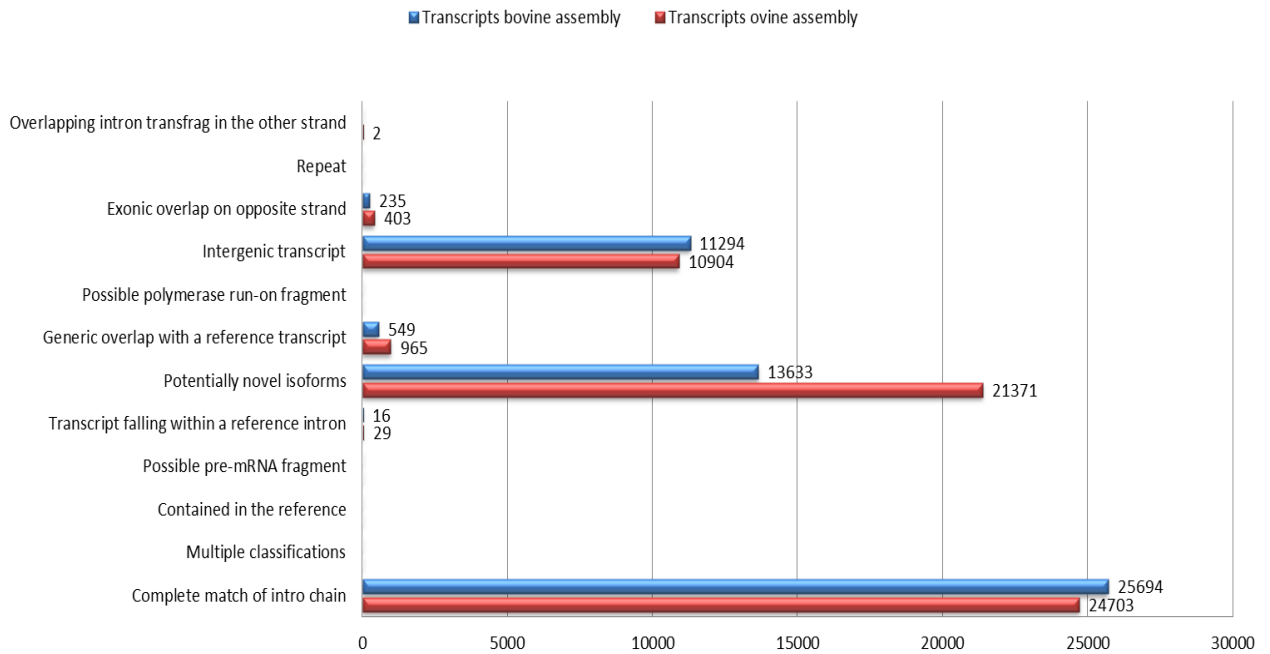
**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

además, de un método sencillo y no invasivo de obtención de RNA, lo que lo hace especialmente interesante en estudios dinámicos de la glándula mamaria en la que se realizan tomas de muestras por animal seriadas en el tiempo (Boutinaud and Jammes, 2002).

En los análisis RNA-Seq basados en el alineamiento de los datos frente al genoma de referencia, la calidad de la secuencia del genoma y su anotación es esencial para obtener una representación óptima de los transcriptomas. Por ello, en un estudio previo se hizo un análisis comparativo en el que se alinearon los datos de RNA-seq de las dos razas en uno de los puntos de lactación (D120) frente a los genomas ovino (Oar\_v3.1) y bovino (UMD 3.1), presentando este último un grado de anotación mucho más avanzado. En el caso del alineamiento con el genoma ovino, el 87,7% de las lecturas fueron alineadas. Cuando se utilizó como referencia el genoma bovino alinearon un 72,8% de las lecturas. El ensamblaje de los transcritos para los dos alineamientos demuestra que la utilización como referencia del genoma ovino permite la detección de un mayor número de nuevas isoformas potenciales, siendo el número de transcritos cuyos intrones concuerdan completamente y el de transcritos intergénicos muy similar al usar uno u otro genoma de referencia (Figura 3). Estos datos nos llevaron a concluir que la utilización del genoma Oar\_v3.1 como referencia era la opción más apropiada para realizar el análisis global de los datos de RNA-seq en el ganado ovino durante la lactación.

En el análisis del transcriptoma global de la oveja durante la lactación se detectaron un alto porcentaje de transcritos intergénicos (61,87 % de todos los transcritos ensamblados). Este hecho nos da una idea de la anotación incompleta de la versión actual del genoma ovino (Oar\_v3.1). En la actualidad, proyectos como el FAANG (*Functional Annotation of Animal Genomes*) (<http://www.faang.org/>), están siendo desarrollados con el objetivo de mejorar la anotación del genoma e identificar todos los elementos funcionales en los genomas de diferentes especies animales, entre ellos el de la oveja.

**Figura 3.** Histograma de la distribución de los transcritos ensamblados utilizando como referencia el genoma bovino (UMD 3.1) y el genoma ovino (Oar\_v3.1).



En el análisis realizado en base a los genes anotados en el genoma ovino Oar\_v3.1, se detectó la expresión de una media de 16.897 y 16.757 genes en las CSL de Churra y Assaf, respectivamente, lo que representa un 67,06% y un 66,51% del total de genes anotados en la versión actual del genoma. La mayoría de los genes expresados en las CSL (más de 10.000 genes en cada punto de muestreo y raza) tienen una expresión baja, de menos de 10 FPKM (Fragmentos por Kilobase de exón por millón de fragmentos mapeados). Para las dos razas estudiadas, los diez genes más expresados en cada punto de muestreo acumulan aproximadamente el 70% del total de FPKM, lo que significa que un pequeño número de genes acumula un porcentaje muy alto de las lecturas del RNA total extraído de las CSL. Estas observaciones concuerdan con el hecho de que estamos realizando un estudio sobre un órgano, la glándula mamaria, que posee una maquinaria altamente especializada propia del estado productivo en que se encuentra, la lactación, cuyas funciones principales son las síntesis de proteínas (en las células del alveolo mamario) y la secreción de grasa (captación y síntesis de

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

*novo*). Nuestros resultados muestran que el perfil de genes altamente expresados en las dos razas es muy similar a lo largo de la lactación. Al analizar los 10 genes más expresados en las dos razas y en todos los puntos de lactación nos encontramos que los genes que codifican para las caseínas (*CSN1S1*, *CSN1S2*, *CSN2* y *CSN3*) y las proteínas del lactosuero (*LGB* y *LALBA*), principales componentes proteicos de la leche, se encuentran siempre en las primeras posiciones de este *top-10*. Otros dos genes altamente expresados en las dos razas y en los cuatro puntos de la lactación analizados son el gen *GLYCAM-1*, componente de la membrana del glóbulo graso de la leche (Kann *et al.*, 1999) y el gen *B2M*, componente estructural de receptores implicados en la transferencia de inmunoglobulinas G del suero a la leche (Anderson *et al.*, 2006). Tres genes se incluían en el *top-10* en las dos razas pero no en todos los puntos de lactación, estos fueron los genes *5.8S rRNA*, *SERP1* y *OST*. El *5.8S rRNA* es un RNA no codificante que forma parte del ribosoma. Estudios realizados en la glándula mamaria de conejos relacionan la expresión de rRNA con niveles altos de síntesis proteica (Teyssot and Houdebine, 1981). El gen *SERP1* codifica para la proteína del retículo endoplásmico asociada al estrés 1, la síntesis de esta proteína aumenta asociada al estrés metabólico, suprimiendo la agregación o degradación de proteínas integrales de membrana recién sintetizadas y favoreciendo su glicosilación cuando el estrés se elimina (Yamaguchi *et al.*, 1999). El gen *OST* (también *SPP1*) codifica para la osteopontina, cuya función en la glándula mamaria no está claramente determinada. Existen diversos estudios que intentan explicar la función del gen *OST* relacionando su expresión con aumentos en la síntesis de las caseínas *CSN2* y *CSN3* (Sheehy *et al.*, 2009), con la morfogénesis de la glándula mamaria (Nemir *et al.*, 2000) o con la transmisión de propiedades inmunitarias al recién nacido (Alain *et al.*, 2009; Schack *et al.*, 2009). Además, en la lista de los diez genes más expresados se identificaron dos genes en Churra el día 120 que no se detectaron en Assaf, estos fueron los genes *CLU* y *SAA*. En la glándula mamaria de ratón, el gen *CLU* se encuentra altamente expresado en el parto y la involución, estando altamente relacionado con procesos de morfogénesis tubuloalveolar y diferenciación de células epiteliales (French *et al.*, 1996). Niveles elevados de expresión del gen *SAA* se relacionan con estados inflamatorios de la glándula mamaria (Pyorala, 2003); sin embargo estudios realizados

en glándulas mamarias sanas de cerdos indican que esta proteína puede tener una función fisiológica normal en la glándula mamaria sana estimulando la respuesta inmune neonatal (de Jesus Rodriguez *et al.*, 2009). Como contraste a nuestros resultados, los datos publicados en el ganado vacuno indican que al final de la lactación las caseínas y las proteínas del lactosuero no se encuentran entre los genes más expresados siendo sustituidos por otros genes relacionados con fenómenos de proteólisis, actividad antiapoptótica y funciones inmunes (Wickramasinghe *et al.*, 2012).

En el estudio de expresión diferencial realizado entre los distintos puntos de muestreo a lo largo de la lactación, el mayor número de genes diferencialmente expresados en las dos razas se detectó entre el inicio y el final de la lactación (días 10 y 150). En total se detectaron 89 genes anotados expresados diferencialmente en Churra y 38 en Assaf. Los análisis de enriquecimiento funcional para los genes sobreexpresados en el día 10 de la lactación en la raza Churra se relacionaban con procesos de respuesta a estímulos, proliferación y diferenciación celular y regulación positiva de fosforilación de proteínas. Procesos que están relacionados con los altos niveles de producción de las primeras etapas de la lactación. En el caso de los genes sobreexpresados en el día 150 de lactación se encuentran términos relacionados con la remodelación de la glándula mamaria (procesos de desarrollo, morfogénesis de órganos, movimiento y migración celular). En el caso de la raza Assaf, sólo se encontró enriquecimiento en términos de ontología de genes en los genes sobreexpresados en el día 150. Estos términos, al igual que en Churra, estaban relacionados con procesos de desarrollo celular, de tejidos, diferenciación celular y morfogénesis. En el caso del ganado ovino al contrario de lo que cabría esperar en un punto de muestreo tardío como es el día 150, no se encuentran términos de ontología génica relacionados con procesos de apoptosis y degeneración tisular. En el ganado vacuno, un estudio realizado por (Holst *et al.*, 1987) describe que a nivel tisular los cambios en la glándula mamaria al final de la lactación se reflejan más como un cambio en el estado de secreción que en una degeneración extensiva del tejido. Sin embargo, en el estudio global de expresión génica a lo largo de la lactación en el ganado vacuno realizado por (Wickramasinghe *et al.*, 2012) sí se detectan genes relacionados con la apoptosis y muerte celular tanto

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

entre los genes más expresados como en los diferencialmente expresados en la última etapa de lactación bovina (día 250). Para entender estas diferencias entre el perfil de expresión génica al final de la lactación en el ganado vacuno y el ganado ovino lo primero que deberíamos tener en cuenta es la aptitud productiva y longitud de las lactaciones en una y otra especie. Por lo general, la aptitud principal del ganado vacuno lechero es la producción láctea y en esta especie los periodos de lactación se llevan al límite (lactaciones normalizadas a 305 días en la raza Holstein). Sin embargo, las dos razas ovinas lecheras analizadas en este trabajo se explotan generalmente con una doble aptitud, producción de leche, principalmente para queso, y producción de corderos para carne. En este sentido los actuales sistemas de manejo están orientados a que se produzca un parto al año en las explotaciones de ovejas de raza Assaf y tres partos en dos años en las explotaciones de ovejas de raza Churra. Esto hace que las lactaciones se normalicen generalmente a 150 días en la raza Assaf y 120 días en la raza Churra. Desde nuestro punto de vista, en base al perfil de expresión génica analizado, en el que la expresión de caseínas y proteínas del lactosuero se mantienen elevadas y no se detectan genes relacionados con fenómenos de degeneración tisular al final de la lactación, las lactaciones podrían ser fisiológicamente más largas en el ganado ovino, sin embargo desde un punto de vista productivo resulta económicamente más interesante tener lactaciones más cortas y sacar beneficios de la venta de corderos.

En el análisis de expresión diferencial realizado entre las dos razas, Churra y Assaf, se detectaron sobreexpresados, en la raza Churra, unos grupos de genes asociados con el transporte de iones y la actividad endopeptidasa. Los genes agrupados bajo estas funciones podrían relacionarse con la mayor aptitud de la leche de Churra para la elaboración de quesos. En el caso de los genes relacionados con el transporte de iones, estos genes pueden influir sobre características de la leche como el pH y la concentración de calcio y otros minerales. Estos factores tienen una influencia directa sobre el proceso de coagulación de la leche para la elaboración de queso (Park *et al.*, 2007, Bittante *et al.*, 2012). En cuanto a la actividad endopeptidasa, en general, las proteasas endógenas de la leche suelen tener efectos sobre las características fisicoquímicas de la leche fresca y sobre la calidad de los productos lácteos (Marino *et*

al., 2005). La proteólisis tiene una influencia directa en el sabor del queso ya que en este proceso se producen péptidos cortos y aminoácidos, algunos de los cuales tienen sabor. También influye de forma indirecta, ya que a lo largo de la proteólisis se liberan aminoácidos que son sustratos para una serie de reacciones catabólicas que generan muchos compuestos importantes para el sabor del queso maduro (McSweeney, 2004). El proceso de proteólisis es muy complejo y en él juegan un papel fundamental diferentes enzimas proteolíticas tanto exógenas (cuajo, microorganismos) como endógenas (presentes en la propia leche). Hay que tener en cuenta que las proteasas endógenas tienen un papel importante en el proceso de maduración del queso ya que son una de las fuentes de enzimas proteolíticas, entre las proteasas endógenas destacan por su actividad en la leche, la plasmina, las catepsinas y la elastasa (McSweeney, 2004). En una revisión inicial de los genes sobreexpresados en Churra, dentro del término actividad peptidasa nos llama la atención la expresión de un grupo de seis genes pertenecientes a la familia de las calicreínas (*KLK5*, *KLK6*, *KLK7*, *KLK10*, *KLK12*, *KLK13*). Estos genes codifican para un grupo de serín-proteasas con actividad tipo tripsina o quimotripsina que son responsables, entre otras, de la coordinación de diversas funciones fisiológicas incluyendo la presión arterial, la licuefacción del semen y la descamación de la piel (Yousef and Diamandis, 2001, Shaw and Diamandis, 2007). El estudio realizado por Shaw and Diamandis (2007) en seres humanos confirma la presencia de las calicreínas *KLK4*, *KLK5*, *KLK6*, *KLK7*, *KLK8*, *KLK9*, *KLK10*, *KLK11*, *KLK12*, *KLK15* en leche. Las funciones de esta familia de proteasas en la leche están aún por determinar y no hemos encontrado ningún trabajo que analice el posible papel de las calicreínas en la glándula mamaria. Sin embargo, sorprende la actividad diferencial de los genes que codifican para estas enzimas proteolíticas en las dos razas analizadas. El posible papel de la concentración de estas proteasas y la diferente aptitud en la maduración de los quesos entre ambas razas deberá ser objeto de análisis en futuros trabajos.

En el estudio preliminar realizado sobre los genes relacionados con el metabolismo lipídico de la glándula mamaria en lactación se seleccionaron un total de siete genes: dos genes implicados en la captación de ácidos grasos por la glándula mamaria (*VLDLR* y *LPL*), dos genes implicados en la formación del glóbulo graso (*XDH* y *BTN1A1*), dos

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

genes implicados en la síntesis *de novo* (*ACACA* y *FASN*) y un gen relacionado con la desaturación de los ácidos grasos (*SCD1*). Ninguno de los genes está diferencialmente expresado entre las dos razas. En general, la mayoría de los genes analizados incrementaron su expresión a lo largo de la lactación, al contrario de lo descrito en el ganado bovino, especie en la que la mayoría de los genes muestran su mayor expresión al inicio de la misma (Wickramasinghe *et al.*, 2012). Los tres genes con mayor expresión fueron los genes *FASN*, *SCD1* y *XDH*, que además se encuentran en el grupo de genes con más de 500 FPKM en el estudio global, por lo tanto en el grupo de genes altamente expresados. Los genes *FASN* y *XDH* están implicados en la producción de ácidos grasos *de novo* y la formación del glóbulo graso. El incremento de expresión de estos genes a lo largo de la lactación implicaría una mayor formación de lípidos, producidos en el retículo endoplásmico liso. En los análisis de expresión diferencial a lo largo de la lactación encontramos sobre-expresados en la raza Churra en los días 120 y 150, genes que codifican proteínas cuya actividad se detecta de forma preferencial en el retículo endoplásmico, lo que podría correlacionarse con la mayor formación de lípidos en etapas finales de la lactación. El gen *SCD1* codifica para la esteroil-CoA desaturasa 1, principal enzima implicada en la biosíntesis de ácidos grasos monoinsaturados en rumiantes (Ntambi and Miyazaki, 2004). En la glándula mamaria de rumiantes, este gen se encarga de la producción de cerca del 80% de la forma más común de ácido linoleico conjugado (CLA, isómero *cis*-9, *trans*-11, C18:2) (Corl *et al.*, 2001). Está demostrado que el consumo de ácidos linoleicos conjugados es beneficioso para la salud humana por sus propiedades anti-carcinogénicas, anti-lipogénicas e inmunomoduladoras (Pariza *et al.*, 1999). El *SCD1*, es el gen más expresado, de los estudiados en este apartado, al final de la lactación en ambas razas, siendo también el gen más expresado en Churra los días 50 y 120 de lactación, lo que implica que pueda estar relacionada con una mayor concentración de CLA en la leche de Churra y, por lo tanto, con una leche potencialmente más saludable. Estudios previos realizados sobre la concentración en leche de CLA (C18:2 *cis*-9, *trans*-11) en ambas razas confirman que la concentración de CLA en Churra es superior (0,89 g/ 100 g de ácidos grasos totales; De La Fuente *et al.*, 2009) a la de Assaf (0,64 g/ 100 g de ácidos grasos totales; Toral *et al.*, 2010).



Este trabajo constituye el primer estudio que analiza el perfil de expresión génica de la glándula mamaria de la oveja en lactación. Se trata de un estudio complejo, con dos razas y cuatro puntos de muestreo, donde los resultados obtenidos nos proporcionan una visión global del transcriptoma de la glándula mamaria de la oveja y nos dan información sobre posibles rutas metabólicas implicadas en la lactación y en caracteres productivos de interés. Sin embargo, la gran cantidad de datos proporcionados por la técnica de RNA-Seq hace que este estudio sólo constituya el inicio y que quede aún mucho trabajo por hacer, como el análisis detallado de los genes implicados en las rutas significativas detectadas, relacionadas *a priori* con caracteres productivos; la identificación de variantes tanto en genes relacionados con estas rutas como en regiones genómicas en las que previamente se han identificado QTLs; así como el análisis de expresión diferencial de isoformas y análisis funcional de transcritos no anotados en el genoma de referencia. De la misma forma, resultaría interesante ahondar en rutas específicas, que aunque no contengan genes diferencialmente expresados, sean interesantes desde el punto de vista productivo (genes implicados en el metabolismo de oligosacáridos, rutas metabólicas implicadas en la degradación de proteínas, etc.) al igual que se ha hecho en el análisis preliminar de los genes implicados en el metabolismo lipídico presentado en esta Tesis Doctoral.

### **5.3. Perspectivas Futuras**

La capacidad de secuenciar y descifrar el material genético de las distintas especies animales de una forma rápida y a un coste asequible se ha logrado gracias al desarrollo de las técnicas NGS. Sin embargo, el almacenamiento, procesado, análisis, interpretación e integración de resultados suponen un reto para los investigadores debido a la enorme cantidad de datos generados.

Las técnicas de secuenciación de alto rendimiento ofrecen oportunidades sin precedentes en la identificación de genes y mutaciones relacionadas con enfermedades. La principal ventaja en la utilización de herramientas NGS es que, al contrario que las plataformas de genotipado, las técnicas de secuenciación de segunda generación permiten la detección de forma directa de las mutaciones causales. En la actualidad la secuenciación de exomas es la técnica más utilizada en la especie

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

humana, sin embargo la disminución progresiva de los costes de secuenciación favorecerá que cada vez se recurra más a la secuenciación de genomas completos, técnica claramente más ventajosa dado el actual estado de anotación en especies como la oveja. Así mismo, el abaratamiento de los costes permitirá la secuenciación de un mayor número de muestras lo que ampliará la potencia de detección. No obstante, al contrario de lo que ocurre cuando se utilizan plataformas de genotipado donde los análisis de GWAS y de identidad por descendencia están altamente estandarizados, en la utilización de técnicas NGS, la capacidad para la generación de datos aún va muy por delante del desarrollo de protocolos estándar para interpretarlos correctamente. Los datos de secuenciación masiva paralela son más propensos a interpretaciones erróneas debido a la gran cantidad de mutaciones causales candidatas que podemos encontrar en cualquier genoma, muchas de las cuales podrían concordar por azar con datos que sugieran que la mutación estaría relacionada con la enfermedad estudiada, un problema que ha sido definido como el “potencial narrativo” de los genomas (Goldstein *et al.*, 2013). Por ello, en la revisión realizada por MacArthur *et al.* (2014) se describen una serie de directrices y se sugieren pautas para el desarrollo y mejora en la investigación de mutaciones causales utilizando técnicas NGS; entre ellas se propone la necesidad de desarrollo y evaluación comparativa de métodos estadísticos estandarizados que permitan la asignación de probabilidades objetivas a los nuevos genes identificados como responsables de enfermedades y mutaciones candidatas (MacArthur *et al.*, 2014). En el mismo estudio, también se sugiere la mejora de las bases de datos en las que se describen las mutaciones detectadas y se anima a los investigadores a que publiquen sus fenotipos, datos de secuenciación masiva y pruebas funcionales con el fin de que toda esta información pueda servir de ayuda en la implicación o exclusión en futuros estudios de nuevas variantes génicas en relación con una determinada enfermedad (MacArthur *et al.*, 2014).

En lo que se refiere a la utilización de datos de RNA-Seq en el estudio del perfil transcripcional de tejidos relacionados con caracteres de interés productivo en las distintas especies domésticas, esta técnica ha demostrado ser de gran utilidad ayudando a caracterizar biológicamente dichos transcriptomas e identificando rutas metabólicas, genes y asociaciones entre alelos marcadores y fenotipos de interés

(Wickramasinghe *et al.*, 2014). Sin embargo, la técnica de RNA-seq también tiene limitaciones que deben ser tenidas en cuenta. En el caso de los animales domésticos los resultados de los estudios de RNA-seq se encuentran en muchas ocasiones condicionados por la anotación actual de los genomas. El proyecto FAANG (<http://www.faang.org/>), es un proyecto reciente (Andersson *et al.*, 2015) que al igual que el ENCODE en seres humanos (ENCODE Project Consortium, 2004) y otras especies modelo (Gerstein *et al.*, 2010; modENCODE Consortium *et al.*, 2010; Shen *et al.*, 2012; Sivasubbu *et al.*, 2013; Yue *et al.*, 2014), nace en los animales domésticos con el objetivo de mapear los elementos funcionales del genoma. En la actualidad, los estudios para la identificación de elementos funcionales se realizan, generalmente, en muestras de tejido que a menudo contiene múltiples células heterogéneas, lo que complica los análisis. Recientemente las técnicas NGS han solventado estos problemas con el desarrollo de plataformas capaces de secuenciar el material genético de una única célula (Nawy, 2014). Así mismo, se han desarrollado técnicas de secuenciación de tercera generación que permiten la secuenciación de una sola molécula de DNA lo que elimina los sesgos producidos en la amplificación inicial (Schadt *et al.*, 2010). Como se deduce de todo lo expuesto anteriormente, el avance de la tecnología en la genómica ha sido espectacular y el reto ahora se encuentra en el desarrollo de herramientas que nos permitan interpretar la complejidad biológica que subyace a la variabilidad genética para poder predecir los fenotipos resultantes o lo que es lo mismo, predecir las consecuencias codificadas en la secuencia genómica. De la misma forma, la colaboración entre diferentes grupos de investigación e integración de datos de las diferentes ciencias “ómicas” en lo que se conoce como biología de sistemas resulta esencial en la interpretación de resultados y su aplicación en el conocimiento de la base genética de caracteres productivos complejos.





## 6. Conclusiones



1ª. El clonado posicional mediante un análisis de asociación en el genoma completo (GWAS) y un análisis de identidad por descendencia nos han permitido identificar la mutación causal de la *lisencefalia con hipoplasia cerebelar* en el ganado ovino de raza Churra. Así:

- Se ha detectado una delección de 31 pb en el exón 36 del gen *RELN* (*c.5410\_5440del*) en estado de homocigosis como la posible mutación responsable de la enfermedad.
- Los efectos de dicha mutación han sido confirmados a nivel funcional mediante la utilización de las técnicas *qRT-PCR* y *Western blot*.
- Se ha desarrollado un método de diagnóstico rápido para la identificación de portadores sanos en las explotaciones de raza Churra.

2ª. La utilización del chip de SNPs de 50K ovino junto con el uso de técnicas de secuenciación de segunda generación han demostrado ser eficaces en el mapeo e identificación de la base genética de la *epidermólisis bullosa juntural* en el ganado ovino de raza Churra. De esta manera:

- Se ha identificado como mutación causal de esta enfermedad una delección de 4 pb en el exón 33 del gen *ITGB4* (*c.4412\_4415del*) que se encontraba en homocigosis en los animales afectados por la enfermedad.
- Los análisis funcionales (*qRT-PCR* e inmunohistoquímica) revelan una disminución en los niveles de mRNA del gen *ITGB4* y la ausencia de integrina  $\beta 4$  en las muestras de piel de los corderos afectados.
- La combinación de la secuenciación masiva paralela de un BAC conteniendo un inserto con la región genómica ovina donde se localiza el gen *ITGB4* y los datos de RNA-Seq de la piel de un control sano ha permitido obtener la secuencia completa del gen *ITGB4* y su correcta anotación.
- Se ha implementado un método de diagnóstico rápido para la identificación de portadores sanos.

***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

3ª. Se ha llevado a cabo el primer estudio integrado de la caracterización de la expresión génica de la glándula mamaria ovina a lo largo de la lactación mediante RNA-Seq. Así hemos podido comprobar que:

- Los diez genes más expresados en cada punto de muestreo acumulan aproximadamente el 70% del total de FPKM, lo que da una idea del alto grado de especialización de la glándula mamaria. Los genes que codifican para las caseínas y las proteínas del lactosuero son los más expresados en todos los puntos analizados.
- Se han detectado un alto porcentaje de transcritos intergénicos, el 61,87 % de todos los transcritos ensamblados, lo que refleja la anotación incompleta de la versión actual del genoma ovino (Oar\_v3.1), así como la complejidad del mismo.
- En los análisis de expresión diferencial longitudinal, las mayores diferencias se detectaron entre el inicio y el final de la lactación. En las etapas finales de este proceso fisiológico se ha detectado una mayor expresión de genes relacionados con la remodelación de la glándula mamaria.
- En las comparaciones de la expresión génica diferencial entre las dos razas ovinas estudiadas, Churra y Assaf, se han identificado genes sobre-expresados en la raza Churra cuya función está relacionada con la actividad endopeptidasa y el transporte de iones. La mayor expresión de estos genes podría asociarse con una mayor aptitud de la leche de Churra para la elaboración de queso.





## **7. Conclusions**



1<sup>st</sup>. The positional cloning with the *OvineSNP50 BeadChip* through a genome-wide association study and an identity-by-descent analysis have enabled the identification of the causal mutation of the *lissencephaly with cerebellar hypoplasia* in Spanish Churra sheep. Thus:

- We have detected a homozygous deletion of 31 bp in the predicted exon 36 of the *RELN* gene as the causal mutation of the disease.
- The effects of the mutation have been confirmed at the functional level by using *qRT-PCR* and *Western blot* techniques.
- A diagnostic method for the rapid detection of healthy carriers has been developed in order to avoid at-risk matings in Churra flocks.

2<sup>nd</sup>. The combination of the use of the *OvineSNP50 BeadChip* and massive parallel sequencing has enabled the characterization of the molecular architecture of the *junctional epidermolysis bullosa* in Spanish Churra Sheep. In that way:

- We herein demonstrated that a 4-bp deletion within exon 33 of the *ITGB4* gene (*c.4412\_4415del*), which was found in homozygosity exclusively in affected lambs, is responsible for the junctional epidermolysis bullosa in Churra sheep.
- The functional analyses performed in relation to this deletion revealed that junctional epidermolysis bullosa skin samples showed decreased levels of mRNA and the absence of integrin  $\beta 4$  labeling in immunohistochemical assays.
- *De novo* BAC clone sequencing combined with the RNA-seq data obtained from the skin sample of a control individual allowed us to obtain the complete *ITGB4* gene sequence and its appropriate annotation.
- A diagnostic method has been implemented in Churra sheep flocks to identify healthy carriers of this mutation.

3<sup>rd</sup>. The RNA-seq approach on sheep milk somatic cells across lactation reported here comprises the first integrated overview on global expression of sheep milk transcriptome. Thus, we have confirmed that:

- The ten highest expressed genes at each time point accumulate approximately 70% of the total gene FPKM reads, which means that a small number of genes contribute to a large fraction of the total RNA extracted from milk somatic cells. As could be expected, due to the high specialization of the lactating mammary gland machinery, genes encoding caseins and whey proteins are part of the most expressed genes across lactation.
- A large proportion of intergenic transcripts (61.87% of the total assembled transcripts) has been detected. This fact underlines the incompleteness of the annotation of the current sheep transcriptome as well as its complexity.
- In the differential expression analysis across lactation, the highest number of differentially expressed genes was found between the extreme time points of lactation. Upregulated genes at final stages of lactation were enriched in terms related to mammary gland remodeling.
- In the comparison between the Churra and Assaf breeds, we found genes selectively upregulated in Churra grouped within endopeptidase and channel activity functions. These genes could be related to the higher cheese yield of this breed when compared with Assaf.



## 8. Resumen



La presente Tesis Doctoral surge dentro del marco de los cambios y la evolución que han sufrido las herramientas genómicas en los últimos diez años, desde la aparición de las plataformas de genotipado masivo derivadas de la información producida en las fases iniciales de los proyectos de secuenciación de los genomas de los animales domésticos, hasta el desarrollo de las técnicas de secuenciación de segunda generación (*Next Generation Sequencing* (NGS)) que surgen debido a la necesidad de tener herramientas de secuenciación de alto rendimiento en el proyecto del Genoma Humano. El objetivo inicial de la presente Tesis Doctoral, coincidiendo con la comercialización del chip de SNPs de 50K ovino, era la utilización del *OvineSNP50 BeadChip* para la caracterización de la base molecular de dos enfermedades de herencia mendeliana, detectadas en rebaños del núcleo de selección de la Asociación de Criadores de Raza Churra (ANCHE). Sin embargo, la rápida evolución, eficacia y asequibilidad de las técnicas de secuenciación de segunda generación en los últimos años nos ha permitido aplicar el uso de estas nuevas metodologías en el ámbito de la mejora genética en producción animal. Así, hemos utilizado las técnicas NGS, en concreto la secuenciación masiva paralela de RNA (RNA-Seq), tanto para la caracterización de la base molecular de una de las enfermedades de herencia mendeliana en estudio, como para el análisis del transcriptoma de la glándula mamaria en lactación. Uno de los principales objetivos de nuestro grupo de investigación es la mejora de la producción láctea en el ganado ovino, principalmente en la raza Churra, por ello, el conocimiento del perfil transcripcional de la glándula mamaria en lactación es de especial interés, ya que permite la identificación de genes que controlan caracteres complejos relacionados con la producción y la composición de la leche, y la detección de variantes génicas que pudieran ser utilizadas en la mejora de la producción lechera. En función de lo expuesto, el objetivo general de esta Tesis Doctoral ha sido utilizar las herramientas genómicas disponibles en el estudio de caracteres de importancia económica en el ganado ovino lechero. En concreto, utilizar el *OvineSNP50 BeadChip* en el mapeo de regiones portadoras de mutaciones causales de enfermedades de herencia mendeliana y utilizar las herramientas de secuenciación masiva paralela en la caracterización de la mutación causal de enfermedades monogénicas y del transcriptoma de la glándula mamaria de la oveja en lactación.

## **Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

En la primera etapa de la presente Tesis Doctoral, utilizamos el chip de SNPs de 50K ovino junto con una aproximación clásica de secuenciación Sanger para identificar la mutación causal de la lisencefalia con hipoplasia cerebelar (*Lisencephaly with cerebellar hypoplasia* (LCH)) en el ganado ovino de raza Churra. El chip ovino *OvineSNP50 BeadChip* se utilizó para genotipar siete corderos afectados de LCH y 33 controles sanos. Los análisis de asociación a nivel genómico y de identidad por descendencia realizados con los datos de genotipado nos permitieron acotar la región de estudio a un intervalo de 4,8 Mpb en el OAR4. La evaluación de la funcionalidad de los genes localizados en el intervalo de homocigosis nos permitió identificar al gen *RELN* como el mejor candidato posicional y funcional. El gen *RELN* juega un papel crítico en la migración neuronal y la formación de las capas cerebrales. La aplicación de la técnica de secuenciación *Sanger* en el mRNA de este gen en un control sano nos permitió obtener la secuencia de cDNA completa del gen *RELN* ovino. La secuenciación del cDNA de este gen en un cordero enfermo y su comparación con el control sano reveló una delección de 31 pb en el exón 36 del gen *RELN* del cordero afectado (*c.5410\_5440del*). Según los estudios de predicción funcional, la mutación detectada (*c.5410\_5440del*) provocaría un cambio en el marco de lectura de la secuencia codificante del gen *RELN* ocasionando un codón de terminación prematuro. Las pruebas funcionales realizadas (*qRT-PCR* y *Western blot*) en animales enfermos revelaron que la mutación provocaba una disminución en la cantidad de mRNA y la ausencia de la proteína en el tejido cerebral.

La segunda enfermedad caracterizada a nivel molecular en la presente Tesis Doctoral fue la epidermólisis bullosa juntural (*junctional epidermolysis bullosa* (JEB)) ovina. La JEB es una genodermatosis autosómica recesiva que cursa con la fragilidad mecánica de la piel y membranas mucosas y la formación de ampollas en la unión dermo-epidérmica, en concreto, a nivel de la lámina lúcida, manifestándose al nacimiento. En el núcleo de selección de ANCHE se detectaron dos rebaños en los que nacían corderos que presentaban erosiones y úlceras en la piel y membranas mucosas y a los que se diagnosticó JEB. El chip ovino de media densidad (*OvineSNP50 BeadChip*) se utilizó para genotipar un total de 96 animales, 48 controles sanos no relacionados y 48 animales de los dos rebaños afectados, de los cuales 20 eran corderos enfermos. Al



igual que en la LCH, los análisis de asociación a nivel genómico y de identidad por descendencia realizados con los datos de genotipado nos permitieron acotar la localización de la mutación causal a un intervalo de homocigosis de 868 Kpb en el OAR11. Dentro del intervalo de homocigosis identificado en los corderos afectados con JEB, se seleccionó el gen *ITGB4*, que codifica para la integrina  $\beta 4$  como mejor candidato posicional y funcional. En el epitelio estratificado, la integrina  $\beta 4$  es un componente estructural de los hemidesmosomas, complejos proteicos que intervienen en la adhesión de las células epiteliales a la membrana basal subyacente. La combinación de la secuencia genómica del *ITGB4*, obtenida mediante la secuenciación de tres clones BAC, y de la información generada al aplicar la técnica de RNA-Seq en la piel de un cordero sano permitió obtener la secuencia anotada del gen *ITGB4* ovino. El análisis de variantes realizado a partir de las muestras de RNA-seq del control sano utilizado para la anotación y de un cordero con JEB nos permitió identificar una delección de 4 pb en el exón 33 del gen *ITGB4* (*c.4412\_4415del*) como responsable de la JEB en el ganado ovino de raza Churra. En este estudio se demuestra que la utilización de plataformas de genotipado de media densidad en combinación con la técnica de RNA-seq resulta ser una estrategia efectiva en la detección de mutaciones causales en enfermedades de herencia mendeliana, constituyendo una alternativa eficaz a la secuenciación de exomas en aquellas especies con un genoma de referencia en desarrollo. Además la técnica de RNA-seq permite conocer el grado de expresión génica de los transcritos expresados en el tejido de interés. Las consecuencias de la mutación identificada se evaluaron mediante pruebas funcionales (*qRT-PCR* e inmunohistoquímica), comprobándose con las mismas una disminución en los niveles de mRNA del gen *ITGB4* y la ausencia de integrina  $\beta 4$  en las muestras de piel de los corderos afectados.

El fin último de la detección de mutaciones causales responsables de enfermedades monogénicas es el control de la aparición de enfermedades genéticas emergentes. La aplicación práctica de la detección de las mutaciones causales en las dos enfermedades caracterizadas en la presente Tesis Doctoral (LCH y JEB) se ve reflejada en las herramientas de diagnóstico desarrolladas y que han sido objeto de patente. Estas herramientas permiten el diagnóstico en animales con síntomas compatibles y

***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

facilitan la aplicación de la selección asistida por genes en el núcleo de selección de ANCHE. La detección de mutantes heterocigotos permite la eliminación de estos animales del programa de selección, ayudando a los criadores a evitar apareamientos de riesgo en sus rebaños. Así mismo, los métodos de diagnóstico desarrollados podrían utilizarse en la detección de portadores en el caso de cruzamientos programados para obtener animales que sirvan como modelo en el estudio de ambas enfermedades.

La disponibilidad de las metodologías de secuenciación de segunda generación y la democratización de los precios de las mismas han permitido que grupos que trabajan en mejora genética animal centren sus objetivos en la caracterización de los transcriptomas de tejidos relacionados con caracteres productivos. Uno de los objetivos de la presente Tesis Doctoral ha sido el análisis global del transcriptoma mamario en el ganado ovino mediante la técnica de RNA-Seq. Este estudio constituye el primer análisis integral del transcriptoma de la glándula mamaria en lactación. Además, en él se han comparado dos razas de ovejas lecheras con distintas características productivas, Churra y Assaf. El planteamiento del trabajo es doble, por un lado ver las diferencias de expresión a lo largo de la lactación en el ganado ovino, y en segundo lugar analizar las diferencias de expresión entre dos razas con distintas características productivas y lactación de distinta duración. Para este estudio se utilizaron muestras de leche procedentes de 8 ovejas libres de mamitis, 4 ovejas de raza Assaf y 4 ovejas de raza Churra. Los puntos de muestreo fueron los días 10, 50, 120 y 150 de lactación. La secuenciación masiva paralela del transcriptoma se realizó en muestras de RNA total procedentes de las células somáticas de la leche. En el análisis global del transcriptoma de la oveja durante la lactación se obtuvieron un total de 1.116 millones de pares de lecturas a partir de la secuenciación del transcriptoma de 30 muestras de RNA. Tras el correspondiente análisis bioinformático a partir de esos datos, se detectaron un total de 107.877 transcritos, de los cuales el 23,95% concordaba exactamente con los transcritos anotados, el 11,18% eran nuevas isoformas potenciales y el 61,87% eran transcritos intergénicos. El alto porcentaje de transcritos intergénicos nos da una idea de la anotación incompleta de la actual versión del genoma ovino (Oar\_v3.1). Nuestros resultados muestran que el perfil de genes altamente expresados en la glándula mamaria es muy similar a lo largo de la

lactación y en las dos razas estudiadas. En la glándula mamaria se expresa un 67% de los genes anotados en el genoma ovino (Oar\_v3.1). Los diez genes más expresados en cada punto de muestreo acumulan aproximadamente el 70% del total de FPKM, lo que significa que un pequeño número de genes acumula un porcentaje muy alto de las lecturas del RNA total extraído de las células somáticas de la leche. Como cabe esperar, dado el alto grado de especialización de la glándula mamaria, los genes que codifican para las caseínas y proteínas del lactosuero se encuentran entre los genes más expresados en las células somáticas de la leche a lo largo de la lactación. En el estudio de expresión diferencial realizado entre los distintos puntos de muestreo a lo largo de la lactación, el mayor número de genes diferencialmente expresados en las dos razas se detectó entre el inicio y el final de lactación (días 10 y 150). En total se detectaron 89 genes anotados expresados diferencialmente en Churra y 38 en Assaf. Los análisis de enriquecimiento funcional para los genes sobreexpresados en el día 10 de la lactación en la raza Churra se relacionaron con procesos de respuesta a estímulos, proliferación y diferenciación celular y regulación positiva de fosforilación de proteínas; todas ellas, actividades de síntesis relacionadas con los altos niveles de producción de las primeras etapas de la lactación. En el caso de la raza Assaf, no se encontró enriquecimiento en términos de ontología de genes en los genes sobreexpresados en el día 10. En el día 150 de la lactación, para las dos razas, se encuentran sobreexpresados genes implicados en procesos de desarrollo, morfogénesis de órganos, movimiento y migración celular, lo que sugeriría procesos de remodelación de la glándula mamaria que podrían estar relacionados con el final de la lactación. En el análisis de expresión diferencial realizado entre las dos razas, Churra y Assaf, se detectaron un total de 256 genes anotados expresados de forma diferencial. En la raza Churra se detectaron varios grupos de genes enriquecidos en funciones relacionadas con transporte de iones y actividad endopeptidasa. Estos genes podrían afectar a las propiedades de coagulación de la leche y maduración de los quesos y explicar, en parte, el mejor rendimiento de la leche de Churra para la producción de quesos.

Centrándonos en rutas metabólicas específicas, el estudio preliminar realizado con una selección de siete genes implicados en el metabolismo de los ácidos grasos en la

***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

glándula mamaria (*LPL, VLDL, ACACA, FASN, SCD, BTN1A1* y *XDH*), observamos que, en general y al contrario de lo descrito en ganado vacuno, en la oveja los genes analizados sufren un incremento progresivo en su expresión a lo largo de la lactación, lo que se correlacionaría con la mayor concentración de lípidos en la leche ovina comparada con la leche de vaca.

Este trabajo nos da una visión global del transcriptoma mamario de la oveja en lactación, proporcionando información fundamental para el desarrollo de futuros análisis detallados de los genes implicados en las rutas significativas detectadas, relacionadas a priori con caracteres productivos, así como el análisis funcional de transcritos no anotados en el genoma de referencia o la identificación de variantes relacionadas con caracteres productivos de interés.



## 9. Summary



This PhD Thesis was set out concurrently with the impressive progress that has been achieved in the field of genomics during the last decade, from whole genome SNPs genotyping platforms derived from the initial phases of the genome sequencing projects of livestock species, to the development of Next Generation Sequencing (NGS) technologies developed by the need of using high-throughput sequencing tools in the Human Genome Project. The initial objective of this PhD Thesis was the use of the *OvineSNP50 BeadChip*, developed by the International Sheep Genomics Consortium and commercialized by *Illumina* in 2009, to characterize the molecular architecture of two Mendelian disorders that had appeared in several commercial flocks of Spanish Churra sheep. However, with the development of NGS, the growing power and the reducing costs of these technologies in the last years sparked an enormous range of applications and enabled their use in animal breeding. Thus, in this work we have taken the advantage of using massive parallel sequencing of RNA not only to determine the basis of monogenic diseases, the initial objective of this PhD Thesis, but also to characterize the sheep mammary gland transcriptome. One of the principal aims of our research group is the improvement of traits related to milk production in Spanish Churra Sheep. On that way, the knowledge of the expression profile of the milk transcriptome is interesting since it allows the identification of genes and variants related to complex milk production traits that could be used in animal breeding. Hence, the final objective of this PhD Thesis was to use the available genomic tools for the study of the molecular architecture of economically important traits in sheep, specifically the use of the *OvineSNP50 BeadChip* for fine-scale mapping of inherited Mendelian diseases and the use of massive parallel RNA sequencing to characterize one of the monogenic diseases under study and the sheep milk transcriptome.

In the first stage of this PhD Thesis, we used of the *OvineSNP50 BeadChip* together with a Sanger sequencing classical approach to identify the causal mutation of the lissencephaly with cerebellar hypoplasia (LCH) in Spanish Churra sheep. The genotyping of seven affected animals and 33 controls with the *OvineSNP50 BeadChip* allowed the localization of the causative mutation for ovine LCH to a 4.8-Mb interval on sheep chromosome 4 using genome-wide association and homozygosity mapping analysis. The *RELN* gene, located within this interval, was considered a strong

positional and functional candidate because it plays critical roles in neuronal migration and layer formation. By performing a sequencing analysis of the cDNA of this gene in a control lamb, we obtained the complete coding sequence of the ovine *RELN* gene. The cDNA sequence of this gene from an LCH-affected lamb revealed a deletion of 31 bp (*c.5410\_5440del*) in the predicted exon 36 of *RELN*, resulting in a premature termination codon. A functional analysis of this mutation revealed decreased levels of *RELN* mRNA and a lack of reelin protein in the brain cortex and blood of affected lambs. This mutation showed a complete concordance with the Mendelian recessive pattern of inheritance observed for the disease.

The second disease analyzed in this PhD Thesis was the junctional epidermolysis bullosa (JEB), which is a recessive genodermatosis that manifests with neonatal mechanical fragility of the skin, blistering confined to the lamina lucida of the basement membrane and severe alteration of the hemidesmosomal junctions. In this study, we demonstrated the use of a genome-wide association mapping together with RNA-seq in a reduced number of samples, as an efficient approach to detect the causal mutation for a Mendelian disease, evidencing that RNA-seq offers cost-effective alternative to identify variants in the species in which high resolution exome-sequencing is not straightforward. In Spanish Churra sheep, JEB was detected in two commercial flocks. The *OvineSNP50 BeadChip* was used to genotype 20 affected animals and 76 controls. The JEB locus was mapped to *Ovis aries* chromosome 11 by GWAS and subsequently fine-mapped to an 868-Kb homozygous segment using the identical-by-descent method. Within the 868-Kb homozygous segment identified in JEB-affected animals, the *ITGB4* gene was identified as the best positional and functional candidate gene. The *ITGB4* gene codes for the integrin  $\beta 4$  protein. In stratified epithelia, integrin  $\beta 4$  is located in specialized epidermal cell-basement membrane adhesion structures, called hemidesmosomes. *De novo* BAC clone sequencing combined with the RNA-seq data obtained from the control skin sample allowed us to obtain the complete *ITGB4* gene sequence and its annotation. The RNA-seq variant analysis of control and JEB-affected skin RNA samples allowed us to find the causal mutation, a 4-bp deletion within exon 33 of the *ITGB4* gene (*c.4412\_4415del*). The *c.4412\_4415del* mutation causes a frameshift resulting in a



premature stop codon at position 1,472 of the amino acid sequence of the integrin  $\beta$ 4 protein. A functional analysis of this deletion revealed decreased levels of mRNA in JEB skin samples and the absence of integrin  $\beta$ 4 labeling in immunohistochemically assays. Genotyping of *c.4412\_4415del* showed complete concordance with the recessive mode of inheritance of the disease phenotype.

The ultimate aim of the identification of the causal mutations in monogenic diseases is the control of emerging recessive defects. The practical implementation of the molecular characterization of these two Mendelian diseases in Churra sheep (LCH and JEB) is reflected in the two diagnostic methods developed by our group that have been patented. These diagnostic methods enabled the application of gene-assisted selection to detect heterozygous mutants, which will help breeders to avoid at-risk matings in their flocks. Moreover, the identification of these naturally occurring mutations provide an opportunity to use Churra sheep as a genetically characterized large animal model for the study of both diseases LCH and JEB.

The availability of NGS technologies and the reductionist trend of sequencing costs have allowed that many animal breeding research groups have focused their objectives on the complete characterization of tissue transcriptomes. The undertaken RNA-seq approach on sheep milk somatic cells across lactation comprises the first integrated overview on global expression of the milk sheep transcriptome. It has also enabled us to explore the mammary gland transcriptome of two dairy sheep breeds with differences in production characteristics, Churra and Assaf. A total of eight healthy sheep were selected to be included in the experiment, four Assaf and four Churra ewes. Transcriptome sequencing (RNA-seq) was performed on total RNA extracted from milk somatic cells from ewes on days 10, 50, 120 and 150 after lambing. A total of 1,116 million paired-end reads were obtained from the transcriptome sequencing of 30 milk samples. A total of 107,877 transcripts were detected, 23.95% of these transcripts matched exactly with annotated transcripts in the reference ovine genome (Oar\_v3.1), whereas 11.18% of them were potentially novel isoforms and 61.87% were annotated as intergenic transcripts. The large number of intergenic transcripts detected in our analysis underlines the incompleteness of the annotation of the current sheep transcriptome (Oar\_v3.1) as well as its complexity.

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

Nearly 67% of the annotated genes in the reference genome (Oar\_v3.1) were expressed in the ovine milk somatic cells. The highest expressed genes across lactation are very similar between the different time points analyzed and in the two breeds. For both breeds, Churra and Assaf, the ten highest expressed genes at each time point accumulate at approximately 70% of the total gene FPKM reads, which means that a small number of genes contribute to a large fraction of the total RNA extracted from milk somatic cells. As could be expected, due to the high specialization of the lactating mammary gland machinery, genes encoding caseins and whey proteins are included within the most expressed genes across lactation. For the differentially expressed genes across lactation time points, the highest number of differences, in both breeds were found between extremes of lactation (day 10 vs. day 150), with Churra autochthonous breed showing the largest number of differences (89 annotated differentially expressed genes in Churra and 38 in Assaf). As expected for the initial stage of lactation, the functional enrichment analysis performed with the upregulated genes at day 10 in Churra revealed terms involving biological regulation and metabolic protein synthesis processes which are clearly related to the extreme metabolic stress associated with high levels of milk production. For Assaf breed, no significant enriched gene ontology terms were found for upregulated genes at day 10. At the last stage of lactation studied (day 150), similar enriched gene ontology terms were found for both breeds. These terms were related to development, morphogenesis, cellular movement and cell migration which suggest processes of mammary gland remodeling that could be related with the decline of lactation. When the milk transcriptome between both breeds, Churra and Assaf, was compared, 256 annotated genes were found as differentially expressed. In this comparison, we found genes selectively upregulated in Churra grouped within endopeptidase and channel activity functions. These genes could influence on renneting properties and flavor during cheese ripening and explain, at least partially, the higher cheese yield of this breed when compared to Assaf. Focusing on specific metabolic pathways, the initial study of the transcriptional profile of genes related to the mammary gland fat metabolism (*LPL*, *VLDL*, *ACACA*, *FASN*, *SCD*, *BTN1A1* and *XDH*), has shown that, in contrast with the evidences reported in dairy cattle, in sheep mostly all of the genes analyzed tend to increase their expression

across lactation. This result could be correlated with the increased concentration of lipids in milk towards late lactation.

The sequencing data and analysis applied on the milk sheep transcriptome that are reported here provide essential information for future studies on genes implicated in significant specific pathways involved in lactation, as well as the functional annotation of novel genes identified and the detection of genetic variants related to milk production traits.





## 10. Bibliografía



**Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A. et al.** (2010) A method and server for predicting damaging missense mutations. *Nature methods* 7, 248-249.

**Ahmad W., Faiyaz ul Haque M., Brancolini V., Tsou H.C., ul Haque S. et al.** (1998) Alopecia universalis associated with a mutation in the human hairless gene. *Science* 279, 720-724.

**Alain K., Karrow N.A., Thibault C., St-Pierre J., Lessard M. & Bissonnette N.** (2009) Osteopontin: an early innate immune marker of Escherichia coli mastitis harbors genetic polymorphisms with possible links with resistance to mastitis. *BMC genomics* 10, 444-2164-10-444.

**Alhaidari Z., Olivry T., Spadafora A., Thomas R.C., Perrin C. et al.** (2005) Junctional epidermolysis bullosa in two domestic shorthair kittens. *Veterinary dermatology* 16, 69-73.

**Al-Shahrour F., Minguez P., Tarraga J., Medina I., Alloza E. et al.** (2007) FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic acids research* 35, W91-6.

**Amaral A.J., Megens H.J., Kerstens H.H., Heuven H.C., Dibbits B. et al.** (2009) Application of massive parallel sequencing to whole genome SNP discovery in the porcine genome. *BMC genomics* 10, 374-2164-10-374.

**Anders S., Pyl P.T. & Huber W.** (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169.

**Anderson R.** (2014) Development of a High Density (600K) Illumina Ovine SNP Chip and Its Use to Fine Map the Yellow Fat Locus. In: *Plant and Animal Genome XXII Conference*, P611. San Diego, CA, USA.

**Anderson C.L., Chaudhury C., Kim J., Bronson C.L., Wani M.A. & Mohanty S. (2006)** Perspective-- FcRn transports albumin: relevance to immunology and medicine. *Trends in immunology* 27, 343-348.

**Andersson L., Archibald A.L., Bottema C.D., Brauning R., Burgess S.C. et al. (2015)** Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome biology* 16, 57-015-0622-4.

**Asmann Y.W., Klee E.W., Thompson E.A., Perez E.A., Middha S. et al. (2009)** 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC genomics* 10, 531-2164-10-531.

**Bamshad M.J., Ng S.B., Bigham A.W., Tabor H.K., Emond M.J. et al. (2011)** Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews.Genetics* 12, 745-755.

**Barkovich A.J., Kuzniecky R.I., Jackson G.D., Guerrini R. & Dobyns W.B. (2005)** A developmental and genetic classification for malformations of cortical development. *Neurology* 65, 1873-1887.

**Becker D., Tetens J., Brunner A., Burstel D., Ganter M. et al. (2010)** Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. *PloS one* 5, e8689.

**Benavides J., Suarez-Vega A., Gutierrez J., de la Hera S., Fuertes M. et al. (2015)** Generalized severe junctional epidermolysis bullosa with congenital absence of skin in churra lambs. *Veterinary dermatology*, 10.1111/vde.12226.

**Bionaz M., Periasamy K., Rodriguez-Zas S.L., Everts R.E., Lewin H.A. et al. (2012)** Old and new stories: revelations from functional analysis of the bovine mammary transcriptome during the lactation cycle. *PloS one* 7, e33268.



**Birnbaum R.Y., Landau D., Elbedour K., Ofir R., Birk O.S. & Carmi R.** (2008) Deletion of the first pair of fibronectin type III repeats of the integrin beta-4 gene is associated with epidermolysis bullosa, pyloric atresia and aplasia cutis congenita in the original Carmi syndrome patients. *American journal of medical genetics* 146A, 1063-1066.

**Bittante G., Penasa M. & Cecchinato A.** (2012) Invited review: Genetics and modeling of milk coagulation properties. *Journal of dairy science* 95, 6843-6870.

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

**Boutinaud M. & Jammes H.** (2002) Potential uses of milk epithelial cells: a review. *Reproduction, nutrition, development* 42, 133-147.

**Bovine Genome Sequencing and Analysis Consortium, Elsik C.G., Tellam R.L., Worley K.C., Gibbs R.A. et al.** (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324, 522-528.

**Boycott K.M., Vanstone M.R., Bulman D.E. & MacKenzie A.E.** (2013) Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature reviews Genetics* 14, 681-691.

**Broom M.F., Zhou C., Broom J.E., Barwell K.J., Jolly R.D. & Hill D.F.** (1998) Ovine neuronal ceroid lipofuscinosis: a large animal model syntenic with the human neuronal ceroid lipofuscinosis variant CLN6. *Journal of medical genetics* 35, 717-721.

**Cachon-Gonzalez M.B., San-Jose I., Cano A., Vega J.A., Garcia N. et al.** (1999) The hairless gene of the mouse: relationship of phenotypic effects with expression profile and genotype. *Developmental dynamics* 216, 113-126.

**Canovas A., Rincon G., Bevilacqua C., Islas-Trejo A., Brenaut P. et al.** (2014) Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. *Scientific reports* 4, 5297.

**Canovas A., Rincon G., Islas-Trejo A., Jimenez-Flores R., Laubscher A. & Medrano J.F.** (2013) RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. *Journal of dairy science* 96, 2637-2648.

**Canovas A., Rincon G., Islas-Trejo A., Wickramasinghe S. & Medrano J.F.** (2010) SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mammalian genome* 21, 592-598.

**Capt A., Spirito F., Guaguere E., Spadafora A., Ortonne J.P. & Meneguzzi G.** (2005) Inherited junctional epidermolysis bullosa in the German Pointer: establishment of a large animal model. *The Journal of investigative dermatology* 124, 530-535.

**Carmi R., Sofer S., Karplus M., Ben-Yakar Y., Mahler D. et al.** (1982) Aplasia cutis congenita in two sibs discordant for pyloric atresia. *American Journal of Medical Genetics* 11, 319-328.

**Chaisson M.J., Huddleston J., Dennis M.Y., Sudmant P.H., Malig M. et al.** (2015) Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 517, 608-611.

**Charlier C., Coppieters W., Rollin F., Desmecht D., Agerholm J.S. et al.** (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature genetics* 40, 449-454.

**Chavanas S., Gache Y., Vailly J., Kanitakis J., Pulkkinen L. et al.** (1999) Splicing modulation of integrin beta4 pre-mRNA carrying a branch point mutation underlies epidermolysis bullosa with pyloric atresia undergoing spontaneous amelioration with ageing. *Human molecular genetics* 8, 2097-2105.

**Chin C.S., Alexander D.H., Marks P., Klammer A.A., Drake J. et al.** (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods* 10, 563-569.

**Cirulli E.T. & Goldstein D.B.** (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews.Genetics* 11, 415-425.

**Clarke J., Wu H.C., Jayasinghe L., Patel A., Reid S. & Bayley H.** (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology* 4, 265-270.

**Cockett N.E., Shay T.L., Beever J.E., Nielsen D., Albrechtsen J. et al.** (1999) Localization of the locus causing Spider Lamb Syndrome to the distal end of ovine Chromosome 6. *Mammalian genome* 10, 35-38.

**Conesa A., Gotz S., Garcia-Gomez J.M., Terol J., Talon M. & Robles M.** (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.

**Corl B.A., Baumgard L.H., Dwyer D.A., Griinari J.M., Phillips B.S. & Bauman D.E.** (2001) The role of Delta(9)-desaturase in the production of cis-9, trans-11 CLA. *The Journal of nutritional biochemistry* 12, 622-630.

**Cox L.A., Glenn J.P., Spradling K.D., Nijland M.J., Garcia R. et al.** (2012) A genome resource to address mechanisms of developmental programming: determination of the fetal sheep heart transcriptome. *The Journal of physiology* 590, 2873-2884.

**Cui X., Hou Y., Yang S., Xie Y., Zhang S. et al.** (2014) Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing. *BMC genomics* 15, 226-2164-15-226.

**Daetwyler H.D., Capitan A., Pausch H., Stothard P., van Binsbergen R. et al.** (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature genetics* 46, 858-865.

**Dalloul R.A., Long J.A., Zimin A.V., Aslam L., Beal K. et al.** (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS biology* 8, 10.1371/journal.pbio.1000475.

**Dalrymple B.P., Kirkness E.F., Nefedov M., McWilliam S., Ratnakumar A. et al.** (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome biology* 8, R152.

**D'Arcangelo G. & Curran T.** (1998) Reeler: new tales on an old mutant mouse. *BioEssays* 20, 235-244.

**De Jesus Rodriguez B., Chevaleyre C., Henry G., Molle D., Virlogeux-Payant I. et al.** (2009) Identification in milk of a serum amyloid A peptide chemoattractant for B lymphoblasts. *BMC immunology* 10, 4-2172-10-4.

**De La Fuente L., Barbosa E., Carriedo J., Gonzalo C., Arenas R. et al.** (2009) Factors influencing variation of fatty acid content in ovine milk. *Journal of dairy science* 92, 3791-3799.

**De la Fuente L., Gabina D., Carolino N. & Ugarte E.** (2006) The Awassi and Assaf breeds in Spain and Portugal. In: *Proc. 57th Annual Meeting of the EAAP*, Paper S14.2. Antalya, Turkey.

**Desai A.N. & Jere A.** (2012) Next-generation sequencing: ready for the clinics? *Clinical genetics* 81, 503-510.

**Diaz-Sanchez S., Hanning I., Pendleton S. & D'Souza D.** (2013) Next-generation sequencing: the future of molecular genetics in poultry production and food safety. *Poultry science* 92, 562-572.

**Dillies M.A., Rau A., Aubert J., Hennequet-Antier C., Jeanmougin M. et al.** (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics* 14, 671-683.

**Djari A., Esquerre D., Weiss B., Martins F., Meersseman C. et al.** (2013) Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC genomics* 14, 307-2164-14-307.

**Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. et al.** (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.

**Dong Y., Xie M., Jiang Y., Xiao N., Du X. et al.** (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature biotechnology* 31, 135-141.

**Eggen A.** (2012) The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers* 2, 10-15.

**Eid J., Fehr A., Gray J., Luong K., Lyle J. et al.** (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133-138.

**ENCODE Project Consortium** (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306, 636-640.

**Fine J.D., Eady R.A., Bauer E.A., Bauer J.W., Bruckner-Tuderman L. et al.** (2008) The classification of inherited epidermolysis bullosa (EB): Report of the Third International Consensus Meeting on Diagnosis and Classification of EB. *Journal of the American Academy of Dermatology* 58, 931-950.

**Finocchiaro R., Portolano B., Damiani G., Caroli A., Budelli E. et al.** (2003) The hairless (hr) gene is involved in the congenital hypotrichosis of Valle del Belice sheep. *Genetics, selection, evolution* 35 Suppl 1, S147-56.

**French L.E., Soriano J.V., Montesano R. & Pepper M.S.** (1996) Modulation of clusterin gene expression in the rat mammary gland during pregnancy, lactation, and involution. *Biology of reproduction* 55, 1213-1220.

**Frotscher M.** (2010) Role for Reelin in stabilizing cortical architecture. *Trends in neurosciences* 33, 407-414.

**Frugier T., Mitchell N.L., Tammen I., Houweling P.J., Arthur D.G. et al.** (2008) A new large animal model of CLN5 neuronal ceroid lipofuscinosis in Borderdale sheep is

caused by a nucleotide substitution at a consensus splice site (c.571+1G>A) leading to excision of exon 3. *Neurobiology of disease* 29, 306-315.

**Garcia-Gamez E., Gutierrez-Gil B., Sahana G., Sanchez J.P., Bayon Y. & Arranz J.J.** (2012a) GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PloS one* 7, e47782.

**Garcia-Gamez E., Sahana G., Gutierrez-Gil B. & Arranz J.J.** (2012b) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC genetics* 13, 43-2156-13-43.

**Garcia-Gamez E., Reverter A., Whan V., McWilliam S.M., Arranz J.J. et al.** (2011) Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in merino sheep. *PloS one* 6, e21158.

**Gerstein M.B., Lu Z.J., Van Nostrand E.L., Cheng C., Arshinoff B.I. et al.** (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* 330, 1775-1787.

**Goldstein D.B., Allen A., Keebler J., Margulies E.H., Petrou S. et al.** (2013) Sequencing studies in human genetics: design and interpretation. *Nature reviews.Genetics* 14, 460-470.

**Gonzaga-Jauregui C., Lupski J.R. & Gibbs R.A.** (2012) Human genome sequencing in health and disease. *Annual Review of Medicine* 63, 35-61.

**Graves K.T., Henney P.J. & Ennis R.B.** (2009) Partial deletion of the LAMA3 gene is responsible for hereditary junctional epidermolysis bullosa in the American Saddlebred Horse. *Animal Genetics* 40, 35-41.

**Greber D., Doherr M., Drogemuller C. & Steiner A.** (2013) Occurrence of congenital disorders in Swiss sheep. *Acta Veterinaria Scandinavica* 55, 27-0147-55-27.

**Haenlein G.F.** (2001) Past, present, and future perspectives of small ruminant dairy research. *Journal of dairy science* 84, 2097-2115.

**Hansen K.D., Brenner S.E. & Dudoit S.** (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research* 38, e131.

**Hatem A., Bozdag D., Toland A.E. & Catalyurek U.V.** (2013) Benchmarking short sequence mapping tools. *BMC bioinformatics* 14, 184-2105-14-184.

**Hirschhorn J.N. & Daly M.J.** (2005) Genome-wide association studies for common diseases and complex traits. *Nature reviews.Genetics* 6, 95-108.

**Hodges E., Xuan Z., Balija V., Kramer M., Molla M.N. et al.** (2007) Genome-wide in situ exon capture for selective resequencing. *Nature genetics* 39, 1522-1527.

**Holst B.D., Hurley W.L. & Nelson D.R.** (1987) Involution of the bovine mammary gland: histological and ultrastructural changes. *Journal of dairy science* 70, 935-944.

**Hong S.E., Shugart Y.Y., Huang D.T., Shahwan S.A., Grant P.E. et al.** (2000) Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations. *Nature genetics* 26, 93-96.

**International Chicken Genome Sequencing Consortium** (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695-716.

**International Human Genome Sequencing Consortium** (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431, 931-945.

**International Sheep Genomics Consortium, Archibald A.L., Cockett N.E., Dalrymple B.P., Faraut T. et al.** (2010) The sheep genome reference sequence: a work in progress. *Animal Genetics* 41, 449-453.

**Jiang Y., Xie M., Chen W., Talbot R., Maddox J.F. et al.** (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168-1173.

**Kann G., Delobelle-Deroide A., Belair L., Gertler A. & Djiane J.** (1999) Demonstration of in vivo mammogenic and lactogenic effects of recombinant ovine placental lactogen and mammogenic effect of recombinant ovine GH in ewes during artificial induction of lactation. *The Journal of endocrinology* 160, 365-377.

**Karageorgos L., Lancaster M.J., Nimmo J.S. & Hopwood J.J.** (2011) Gaucher disease in sheep. *Journal of inherited metabolic disease* 34, 209-215.

**Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R. et al.** (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS biology* 10, e1001258.

**Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F. et al.** (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS one* 4, e4668.

**Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R. & Salzberg S.L.** (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* 14, R36-2013-14-4-r36.

**Koren S., Schatz M.C., Walenz B.P., Martin J., Howard J.T. et al.** (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature biotechnology* 30, 693-700.

**Ku C.S., Cooper D.N., Polychronakos C., Naidoo N., Wu M. & Soong R.** (2012) Exome sequencing: dual role as a discovery and diagnostic tool. *Annals of Neurology* 71, 5-14.

**Kumar P., Henikoff S. & Ng P.C.** (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 4, 1073-1081.

**Kuska B.** (1998) Beer, Bethesda, and biology: how "genomics" came into being. *Journal of the National Cancer Institute* 90, 93.



**Kvam V.M., Liu P. & Si Y.** (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American Journal of Botany* 99, 248-256.

**Kwon J.M. & Goate A.M.** (2000) The candidate gene approach. *Alcohol Research & Health* 24, 164-168.

**Laird P.W.** (2010) Principles and challenges of genomewide DNA methylation analysis. *Nature reviews.Genetics* 11, 191-203.

**Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C. et al.** (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

**Larkin D.M., Daetwyler H.D., Hernandez A.G., Wright C.L., Hetrick L.A. et al.** (2012) Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proceedings of the National Academy of Sciences of the United States of America* 109, 7693-7698.

**Lemkin P.F., Thornwall G.C., Walton K.D. & Hennighausen L.** (2000) The microarray explorer tool for data mining of cDNA microarrays: application for the mammary gland. *Nucleic acids research* 28, 4452-4459.

**Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L. et al.** (2007) The diploid genome sequence of an individual human. *PLoS biology* 5, e254.

**Li H.** (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987-2993.

**Li H. & Durbin R.** (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

**Li H., Handsaker B., Wysoker A., Fennell T., Ruan J. et al.** (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

**Li R., Fan W., Tian G., Zhu H., He L. et al.** (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463, 311-317.

**Li Y., Chen W., Liu E.Y. & Zhou Y.H.** (2013) Single Nucleotide Polymorphism (SNP) Detection and Genotype Calling from Massively Parallel Sequencing (MPS) Data. *Statistics in biosciences* 5, 3-25.

**Lindblad-Toh K., Wade C.M., Mikkelsen T.S., Karlsson E.K., Jaffe D.B. et al.** (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803-819.

**Liu L., Li Y., Li S., Hu N., He Y. et al.** (2012) Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology* 2012, 251364.

**Loman N.J., Quick J. & Simpson J.T.** (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature methods* 12, 733–735.

**Love M.I., Huber W. & Anders S.** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, 550.

**Luhken G.** (2012) Genetic testing for phenotype-causing variants in sheep and goats. *Molecular and cellular probes* 26, 231-237.

**Luhken G., Fleck K., Pauciuolo A., Huisinga M. & Erhardt G.** (2012) Familiar hypopigmentation syndrome in sheep associated with homozygous deletion of the entire endothelin type-B receptor gene. *PloS one* 7, e53020.

**MacArthur D.G., Manolio T.A., Dimmock D.P., Rehm H.L., Shendure J. et al.** (2014) Guidelines for investigating causality of sequence variants in human disease. *Nature* 508, 469-476.

**Madoui M.A., Engelen S., Cruaud C., Belser C., Bertrand L. et al.** (2015) Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC genomics* 16, 327-015-1519-z.

**Maningat P.D., Sen P., Rijnkels M., Sunehag A.L., Hadsell D.L. et al.** (2009) Gene expression in the human mammary epithelium during lactation: the milk fat globule transcriptome. *Physiological genomics* 37, 12-22.

**Mardis E.R.** (2008) Next-generation DNA sequencing methods. *Annual review of genomics and human genetics* 9, 387-402.

**Marino R., Considine T., Sevi A., McSweeney P. & Kelly A.** (2005) Contribution of proteolytic activity associated with somatic cells in milk to cheese ripening. *International Dairy Journal* 15, 1026-1033.

**Marioni J.C., Mason C.E., Mane S.M., Stephens M. & Gilad Y.** (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* 18, 1509-1517.

**Maxam A.M. & Gilbert W.** (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* 74, 560-564.

**McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K. et al.** (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20, 1297-1303.

**McKusick V.A. & Ruddle F.H.** (1987) Toward a complete map of the human genome. *Genomics* 1, 103-106.

**McLaren W., Pritchard B., Rios D., Chen Y., Flicek P. & Cunningham F.** (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26, 2069-2070.

**McSweeney P.** (2004) Biochemistry of cheese ripening. *International Journal of Dairy Technology* 57, 127-144.

**Medeiros G.X., Riet-Correa F., Armien A.G., Dantas A.F., de Galiza G.J. & Simoes S.V.** (2012) Junctional epidermolysis bullosa in a calf. *Journal of veterinary diagnostic investigation* 24, 231-234.

**Medrano J.F., Cánovas A. & Islas-Trejo A.** (2014) RNA Sequencing for the Analysis of Complex Traits in Milk: Detection of Bacteria. In: *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*, No. 0137. Vancouver, Canada.

**Medrano J., Rincon G. & Islas-Trejo A.** (2010) Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. In: *Proceedings, 9th World Congress of Genetics Applied to Livestock Production*, No. 0852. Leipzig, Germany.

**Menoud A., Welle M., Tetens J., Lichtner P. & Drogemuller C.** (2012) A COL7A1 mutation causes dystrophic epidermolysis bullosa in Rotes Höhenvieh cattle. *PLoS one* 7, e38823.

**Metzker M.L.** (2010) Sequencing technologies - the next generation. *Nature reviews Genetics* 11, 31-46.

**Milenkovic D., Chaffaux S., Taourit S. & Guerin G.** (2003) A mutation in the LAMC2 gene causes the Herlitz junctional epidermolysis bullosa (H-JEB) in two French draft horse breeds. *Genetics, selection, evolution* 35, 249-256.

**modENCODE Consortium, Roy S., Ernst J., Kharchenko P.V., Kheradpour P. et al.** (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 330, 1787-1797.

**Momke S., Kerkmann A., Wohlke A., Ostmeier M., Hewicker-Trautwein M. et al.** (2011) A frameshift mutation within LAMC2 is responsible for Herlitz type junctional epidermolysis bullosa (HJEB) in black headed mutton sheep. *PLoS one* 6, e18943.

**Monteagudo L.V., Tejedor M.T., Ramos J.J., Lacasta D. & Ferrer L.M.** (2015) Ovine congenital myotonia associated with a mutation in the muscle chloride channel gene. *Veterinary journal* 204, 128-129.

**Montgomery G.W., Lord E.A., Penty J.M., Dodds K.G., Broad T.E. et al.** (1994) The Booroola fecundity (FecB) gene maps to sheep chromosome 6. *Genomics* 22, 148-153.

**Morozova O. & Marra M.A.** (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255-264.

**Mortazavi A., Williams B.A., McCue K., Schaeffer L. & Wold B.** (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5, 621-628.

**Mulsant P., Lecerf F., Fabre S., Schibler L., Monget P. et al.** (2001) Mutation in bone morphogenetic protein receptor-1B is associated with increased ovulation rate in Booroola Merino ewes. *Proceedings of the National Academy of Sciences of the United States of America* 98, 5104-5109.

**Murgiano L., Jagannathan V., Calderoni V., Joechler M., Gentile A. & Drogemuller C.** (2014) Looking the cow in the eye: deletion in the NID1 gene is associated with recessive inherited cataract in Romagnola cattle. *PLoS one* 9, e110628.

**Nawy T.** (2014) Single-cell sequencing. *Nature methods* 11, 18.

**Nemir M., Bhattacharyya D., Li X., Singh K., Mukherjee A.B. & Mukherjee B.B.** (2000) Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency. *The Journal of biological chemistry* 275, 969-976.

**Nezamzadeh R., Seubert A., Pohlenz J. & Brenig B.** (2005) Identification of a mutation in the ovine uroporphyrinogen decarboxylase (UROD) gene associated with a type of porphyria. *Animal Genetics* 36, 297-302.

**Ng S.B., Buckingham K.J., Lee C., Bigham A.W., Tabor H.K. et al.** (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics* 42, 30-35.

**Nicholas F.W. & Hobbs M.** (2014) Mutation discovery for Mendelian traits in non-laboratory animals: a review of achievements up to 2012. *Animal Genetics* 45, 157-170.

**Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino**

**Nicholas F.W.** (2010) A revolution in mapping and identifying the causal mutation of single-gene disorders. In: *Introduction to Veterinary Genetics, 3rd Edition*. (ed. by F.W. Nicholas), pp. 96-97, 98. Wiley-Blackwell, Oxford, UK.

**Ntambi J.M. & Miyazaki M.** (2004) Regulation of stearoyl-CoA desaturases and role in metabolism. *Progress in lipid research* 43, 91-104.

**Olivry T., Poujade-Delverdier A., Dunston S.M., Fine J. & Ortonne J.** (1997) Absent expression of collagen XVII (BPAG2, BP180) in canine familial localized junctional epidermolysis bullosa. *Veterinary dermatology* 8, 203-212.

**Oshlack A. & Wakefield M.J.** (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biology direct* 4, 14-6150-4-14.

**Pareek C.S., Smoczynski R. & Tretyn A.** (2011) Sequencing technologies and genome sequencing. *Journal of Applied Genetics* 52, 413-435.

**Pariza M.W., Park Y. & Cook M.E.** (1999) Conjugated linoleic acid and the control of cancer and obesity. *Toxicological sciences* 52, 107-110.

**Park Y., Juárez M., Ramos M. & Haenlein G.** (2007) Physico-chemical characteristics of goat and sheep milk. *Small ruminant research* 68, 88-113.

**Park P.J.** (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews.Genetics* 10, 669-680.

**Peters M., Reber I., Jagannathan V., Raddatz B., Wohlsein P. & Drogemuller C.** (2015) DNA-based diagnosis of rare diseases in veterinary medicine: a 4.4 kb deletion of ITGB4 is associated with epidermolysis bullosa in Charolais cattle. *BMC veterinary research* 11, 366-015-0366-0.

**Pettersson E., Lundeberg J. & Ahmadian A.** (2009) Generations of sequencing technologies. *Genomics* 93, 105-111.

**Pontius J.U., Mullikin J.C., Smith D.R., Agencourt Sequencing Team, Lindblad-Toh K. et al.** (2007) Initial sequence and comparative analysis of the cat genome. *Genome research* 17, 1675-1689.

**Porada C.D., Sanada C., Long C.R., Wood J.A., Desai J. et al.** (2010) Clinical and molecular characterization of a re-established line of sheep exhibiting hemophilia A. *Journal of thrombosis and haemostasis* 8, 276-285.

**Poulter J.A., El-Sayed W., Shore R.C., Kirkham J., Inglehearn C.F. & Mighell A.J.** (2014) Whole-exome sequencing, without prior linkage, identifies a mutation in LAMB3 as a cause of dominant hypoplastic amelogenesis imperfecta. *European journal of human genetics* 22, 132-135.

**Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A. et al.** (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559-575.

**Pyorala S.** (2003) Indicators of inflammation in the diagnosis of mastitis. *Veterinary research* 34, 565-578.

**Qiao R., He Y., Pan B., Xiao S., Zhang X. et al.** (2015) Understanding the molecular mechanisms of human microtia via a pig model of HOXA1 syndrome. *Disease Models & Mechanisms*, dmm. 018291.

**Rabbani B., Tekin M. & Mahdieh N.** (2014) The promise of whole-exome sequencing in medical genetics. *Journal of human genetics* 59, 5-15.

**Ratnakumar A., Kirkness E.F. & Dalrymple B.P.** (2010) Quality control of the sheep bacterial artificial chromosome library, CHORI-243. *BMC research notes* 3, 334-0500-3-334.

**Rehm H.L.** (2013) Disease-targeted sequencing: a cornerstone in the clinic. *Nature reviews.Genetics* 14, 295-300.

***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

**Reicher S., Seroussi E. & Gootwine E.** (2010) A mutation in gene CNGA3 is associated with day blindness in sheep. *Genomics* 95, 101-104.

**Robinson M.D., McCarthy D.J. & Smyth G.K.** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

**Robinson M.D. & Oshlack A.** (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* 11, R25-2010-11-3-r25.

**Rougemont J. & Naef F.** (2012) Computational analysis of protein-DNA interactions from ChIP-seq data. *Methods in molecular biology* 786, 263-273.

**Royaux I., Bernier B., Montgomery J.C., Flaherty L. & Goffinet A.M.** (1997) Reln(rl-Alb2), an allele of reeler isolated from a chlorambucil screen, is due to an IAP insertion with exon skipping. *Genomics* 42, 479-482.

**Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R. et al.** (1977a) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265, 687-695.

**Sanger F., Nicklen S. & Coulson A.R.** (1977b) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5463-5467.

**Sartelet A., Li W., Pailhoux E., Richard C., Tamma N. et al.** (2015) Genome-wide next-generation DNA and RNA sequencing reveals a mutation that perturbs splicing of the phosphatidylinositol glycan anchor biosynthesis class H gene (PIGH) and causes arthrogyrosis in Belgian Blue cattle. *BMC genomics* 16, 316-015-1528-y.

**Saunders C.J., Miller N.A., Soden S.E., Dinwiddie D.L., Noll A. et al.** (2012) Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Science translational medicine* 4, 154ra135.



**Schaapveld R.Q., Borradori L., Geerts D., van Leusden M.R., Kuikman I. et al.** (1998) Hemidesmosome formation is initiated by the beta4 integrin subunit, requires complex formation of beta4 and HD1/plectin, and involves a direct interaction between beta4 and the bullous pemphigoid antigen 180. *The Journal of cell biology* 142, 271-284.

**Schack L., Stapulionis R., Christensen B., Kofod-Olsen E., Skov Sorensen U.B. et al.** (2009) Osteopontin enhances phagocytosis through a novel osteopontin receptor, the alphaXbeta2 integrin. *Journal of immunology* 182, 6943-6950.

**Schadt E.E., Turner S. & Kasarskis A.** (2010) A window into third-generation sequencing. *Human molecular genetics* 19, R227-40.

**Schena M., Shalon D., Davis R.W. & Brown P.O.** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.

**Schloss J.A.** (2008) How to get genomes at one ten-thousandth the cost. *Nature biotechnology* 26, 1113-1115.

**Shaw J. & Diamandis E.** (2007) Distribution of 15 human kallikreins in tissues and biological fluids. *Clinical chemistry* 53, 1423-1432.

**Sheehy P.A., Riley L.G., Raadsma H.W., Williamson P. & Wynn P.C.** (2009) A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Animal Genetics* 40, 492-498.

**Shen Y., Yue F., McCleary D.F., Ye Z., Edsall L. et al.** (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116-120.

**Shu D.P., Chen B.L., Hong J., Liu P.P., Hou D.X. et al.** (2012) Global transcriptional profiling in porcine mammary glands from late pregnancy to peak lactation. *Omic* 16, 123-137.

**Sivasubbu S., Sachidanandan C. & Scaria V.** (2013) Time for the zebrafish ENCODE. *Journal of genetics* 92, 695-701.

**Soon W.W., Hariharan M. & Snyder M.P.** (2013) High-throughput sequencing for biology and medicine. *Molecular systems biology* 9, 640.

**Spirito F., Charlesworth A., Linder K., Ortonne J.P., Baird J. & Meneguzzi G.** (2002) Animal models for skin blistering conditions: absence of laminin 5 causes hereditary junctional mechanobullous disease in the Belgian horse. *The Journal of investigative dermatology* 119, 684-691.

**Stepp M.A., Spurr-Michaud S., Tisdale A., Elwell J. & Gipson I.K.** (1990) Alpha 6 beta 4 integrin heterodimer is a component of hemidesmosomes. *Proceedings of the National Academy of Sciences of the United States of America* 87, 8970-8974.

**Suarez-Vega A., Gutierrez-Gil B., Benavides J., Perez V., Tosser-Klopp G. et al.** (2015) Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. *PloS one* 10, e0126416.

**Suarez-Vega A., Gutierrez-Gil B., Cuchillo-Ibanez I., Saez-Valero J., Perez V. et al.** (2013) Identification of a 31-bp deletion in the RELN gene causing lissencephaly with cerebellar hypoplasia in sheep. *PloS one* 8, e81072.

**Takahara T., Ohsumi T., Kuromitsu J., Shibata K., Sasaki N. et al.** (1996) Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. *Human molecular genetics* 5, 989-993.

**Takeichi T., Liu L., Fong K., Ozoemena L., McMillan J.R. et al.** (2014) Whole-exome sequencing improves mutation detection in a diagnostic epidermolysis bullosa laboratory. *The British journal of dermatology* 172(1), 94-100.

**Tammen I., Houweling P.J., Frugier T., Mitchell N.L., Kay G.W. et al.** (2006) A missense mutation (c.184C>T) in ovine CLN6 causes neuronal ceroid lipofuscinosis in Merino sheep whereas affected South Hampshire sheep have reduced levels of CLN6 mRNA. *Biochimica et biophysica acta* 1762, 898-905.

- Tan P., Allen J.G., Wilton S.D., Akkari P.A., Huxtable C.R. & Laing N.G.** (1997) A splice-site mutation causing ovine McArdle's disease. *Neuromuscular disorders* 7, 336-342.
- Taylor J., Decker J. & Schnabel R.** (2014) Animal Breeding in the Next-Generation Sequencing Era. In: *Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Plenary-1. Vancouver, Canada.*
- Teyssot B. & Houdebine L.M.** (1981) Role of progesterone and glucocorticoids in the transcription of the beta-casein and 28-S ribosomal genes in the rabbit mammary gland. *European journal of biochemistry* 114, 597-608.
- Toral P.G., Hervas G., Gomez-Cortes P., Frutos P., Juarez M. & de la Fuente M.A.** (2010) Milk fatty acid profile and dairy sheep performance in response to diet supplementation with sunflower oil plus incremental levels of marine algae. *Journal of dairy science* 93, 1655-1667.
- Torres P.A., Zeng B.J., Porter B.F., Alroy J., Horak F. et al.** (2010) Tay-Sachs disease in Jacob sheep. *Molecular genetics and metabolism* 101, 357-363.
- Trapnell C., Roberts A., Goff L., Pertea G., Kim D. et al.** (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 7, 562-578.
- Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G. et al.** (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28, 511-515.
- Tynnela J., Sohar I., Sleat D.E., Gin R.M., Donnelly R.J. et al.** (2001) Congenital ovine neuronal ceroid lipofuscinosis--a cathepsin D deficiency with increased levels of the inactive enzyme. *European journal of paediatric neurology* 5 Suppl A, 43-45.
- van der Neut R., Krimpenfort P., Calafat J., Niessen C.M. & Sonnenberg A.** (1996) Epithelial detachment due to absence of hemidesmosomes in integrin beta 4 null mice. *Nature genetics* 13, 366-369.

**van Dijk E.L., Auger H., Jaszczyszyn Y. & Thermes C.** (2014) Ten years of next-generation sequencing technology. *Trends in genetics* 30, 418-426.

**Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J. et al.** (2001) The sequence of the human genome. *Science* 291, 1304-1351.

**Vilain C. & Vassart G.** (2004) Small amplified RNA-SAGE. *Methods in molecular biology* 258, 135-152.

**Wade C.M., Giulotto E., Sigurdsson S., Zoli M., Gnerre S. et al.** (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326, 865-867.

**Walker K.M.** (2012) Molecular Characterization of Ovine GM1 Gangliosidosis. *Doctoral dissertation*. University of Illinois at Urbana-Champaign, IL, USA. URL: [https://www.ideals.illinois.edu/bitstream/handle/2142/42257/Kristen\\_Walker.pdf?sequence=1](https://www.ideals.illinois.edu/bitstream/handle/2142/42257/Kristen_Walker.pdf?sequence=1)

**Wang J., Duncan D., Shi Z. & Zhang B.** (2013) WEB-based GENE SeT Analysis Toolkit (WebGestalt): update 2013. *Nucleic acids research* 41, W77-83.

**Wang K., Li M. & Hakonarson H.** (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 38, e164.

**Wang Z., Gerstein M. & Snyder M.** (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews.Genetics* 10, 57-63.

**Wickramasinghe S., Canovas A., Rincón G. & Medrano J.F.** (2014) RNA-sequencing: a tool to explore new frontiers in animal genetics. *Livestock Science* 166, 206-216.

**Wickramasinghe S., Rincon G., Islas-Trejo A. & Medrano J.F.** (2012) Transcriptional profiling of bovine milk using RNA sequencing. *BMC genomics* 13, 45-2164-13-45.

**Wickramasinghe S., Hua S., Rincon G., Islas-Trejo A., German J.B. et al.** (2011) Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-sequencing. *PloS one* 6, e18895.

**Willig L.K., Petrikin J.E., Smith L.D., Saunders C.J., Thiffault I. et al.** (2015) Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *The Lancet. Respiratory medicine* 3, 377-387.

**Wong G.K., Liu B., Wang J., Zhang Y., Yang X. et al.** (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432, 717-722.

**Yamaguchi A., Hori O., Stern D.M., Hartmann E., Ogawa S. & Tohyama M.** (1999) Stress-associated endoplasmic reticulum protein 1 (SERP1)/Ribosome-associated membrane protein 4 (RAMP4) stabilizes membrane proteins during stress and facilitates subsequent glycosylation. *The Journal of cell biology* 147, 1195-1204.

**Yang Y., Muzny D.M., Reid J.G., Bainbridge M.N., Willis A. et al.** (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England journal of medicine* 369, 1502-1511.

**Yokoi N., Namae M., Wang H.Y., Kojima K., Fuse M. et al.** (2003) Rat neurological disease creeping is caused by a mutation in the reelin gene. *Brain research. Molecular brain research* 112, 1-7.

**Yousef G. & Diamandis E.** (2001) The New Human Tissue Kallikrein Gene Family: Structure, Function, and Association to Disease 1. *Endocrine reviews* 22, 184-204.

**Yue F., Cheng Y., Breschi A., Vierstra J., Wu W. et al.** (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355-364.

**Zhang W., Chen J., Yang Y., Tang Y., Shang J. & Shen B.** (2011) A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. *PLoS one* 6, e17915.

**Zhao X., Onteru S.K., Dittmer K.E., Parton K., Blair H.T. et al.** (2012a) A missense mutation in AGTPBP1 was identified in sheep with a lower motor neuron disease. *Heredity* 109, 156-162.

***Utilización de herramientas genómicas en el estudio de la arquitectura molecular de caracteres de importancia económica en el ganado ovino***

**Zhao X., Onteru S.K., Piripi S., Thompson K.G., Blair H.T. et al.** (2012b) In a shake of a lamb's tail: using genomics to unravel a cause of chondrodysplasia in Texel sheep. *Animal Genetics* 43 Suppl 1, 9-18.

**Zhao X., Dittmer K.E., Blair H.T., Thompson K.G., Rothschild M.F. & Garrick D.J.** (2011) A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. *PloS one* 6, e21739.

**Zhou H., Hickford J.G. & Fang Q.** (2012) A premature stop codon in the ADAMTS2 gene is likely to be responsible for dermatosparaxis in Dorper sheep. *Animal Genetics* 43, 471-473.

## **PÁGINAS WEB CITADAS**

**ANCHE - Asociación de Criadores de Ganado Selecto de la raza Churra**

<http://www.anche.org/>

**ENSEMBL- Ensembl Genome Browser**

<http://www.ensembl.org/>

**ENSEMBL- Sheep assembly and gene annotation**

[http://www.ensembl.org/Ovis\\_aries/Info/Annotation](http://www.ensembl.org/Ovis_aries/Info/Annotation)

**FAANG – Functional Annotation of Animal Genomes**

<http://www.faang.org/>

**FAOSTAT - Statistics Division of the Food and Agriculture Organization of the United Nations**

<http://faostat3.fao.org/home/E>

**FASTQC - A quality control tool for high throughput sequence data**

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

**FASTX-Toolkit - FASTQ/A short-reads pre-processing tools**

[http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)

**HUMAN GENOME SEQUENCING CENTER- Bovine Genome Project**

<https://www.hgsc.bcm.edu//other-mammals/bovine-genome-project>

**ISGC - International Sheep Genomics Consortium**

<http://www.sheephapmap.org/>

**MAGRAMA - Ministerio de Agricultura, Alimentación y Medio Ambiente**

<http://www.magrama.gob.es/es/>

**OMIA – Online Mendelian Inheritance in Animals**

<http://omia.angis.org.au/>

**PICARD - A set of tools (in Java) for working with next generation sequencing data**

<http://picard.sourceforge.net>





## 11. Anexo I

### Artículos Adicionales:

**Pérez V., Suárez-Vega A., Fuertes M., Benavides J., Delgado L., Ferreras M.C. & Arranz J.J. (2013)** Hereditary lissencephaly and cerebellar hypoplasia in Churra lambs. BMC Veterinary Research: 9:156. doi: 10.1186/1746-6148-9-156

**Benavides J., Suárez-Vega A., Gutiérrez J., de la Hera S., Fuertes M., Delgado L., Fernández M., Ferreras M.C., Arranz J.J. & Pérez V. (2015)** Generalized severe junctional epidermolysis bullosa with congenital absence of skin in churra lambs. Veterinary Dermatology. doi: 10.1111/vde.12226





## **Hereditary lissencephaly and cerebellar hypoplasia in Churra lambs.**

Valentín Pérez<sup>1</sup>, Aroa Suárez-Vega<sup>2</sup>, Miguel Fuertes<sup>1</sup>, Julio Benavides<sup>1</sup>, Laetitia Delgado<sup>1</sup>, M Carmen Ferreras<sup>1</sup> and Juan José Arranz<sup>2</sup>.

<sup>1</sup>Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE), Facultad de Veterinaria, Universidad de León, León, Spain;

<sup>2</sup>Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain.

BMC Veterinary Research: 9:156. doi: 10.1186/1746-6148-9-156.



RESEARCH ARTICLE

Open Access

# Hereditary lissencephaly and cerebellar hypoplasia in Churra lambs

Valentín Pérez<sup>1\*</sup>, Aroa Suárez-Vega<sup>2</sup>, Miguel Fuertes<sup>1</sup>, Julio Benavides<sup>1</sup>, Laetitia Delgado<sup>1</sup>, M Carmen Ferreras<sup>1</sup> and Juan José Arranz<sup>2</sup>

## Abstract

**Background:** Lissencephaly is a rare developmental brain disorder in veterinary and human medicine associated with defects in neuronal migration leading to a characteristic marked reduction or absence of the convolitional pattern of the cerebral hemispheres. In many human cases the disease has a genetic basis. In sheep, brain malformations, mainly cerebellar hypoplasia and forms of hydrocephalus, are frequently due to *in utero* viral infections. Although breed-related malformations of the brain have been described in sheep, breed-related lissencephaly has not been previously recorded in a peer reviewed publication.

**Results:** Here we report neuropathological findings in 42 newborn lambs from a pure Churra breed flock, with clinical signs of weakness, inability to walk, difficulty in sucking and muscular rigidity observed immediately after birth. All the lambs showed near-total agyria with only a rudimentary formation of few sulci and gyri, and a severe cerebellar hypoplasia. On coronal section, the cerebral grey matter was markedly thicker than that of age-matched unaffected lambs and the ventricular system was moderately dilated. Histologically, the normal layers of the cerebral cortex were disorganized and, using an immunohistochemical technique against neurofilaments, three layers were identified instead of the six present in normal brains. The hippocampus was also markedly disorganised and the number and size of lobules were reduced in the cerebellum. Heterotopic neurons were present in different areas of the white matter. The remainder of the brain structures appeared normal. The pathological features reported are consistent with the type LCH-b (lissencephaly with cerebellar hypoplasia group b) defined in human medicine. No involvement of pestivirus or bluetongue virus was detected by immunohistochemistry. An analysis of pedigree data was consistent with a monogenic autosomal recessive pattern inheritance.

**Conclusions:** The study describes the clinical and pathological findings of lissencephaly with cerebellar hypoplasia in Churra lambs for which an autosomal recessive inheritance was the most likely cause. Histopathological features observed in the cerebral cortex and hippocampus are consistent with a possible failure in neuronal migration during brain development. This report suggests that lissencephaly should be considered in the differential diagnosis of congenital neurological disease in newborn lambs showing weakness, inability to walk and difficulty sucking.

**Keywords:** Lissencephaly, Cerebellar hypoplasia, Agyria-pachygyria, Sheep, Autosomal recessive

## Background

Neuronal migration disorders represent a group of rare brain malformations which account for a wide range of cortical anomalies, the most severe forms being represented by lissencephalies [1-3]. Lissencephaly or agyria-pachygyria designates a malformation characterized by a

simplified convolitional pattern in which there may be no or only a few broad gyri separated by rudimentary primary fissures and sulci; agyria refers to the extreme in which the brain shows a smooth surface over all the cerebral hemispheres [4]. This term was proposed in humans to distinguish the flat brains of lower mammalian species and at present, several cases have been reported and different studies made in order to clarify its origin [1-3,5,6]. Two main forms of lissencephaly have been described: i) classical or type I lissencephaly, characterized by agyria (total loss of gyri) or pachygyria

\* Correspondence: vperp@unileon.es

<sup>1</sup>Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE), Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain

Full list of author information is available at the end of the article

(few and broadened gyri) together with a very thick cortical gray matter layer, normal brainstem and grossly normal cerebellum or mild vermis hypoplasia, and ii) cobblestone or type II lissencephaly, in which affected children have an agyric brain, with a verrucous appearance and a less marked thickening of the cortex than in type I lissencephaly [1,6,7]. Microscopically, in type I lissencephaly the agyric cortex is mostly composed of four coarse layers: a molecular layer, an outer cellular layer, a sparsely cellular layer and a thick inner cellular layer [7,8]. Type II lissencephaly is characterized by a complete disorganization of the cortex, without discernible layering, and by the presence of gliomesenchymal cell proliferations and neuroglial heterotopia in the leptomeninges [8].

More recently, lissencephaly with cerebellar hypoplasia (LCH) has been recognized as a distinct category of brain malformation, consisting of type I or II lissencephaly associated with cerebellar underdevelopment [1,9-11]. Among LCH, six different phenotypes have been described in children, according to the degree and characteristics of cerebellar hypoplasia and the presence of alterations in other areas such as the corpus callosum and the hippocampus [11].

Lissencephaly in humans is a congenital and genetically heterogeneous disease [3,5,12]. Mutations or deletions in different genes have been identified as involved in lissencephaly cases: the *LIS* gene is responsible for the autosomal form of classical lissencephaly [13] and the doublecortin gene (*DCX* or *X-LIS*) is X-linked [14] whereas homozygous *RELN* gene has been identified in cases of LCH [10]. Recently, mutations in gene *TUBA1A*, coding for alpha I tubulin, have been also described in some cases of LCH [15].

In domestic animals, lissencephaly is an extremely rare brain malformation. Sporadic cases have been described in dogs [16-18], where a breed predisposition for this malformation has been reported in Lhasa Apso dogs [16,17] and more recently in cats, in which lissencephaly was detected together with microencephaly and hypoplasia of cerebellum and corpus callosum [18,19]. The developmental mechanisms responsible for the absence of gyral formation in animals are not known but a genetic basis in many cases is presumed [18].

In sheep, brain malformations, mainly cerebellar hypoplasia, have been widely described as associated with *in utero* viral infections caused by orbiviruses [18], pestiviruses [20] and arboviruses such as Cache Valley [21] or Schmallenberg viruses [22], variably accompanied by arthrogryposis or other anomalies, such as hydrocephalus or porencephaly.

The purpose of the present report is to describe the clinical, pathologic and pedigree analyses carried out in several cases of lissencephaly in newborn lambs from a pedigree Churra dairy sheep flock.

## Results

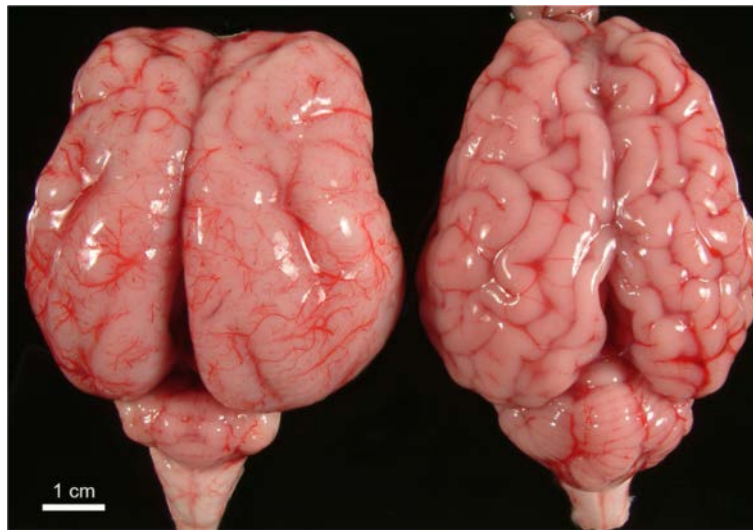
### Case presentation

The affected lambs were from a dairy flock of approximately 1600 adult pedigree Churra breed sheep, including 35 purebred males. The flock is located in the Valladolid province of Castilla y León region (North-West of Spain) and is managed under a semi-extensive system, where the animals graze the pasture in appropriate weather conditions and are kept indoors during the lambing and lactation periods. On average ~2.500 lambs are born in the flock each year. In 2005, the farmer observed some newborn lambs with signs of weakness, shivering, muscular rigidity and recumbency. The animals were unable to stand, had difficulties in sucking their dams and lacked coordination. A total of 23 affected lambs (approximately 0.92% of the total crop in that lambing season) were examined and diagnosed as affected by the same clinicopathological entity. Additionally, another 19 affected lambs born in subsequent years were included in this study. A total of 11 newborn lambs from the same farm but dying from unrelated diseases (watery mouth disease, hypothermia) were used as unaffected controls, after checking that no malformations were present in the brain.

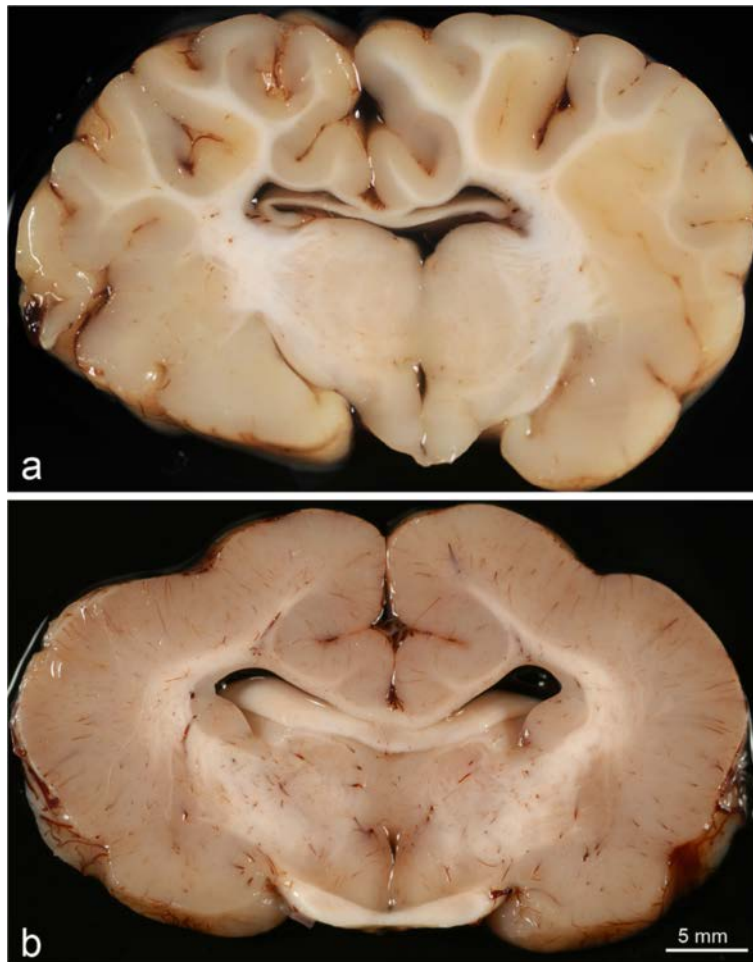
### Pathological findings

All the affected lambs showed the same pathological features. A slight reduction in cerebrum size was noted for affected lambs (mean values 5.2 cm length  $\times$  6.0 cm width  $\times$  3.5 cm height) compared to control lambs (6.4  $\times$  5.7  $\times$  4.3). Agyria was the most remarkable change in the cerebral hemispheres, being particularly evident in the temporal, parietal and occipital lobes, where the cerebral surface had a smooth appearance (Figure 1). Rudimental formation of gyri and sulci were observed in the lateral and rostral areas (Figure 1), probably corresponding to suprasylvian, presylvian and ansate sulci. The cerebellum showed a marked hypoplasia affecting both cerebellar hemispheres and vermis with loss of discernible folial pattern (Figure 1). The mean measurements of the cerebellum of affected lambs (1.4  $\times$  2.7  $\times$  0.8 cm) were clearly reduced compared with unaffected lambs (3.1  $\times$  4.2  $\times$  2 cm).

On coronal sections of the brain, in addition to agyria, a marked pachygyria was observed in all the regions of the cerebral cortex. It was characterized by the presence of a thick grey matter layer, with a depth of 9 mm in average, with broad gyri and very shallow sulci, when present (Figure 2). The white matter was narrow, leading to an inverted grey to white matter ratio. The cerebral grey matter-white matter interface was linear and well defined with rare interdigitations. Macroscopically, the remaining areas of the forebrain (corpus callosum, cingulate gyrus (Figure 2) and hippocampus) and brain stem appeared normally developed. The posterior part



**Figure 1 Affected and normal brains.** Comparative image of a brain showing lissencephaly (left), with few and poorly developed sulci, and a marked cerebellar hypoplasia, and a brain from a newborn control lamb (right).



**Figure 2 Coronal section (fixed tissue) of normal (a) and lissencephalic (b) brains.** A marked increase of thickness of the cerebral cortex is seen in the affected lamb (b). Note that the corpus callosum is normally developed.



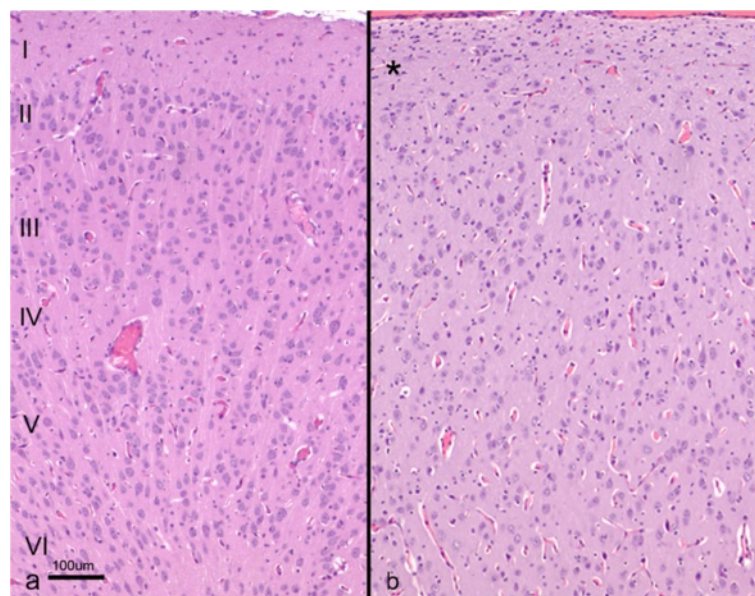
of the lateral ventricles was moderately dilated in 13 lambs but no physical obstruction of CSF flow was detected. There was a substantial reduction in the depth of the cerebellar cortex and the folia were markedly decreased in size and number.

Histopathological examination was carried out in samples from the cerebral cortex of the most severely affected area (temporo-parietal region) in all the animals. The sagittal sections of the cerebral cortex lacked the typical six-layered structure characteristic of this area, seen in the control lambs (Figure 3). Neither glial cells nor neurons were detected in the meningeal layers (Figure 3). In both HE and cresyl violet stained sections, beneath the pial surface, a cell-sparse layer containing neurons smaller than those present in the rest of the cortex and resembling morphologically the molecular or layer I of normal brains, was observed (Figure 3). Below this layer, a wide, more densely populated cellular layer was formed by neuronal bodies of different sizes, some of which had a pyramidal appearance (Figure 3). No laminar organisation or vertical linear array of neurones was identified within this wide layer. The narrow white matter layer was normally myelinated and contained heterotopic neuronal cell bodies.

In the control brain sections immunolabelled with antibodies against NF-L neurofilaments, the six conventional layers were identified in the cerebral cortex, and the two pyramidal neurone layers (III and V) had the strongest intensity of immunostaining (Figure 4a). In

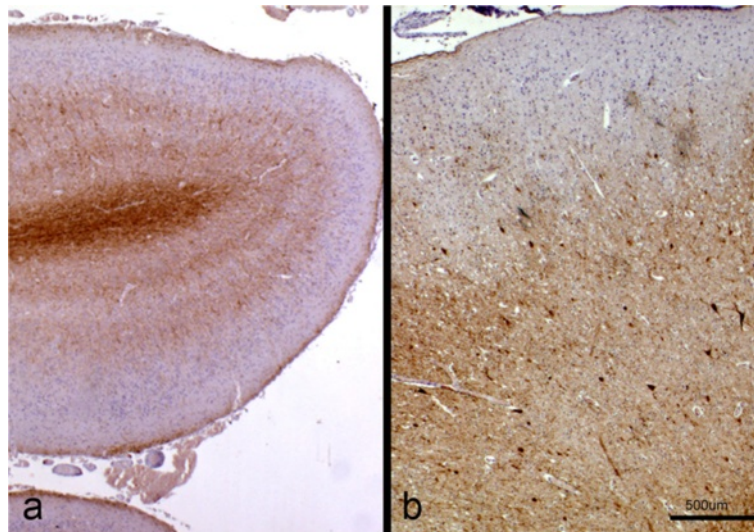
contrast, in the sections from the affected brains, only three layers were discernible (Figure 4b). Besides the outer marginal zone, corresponding to the molecular layer, that did not show immunolabelling, a thick layer formed mainly by pyramidal neurons of different sizes, some of them showing intense neurofilament expression, was present (Figure 4b). Finally, an internal layer not discernible in HE sections, adjacent to the white matter, with a more reduced immunostaining, was identified (Figure 4b).

The corpus callosum was present and normally developed in all the lambs, while the hippocampus was disorganized (Figure 5). The pyramidal layer was sharply delineated in the hippocampus of unaffected lambs, whereas in affected lambs several layers of pyramidal neurones were identified interspersed with large numbers of other neurons and glial cells. The dentate gyrus could be discerned in the majority of the lambs, but in some cases it was not clearly evident. The cerebellum of the affected animals showed a marked reduction in size, complexity and number of folia. Lobules were poorly developed and usually the cerebellar cortex was reduced to a single band of folia, which were notably shortened and separated by shallow and wide sulci (Figure 6). Only occasionally up to two folds of folia were seen (Figure 6). The three layers, characteristic of the cerebellar cortex histological architecture, were easily detected in each folium, but Purkinje neuronal heterotopia was commonly seen (Figure 7).



**Figure 3** Histological section of the cerebral cortex from a control (a) and lissencephalic (b) brain. Both sections were taken at the same magnification. Whereas in the control brain the whole cortex can be seen in the picture and organized in six layers (I to VI), in the lissencephalic brain only the most superficial part of the cortex is shown (aprox. 40% of the whole thickness), due to the thickening of this layer. A sparse-cellular layer underneath the piamater can be identified (\*), whereas in the rest of the gray matter the neurons appear disorganized. H-E.





**Figure 4 Immunohistochemistry for neurofilaments.** Sections of the cerebral cortex from a control (a) and a lissencephalic (b) brain, taken at the same magnification and immunolabelled for neurofilaments. In the control brain, the typical layering of this area (I to VI) can be seen, where the neurons from the pyramidal layers show the strongest signal. In the lissencephalic brain, three different layers can be identified according to neurofilament expression. Immunoperoxidase staining for neurofilaments.

No macroscopic or microscopic changes were detected in any of the other organs of the affected lambs. Pestivirus and bluetongue virus antigens were not detected by immunohistochemical methods.

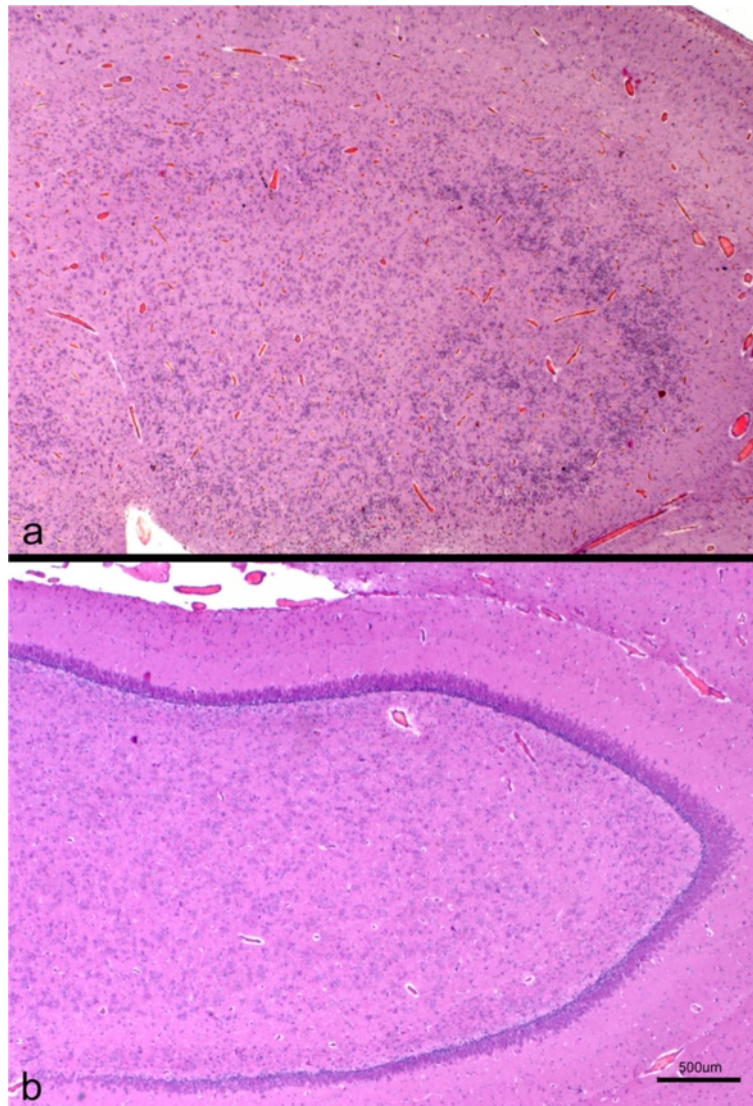
#### Pedigree analysis

The analysis of the three pedigrees, including 14 affected animals, with available DNA, was consistent with a monogenic autosomal recessive inheritance (Figure 8). In the three families some additional healthy lambs were analyzed. Overall, four rams were involved: ram #11 mated 46 ewes that lambed 63 unaffected lambs, 7 ewes that lambed 8 affected lambs and one ewe that lambed a non-affected ewe which, when mated to ram #112 produced an affected lamb; ram #21 mated 23 ewes that produced 30 non-affected lambs and 4 ewes that lambed 4 affected and 1 non-affected lambs; ram #31 mated 16 ewes that produced 20 non-affected lambs and one ewe that produced an affected lamb in one pregnancy and a non-affected lamb in a subsequent parturition.

#### Discussion

This study describes forty-two cases of lissencephaly with cerebellar hypoplasia in newborn Churra lambs, occurring over consecutive years, in the same flock. The occurrence of brain malformations of similar type in multiple lambs is not commonly documented in sheep flocks. Most of such cases have been related to *in utero* viral infections by pestiviruses (Border Disease virus), orthobunyaviruses (Schmallenberg, Akabane or Cache valley viruses) or orbiviruses (bluetongue virus) [20-22,24,25].

In our study, evidence of a possible viral etiology of these cases involving the main teratogenic viruses present in our area was not detected by immunohistochemical analyses. Besides, the histological characteristics observed in the cerebral cortex of these lambs are distinct from those occurring in the cerebrum in cases of viral teratogenesis. Porencephaly and hydranencephaly, cerebellar lesions and arthrogryposis are the most common malformations associated with orbiviruses, pestiviruses and orthobunyaviruses respectively. Moreover, immunohistochemical evidence of border disease virus and bluetongue virus was not detected, and the occurrence of these cases predates the emergence of Schmallenberg virus. Although lissencephaly has been reported to occur in association with ovine pestivirus [26] and Cache valley virus [21] infections, the lesions were not described in detail and the lack of gyral formation may have been associated with underlying cortical lesions such as porencephaly. This lesion would therefore differ from the strict definition of lissencephaly. No other cases of lissencephaly were reported from other flocks of Churra or Assaf breed in the area over the same time period. Furthermore, pedigree analysis has demonstrated that lissencephaly cases in this study were compatible with a monogenic trait following an autosomal recessive inheritance, suggesting that lissencephaly is associated with a genetic defect segregating in this flock, rather than a teratogenic infectious disease. This hypothesis has been also proposed in the cases of lissencephaly previously reported in veterinary medicine, that have occurred predominantly in Lhasa Apso dogs [16-18], suggesting a marked breed predisposition. In this study all the cases



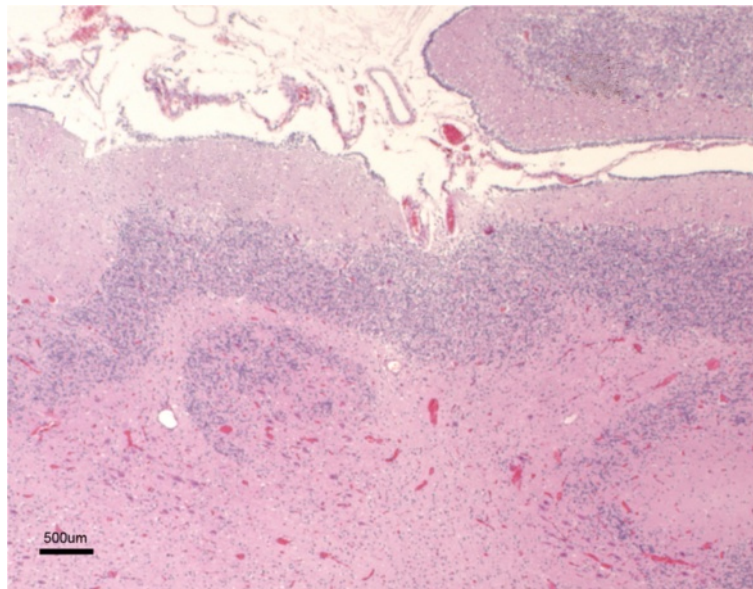
**Figure 5** Cross section of the hippocampus from a lissencephalic (a) and control (b) brain. Both sections were taken at the same magnification. Whereas in the control brain there is an evident and organized layer of pyramidal neurons, this region shows a marked cellular disorganization in the lissencephalic brain, with several layers of neuronal and glial cells interspersed. H-E.

occurred in purebred Churra lambs, as well as a previously documented case [27] that occurred in the same region but in a different flock with no known direct connection with the flock in this report. However, since many other Churra flocks do not appear to suffer from this condition, it would be premature to hypothesize about a breed-related genetic defect or predisposition.

The exact date of the first occurrence of lissencephaly in this flock is not known. It was not until the farmer noticed that a significant number of animals suffered the disease that veterinary assistance was sought. Given the number of cases in the first year we can make a rough estimate of the frequency of the putative allele causing the disease in the flock. Assuming that the population is

in H-W equilibrium, with 23 affected lambs in a total of 2500 lambs born, the allele frequency for the mutated allele in the flock would be approximately 0.09. Since no new rams from other farms have joined the flock in recent years, the most likely reason for this high frequency for a lethal allele will be inbreeding leading to an increased proportion of the flock being carriers of the putative autosomal recessive trait and hence increase the likelihood of mating of carriers leading to birth of affected lambs. Furthermore it should be noted that this problem has not been detected in any other flock belonging to the Churra breeders association. It is also important to note that none of the rams used in the artificial insemination program carried out by the Churra



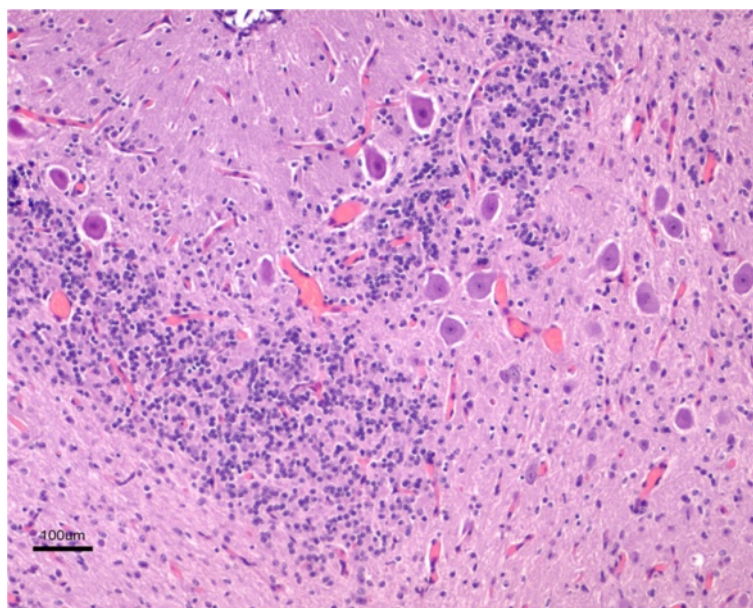


**Figure 6 Marked cerebellar hypoplasia.** The cerebellum shows a reduced number of very short folia. Note that the typical layering of the cortex can be identified. H-E.

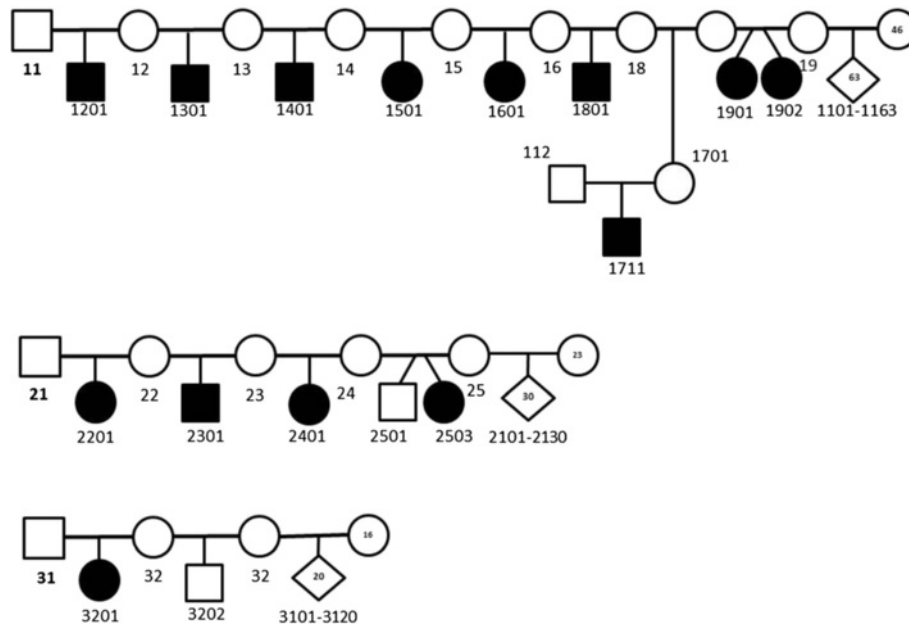
selection scheme has noticed to have any offspring affected by lissencephaly.

In humans, although it is considered a rare disorder, lissencephaly is a well-recognized entity that belongs to the group of malformations caused by abnormal neuronal migration, with two main pathological syndromes described: type I (classical) and type II (cobblestone) [1,7,8]. According to the gross (absence of a verrucous appearance of the cortical surface) and microscopic

findings, with no leptomenigeal neuronal heterotopia, the lissencephaly cases recorded in this study resemble the so-called type I or classical lissencephaly of humans. However, the constant presence of cerebellar hypoplasia in these lambs, is consistent with the more recently recognized category of human lissencephaly associated with cerebellar underdevelopment [LCH; 1,9,11], in which an heterogeneity of clinical presentation and imaging findings have been reported [11,15,28,29], probably associated with



**Figure 7 Cerebellar neuronal heterotopias.** Groups of large neurons, consistent with Purkinje cells, appear in the cerebellar white matter. H-E.



**Figure 8 Pedigree of the lissencephaly and cerebellar hypoplasia affected Churra sheep.** The pattern is compatible with an autosomal recessive inheritance model. The pedigree relationships were confirmed using 19 microsatellite markers. Symbols used in the pedigree are those proposed in humans by the "Pedigree Standardization Work Group of the National Society of Genetic Counselors" [23]. Briefly, square represent male, circle female, shaded square-affected male, shaded circle-affected female. A diamond for the individual symbol is used to indicate animals with gender not specified (males and females). Numbers inside the symbol represent multiple individuals.

different gene mutations responsible for malformations in cerebral and cerebellar cortex [1,11]. Ross et al. [11] have proposed a classification of LCH based on clinical and morphological features and six types, so-called LCH groups a-f, have been recognized and some of them related to mutation in some genes. The cases recorded in this study, according to their pathological features, especially the presence of pachygyria, slight microencephaly, a marked thickening of the cerebral cortex, an abnormal hippocampus configuration and a normally developed corpus callosum, most closely resemble the LCH-b group. In humans, this phenotype has been related to mutations in the *RELN* gene that encodes reelin [10,11], a protein secreted by the first born neurons in neocortex and cerebellum [30]. Reelin is an extracellular matrix protein that plays a key role in the organization of architectonic pattern in cerebral and cerebellar cortices, particularly in the radial cortical organization [31] and controlling cell to cell interactions, critical for cell positioning in the brain [2]. Furthermore, lesions found in this study are in accordance with the phenotypic characteristics of the reeler mice, a mouse strain that have a mutation in the *RELN* gene causing a loss of reelin associated with an abnormal neurogenesis in the brain cortex [32]. In mice, mutations in genes that take part of the reelin signaling pathway like mutations in Disabled 1 (*Dab1*) and double mutations in two lipoprotein receptors, very-low-density lipoprotein receptor (*Vldlr*) and apolipoprotein E receptor (*ApoER2*)

generate a similar phenotype that is known as reeler-like phenotype [31,33-35]. More recently, a mutation in the *TUBA1A* gene, encoding for the alpha I tubulin protein, has been found in cases of LCH [15,36]. These mutations cause defective interactions in the tubulin heterodimer assembly or in the three-dimensional conformation that will impair the action of these proteins, necessary for central nervous system development [37,38]. Although mutations in this gene result in a variety of phenotypical presentations [15], hypoplasia or complete absence of the corpus callosum, has been reported together with lissencephaly and cerebellar hypoplasia, in the majority of the cases [15,29], including one reported in a cat [19] but was not detected in the lambs in this study. Further studies are necessary to investigate the possible gene mutations involved in this outbreak of lissencephaly in lambs.

Children suffering lissencephaly show mental and psychomotor retardation, seizures that are refractory to treatment and hypertonus in the limbs [7,8]. These clinical signs are broadly consistent with those observed in these lambs, which were unable to walk, had difficulty sucking their dams and show shivering and muscular rigidity.

The gross and, particularly, the histopathological findings of this study consisting of an abnormal layering of the cerebral and cerebellar cortices and the hippocampus, together with the presence of heterotopic neurons in the white matter, are in agreement with a neuronal migration defect during brain development [1,2,6,12]. In

classical lissencephaly in humans, four layers are recognized in most of the cases [2,8,36] whereas in LCH a wide variation has been reported, from patients in which no layers can be determined to other showing three or four layers [15,19,28]. In our study, the posterior area of the brain, the most severely affected macroscopically, showed markedly abnormal microscopic architecture. Whereas conventional methods identified two layers, the most external one corresponding to the molecular layer, by using immunohistochemical labelling for neurofilaments, three zones were distinguished. The widest layer was formed by positively immunolabelled neurons, consistent with the pyramidal layers seen in the control brains that also show a strong immunoreactivity. This picture suggests that the most external layer -the molecular area- was formed, and the subsequent migration of the neurons has failed. This finding is consistent with a reeler-like phenotype since the reelin signaling pathway plays a critical role in neuronal migration [39]. A complete and more detailed study of the lesions present in the different areas of the brain in these cases of lissencephaly with cerebellar hypoplasia and the analysis of the genetic basis of the phenotype is necessary to further characterize this entity, and is currently under progress.

## Conclusions

This study reports the clinical and pathological findings of several cases of lissencephaly with cerebellar hypoplasia in pedigree Churra lambs, in which evidence for an autosomal recessive mode of inheritance was demonstrated. The histopathological features observed in the cerebral cortex hippocampus and cerebellum are consistent with a failure in neuronal migration during brain development. Lissencephaly should be considered as a possible cause when newborn lambs show weakness and inability to walk or to suck. Further studies are needed to better characterize this disease in lambs. The possible use of lissencephalic lambs as model for studying human lissencephalies should be also considered.

## Methods

### Pathological examination

Thirty-three out of the 53 lambs examined were submitted alive. After intravenous injection of a veterinary euthanasia drug (T-61, Intervet), a complete necropsy was performed in all the lambs. All the organs were examined and gross lesions recorded. Coronal sections of the brain were made at different levels. Samples from several areas of the brain (cerebral frontal, parietal, occipital and temporal cortex, thalamus and corpus callosum, mesencephalon, hippocampus, pons and cerebellar peduncles, cerebellar cortex and medulla oblongata) and spinal cord together with other tissue samples (liver, lungs, heart, kidneys, thymus, lymph nodes and intestine with Peyer's

patches) were taken. The tissue samples were fixed in 10% neutral buffered formalin, dehydrated through graded alcohols and embedded in paraffin wax. Sections (4 µm) were stained with haematoxylin and eosin (HE) and cresyl violet method for identification of the neuron structure in the brain.

### Immunohistochemistry

Selected sections from the brain cortex were immunohistochemically labelled with a monoclonal antibody specific for the 70 kDa subunit of the NF-L human neurofilament (clone 2 F11, Dako, Golstrupp, Denmark), at a dilution 1:100, to further characterize brain morphology. Additionally, to rule out the teratogenic infectious agents present in Spain at the moment in which this study was carried out, in selected sections from the brain, thymus, mesenteric lymph nodes and Peyer's patches obtained from eight affected lambs, immunohistochemistry was applied, using monoclonal antibodies directed against ruminant pestivirus, known to react with border disease virus (PA0801, AHVLA, Addlestone, UK), and bluetongue virus (clone 2E9, Ingenasa, Madrid, Spain) diluted 1:100 and 1:20 respectively. In all the cases, a polymer-based detection system (EnVision + System Labelled Polymer-HRP antimouse; Dako, Golstrupp, Denmark) was employed, following the manufacturer instructions. Subsequently, immunolabelling was developed with a solution of 3,3'-diaminobenzidine (DAB) (Vector Laboratories, Burlingame, USA). The slides were counterstained with haematoxylin and mounted. Technique specificity was controlled by omission of the primary antibody and substitution by a normal mouse serum, omission of the Envision polymer and omission of diaminobenzidine. All these controls gave negative results. To evaluate the specificity of the antibodies, tissue samples of a normal lamb brain and several organs of a lamb and sheep previously known to be infected with border disease (where positive signal was found in neurons and lymphocytes) and bluetongue (positive labelling was found in macrophages and endothelial cells) viruses respectively, were employed.

### Pedigree data and analysis

To confirm the pedigree information from the affected lambs and their parents we collected blood samples in a 10 ml Venoject tubes with EDTA (Terumo Europe N.V., Leuven, Belgium) of a total of 33 animals from the lissencephaly affected flock. Fourteen of these animals were affected lambs (cases), three animals were unaffected half-sibs of the cases and 17 were healthy parents of the affected offspring, which were classified as obligate carriers. DNA was extracted from blood samples using the salting out procedure previously described [40]. All the samples were analysed for a set of 19 microsatellite markers amplified in a single PCR reaction



following the Multiplex-PCR protocol described by [41]. The PCR products were separated by capillary electrophoresis in an automatic ABI3130xl DNA Sequencer (Applied Biosystems). The results of the capillary electrophoresis were analyzed using GeneMapper v.4.0 software (Applied Biosystems). Finally, a segregation analysis was performed in order to estimate the genetic architecture of the trait. We followed the parsimony principle and, after examining the segregation in the pedigree, we proposed the compatible pattern that makes the fewest assumptions.

#### Abbreviations

LCH: Lissencephaly with cerebellar hypoplasia; CSF: Cerebrospinal fluid; H-W: Hardy-Weinberg.

#### Competing interests

Authors declare that they have no competing interests.

#### Authors' contributions

VP carried out the pathological examination, participated in the sample collection, interpreted the pathological and immunohistochemical analysis, and wrote the manuscript. ASV performed the pedigree data analysis and helped in the sample collection. MF collaborated in the sample collection and carried out the immunohistochemical analysis. JB participated in the pathological studies, interpretation of the results and helped to draft the manuscript. LD contributed to sample collection and to the pathological studies. MCF collaborated in the interpretation of pathological results and sample collection. JJA interpreted the pedigree data analysis and participated in writing the manuscript. All the authors read and approved the final manuscript.

#### Acknowledgements

The authors wish to thank the veterinary practitioners Sandra de la Hera and Jorge Gutiérrez, and the farmer of the flock, for their kind and enthusiastic collaboration. The technical help of Gloria Belver is also acknowledged. We are also grateful to Dr Sandra Scholes for reviewing the manuscript and valuable comments. A. Suárez-Vega is funded by a FPU contract from the Spanish Ministry of Education. J. Benavides is supported by CSIC through the JAE-Doc program, financed in part by European Social Fund (ESF). We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

#### Author details

<sup>1</sup>Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE), Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain. <sup>2</sup>Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain.

Received: 18 December 2012 Accepted: 8 August 2013

Published: 9 August 2013

#### References

- Ross ME, Walsh CA: Human brain malformations and their lessons for neuronal migration. *Annu Rev Neurosci* 2001, **24**:1041–1470.
- Guerrini R, Parrini E: Neuronal migration disorders. *Neurobiol Dis* 2010, **38**:154–166.
- Barkovich AJ, Guerrini R, Kuzniecky RI, Jackson GD, Dobyns WB: A developmental and genetic classification for malformations of cortical development: update 2012. *Brain* 2012, **135**:1348–1369.
- Norman MG, Ludwin SK: Congenital malformations of the nervous system. In *Textbook of Neuropathology*. 3rd edition. Edited by Davis RL, Robertson DM. Baltimore: Williams and Wilkins; 1997:265–330.
- Walsh CA: Genetic malformations of the human cerebral cortex. *Neuron* 1999, **23**:19–29.
- Pilz DT, Stoodley N, Golden JA: Neuronal migration, cerebral cortical development and cerebral cortical anomalies. *J Neuropathol Exp Neurol* 2002, **61**:1–11.
- Pilz DT, Quarrell WJ: Syndromes with lissencephaly. *J Med Genet* 1996, **33**:319–323.
- Kuchelmeister K, Bergmann M, Gullotta F: Neuropathology of lissencephalies. *Childs Nerv Syst* 1993, **9**:394–399.
- Kroon AA, Smit BJ, Barth PG, Hennekam RS: Lissencephaly with extreme cerebral and cerebellar hypoplasia. A magnetic resonance imaging study. *Neuropediatrics* 1996, **27**:273–276.
- Hong SE, Shugart YY, Huang DT, Shahwan SA, Grant PE, Hourihane JO, Martin NDT, Walsh CA: Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human *RELN* mutations. *Nat Genet* 2000, **26**:93–96.
- Ross ME, Swanson K, Dobyns WB: Lissencephaly with cerebellar hypoplasia (LCH): a heterogeneous group of cortical malformations. *Neuropediatrics* 2001, **32**:256–263.
- Gleeson JG, Walsh CA: Neuronal migration disorders: from genetic diseases to developmental mechanisms. *TINS* 2000, **23**:352–359.
- Reiner O, Albrecht U, Gordon M, Chianese KA, Wong C, Gal-Gerber O, Sapir T, Siracusa LD, Buchberg AM, Caskey CT, Eichele G: Lissencephaly gene (*LIS1*) expression in the CNS suggest a role in neuronal migration. *J Neurosci* 1995, **15**:3730–3738.
- Gleeson JG, Allen KM, Fox JW, Lamperti ED, Berkovic S, Scheffer I, Cooper EC, Dobyns WB, Minnerath SR, Ross ME, Walsh CA: Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein. *Cell* 1998, **92**:63–72.
- Fallet-Bianco C, Loeuillet L, Poirier K, Loget P, Chapon F, Pasquier L, Saillour Y, Beldjord C, Chelly J, Francis F: Neuropathological phenotype of a distinct form of lissencephaly associated with mutations in *TUBA1A*. *Brain* 2008, **131**:2304–2320.
- Zaki FA: Lissencephaly in Lhasa Apso dogs. *J Am Vet Med Assoc* 1976, **169**:1165–1168.
- Greene CE, Vandeveld M, Braund K: Lissencephaly in two Lhasa Apso dogs. *J Am Vet Med Assoc* 1976, **169**:405–410.
- Summers BA, Cummings JF, de Lahunta A: Malformations of the central nervous system. In *Veterinary Neuropathology*. Edited by Summers BA, Cummings JF, de Lahunta A. St Louis: Mosby; 1995:68–94.
- Herrmann A, Hecht W, Herden C: Lissencephaly and microencephaly combined with hypoplasia or corpus callosum and cerebellum in a domestic cat. *Tierarztl Prax Ausg K Kleintiere Heimtiere* 2011, **39**:116–120.
- Nettleton PF, Willoughby K: Border disease. In *Diseases of Sheep*. 4th edition. Edited by Aitken ID. Oxford: Blackwell; 2007:119–126.
- Edwards JF, Livingston CW, Chung SI, Collisson EC: Ovine arthrogryposis and central nervous system malformations associated with in utero Cache Valley virus infection: spontaneous disease. *Vet Pathol* 1989, **26**:33–39.
- Herder V, Wohlsein P, Peters M, Hansmann F, Baumgärtner W: Salient lesions in domestic ruminants infected with the emerging so-called Schmallenberg virus in Germany. *Vet Pathol* 2012, **49**:588–591.
- Bennett RL, French KS, Resta RG, Doyle DL: Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. *J Genet Couns* 2008, **17**:424–433.
- Kirkland PD: Akabane disease. In *Diseases of Sheep*. 4th edition. Edited by Aitken ID. Oxford: Blackwell; 2007:473–479.
- Maclachlan NJ, Drew CP, Darpel KE, Worwa G: The pathology and pathogenesis of bluetongue. *J Comp Pathol* 2009, **141**:1–16.
- Plant JW, Walker KH, Acland HM, Gard GP: Pathology in the ovine foetus caused by an ovine pestivirus. *Aus Vet J* 1983, **60**:137–140.
- García Marín JF, Pérez V, Gutiérrez MM, Corpa JM, Gómez N, Escudero A, del Río E: Lissencephaly and cerebellar hypoplasia in lambs. In *Proceedings of the 15th Meeting of the European Society of Veterinary Pathology*. Edited by Università degli Studi di Sassari. Italy: Sassari-Alghero; 1997:132.
- Kerner B, Graham JM, Golden JA, Pepkowitz SH, Dobyns WB: Familial lissencephaly with cleft palate and severe cerebellar hypoplasia. *Am J Med Genet* 1999, **87**:40–445.
- Miyata H, Chute DUJ, Fink J, Villablanca P, Vinters HV: Lissencephaly with agenesis of corpus callosum and rudimentary dysplastic cerebellum: a subtype of lissencephaly with cerebellar hypoplasia. *Acta Neuropathol* 2004, **107**:69–81.

30. D'Arcangelo G, Miao GG, Chen SC, Soares HD, Morgan JL, Curran T: A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* 1995, **374**:719–723.
31. Tissir F, Goffinet AM: Reelin and brain development. *Nat Rev Neurosci* 2003, **4**:496–505.
32. González JL, Russo CJ, Goldowitz D, Sweet HO, Davisson MT, Walsh CA: Birthdate and cell marker analysis of scrambler: a novel mutation affecting cortical development with a reeler-like phenotype. *J Neurosci* 1997, **17**:9204–9211.
33. Howell BW, Hawkes R, Soriano P, Cooper JA: Neuronal position in the developing brain is regulated by mouse disabled-1. *Nature* 1997, **389**:733–737.
34. Trommsdorff M, Gotthardt M, Hiesberger T, Shelton J, Stockinger W, Nimpf J, Hammer RE, Richardson JA, Herz J: Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. *Cell* 1999, **97**:689–701.
35. Ayala R, Shu T, Tsai LH: Trekking across the brain: the journey of neuronal migration. *Cell* 2007, **128**:29–43.
36. Lecourtois M, Poirier K, Friocourt G, Jaglin X, Goldenberg A, Saugier-Verber P, Chelly J, Laquerrière A: Human lissencephaly with cerebellar hypoplasia due to mutations in *TUBA1A*: expansion of the foetal neuropathological phenotype. *Acta Neuropathol* 2010, **119**:779–789.
37. Kumar RA, Pilz DT, Babatz TD, Cushion TD, Harvey K, Topf M, Yates L, Robb S, Uyanik G, Mancini GM, Rees MI, Harvey RJ, Dobyns WB: *TUBA1A* mutations cause wide spectrum lissencephaly (smooth brain) and suggest that multiple neuronal migration pathways converge on alpha tubulins. *Hum Mol Genet* 2010, **19**:2817–2827.
38. Friocourt G, Marcorelles P, Saugier-Verber P, Quille ML, Marret S, Laquerrière A: Role of cytoskeletal abnormalities in the neuropathology and pathophysiology of type I lissencephaly. *Acta Neuropathol* 2011, **121**:149–270.
39. Rice DS, Curran T: Role of the reelin signaling pathway in central nervous system development. *Annu Rev Neurosci* 2001, **24**:1005–1039.
40. Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988, **16**:1215.
41. Glowatzki-Mullis ML, Muntwyler J, Gaillard D: Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. *Anim Genet* 2007, **38**:86–88.

doi:10.1186/1746-6148-9-156

Cite this article as: Pérez et al.: Hereditary lissencephaly and cerebellar hypoplasia in Churra lambs. *BMC Veterinary Research* 2013 **9**:156.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit







# Veterinary Dermatology

## **Generalized severe junctional epidermolysis bullosa with congenital absence of skin in churra lambs.**

Julio Benavides<sup>1</sup>, Aroa Suárez-Vega<sup>2</sup>, Jorge Gutiérrez<sup>3</sup>, Sandra de la Hera<sup>3</sup>, Miguel Fuertes<sup>1</sup>, Laetitia Delgado<sup>1</sup>, Miguel Fernández<sup>1</sup>, María del Carmen Ferreras<sup>1</sup>, Juan José Arranz<sup>2</sup> and Valentín Pérez<sup>1</sup>.

<sup>1</sup>Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE), Facultad de Veterinaria, Universidad de León, León, Spain;

<sup>2</sup>Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, León, Spain; <sup>3</sup>Asociación nacional de criadores de ganado ovino selecto de raza churra (ANCHE), Palencia, Spain

Veterinary Dermatology. doi: 10.1111/vde.12226



## Generalized severe junctional epidermolysis bullosa with congenital absence of skin in churra lambs

Julio Benavides\*, Aroa Suárez-Vega†, Jorge Gutiérrez‡, Sandra de la Hera‡, Miguel Fuertes\*, Laetitia Delgado\*, Miguel Fernández\*, María del Carmen Ferreras\*, Juan José Arranz† and Valentín Pérez\*

\*Departamento de Sanidad Animal (Anatomía Patológica), Instituto de Ganadería de Montaña (CSIC-ULE),

†Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León, 24071, Spain

‡Asociación nacional de criadores de ganado ovino selecto de raza churra (ANCHE), Avd Casado del Alisal 21, Palencia, 34001, Spain

Correspondence: Valentín Pérez, Departamento de Sanidad Animal (Anatomía Patológica), Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, Spain. E-mail: vperp@unileon.es

**Background** – Up to 0.5% of churra lambs from two genetically related flocks showed congenital skin lesions of variable severity, jeopardizing the life of the lambs in the most severe cases.

**Hypothesis/Objectives** – The primary objective of this study was to classify the type of congenital epithelial disease suffered by these animals, based on the description of the macroscopic skin defects, the histological and ultrastructural changes and the hereditary nature of the condition.

**Animals** – Thirty affected newborn lambs from two genetically related flocks were studied. Three additional lambs acquired from two other flocks, which had no grossly apparent skin lesions and had died of infectious diseases, were studied as unaffected control animals.

**Methods** – Histological and ultrastructural examinations of skin and oral mucosa samples were performed. Pedigree analyses were used to investigate genealogical relationships.

**Results** – Generalized severe junctional epidermolysis bullosa with congenital absence of skin was described in all lambs studied and an autosomal recessive mode of inheritance was identified.

**Conclusions and clinical importance** – The pathological findings and mode of inheritance in these lambs are similar to an inherited epidermolysis bullosa subtype of humans, which has not been reported previously in veterinary medicine.

### Introduction

Epidermolysis bullosa (EB) refers to a heterogeneous group of skin disorders that are characterized by increased skin and mucosal fragility, with formation of vesicles and bullae following minor trauma. The disease results from structural defects in the basement membrane. The rupture of bullae leads to loss of the epithelium and ulceration.<sup>1,2</sup> Hereditary EB is classified into three main types based on the location of the basement membrane defect.<sup>3</sup> In EB simplex, the separation occurs within the basal keratinocyte layer, above the lamina lucida; in junctional EB (JEB), blister formation occurs in the lamina lucida; and in dystrophic EB, the separation occurs below the lamina densa. All three types of EB, with different subtypes, have been described in humans and genetically engineered mice, while only certain types have been reported in livestock or companion animals.<sup>2,4,5</sup> There have been several reports of this condition affecting different breeds of sheep in the UK, New Zealand, Switzerland, Spain and Germany.<sup>6–10</sup> Some cases have been characterized as dystrophic EB,<sup>7,10</sup> while others have been reported as JEB.<sup>5</sup>

A review of EB nomenclature has proposed a progressive method to classify subtypes of the disease. After classification into one of the major types (simplex, junctional, dystrophic or Kindler syndrome), it then accounts for clinical phenotypic features, such as distribution and cutaneous or extracutaneous involvement.<sup>11</sup> One of the less common cutaneous lesions that may accompany bullae is the congenital absence of skin. The association of EB with large areas of congenitally absent epithelium has not been reported in animals; it is rare in humans, where it has been described under the eponym of Bart syndrome.<sup>12</sup> It was subsequently classified as aplasia cutis congenita type 6,<sup>13</sup> where both lesions can coexist with other abnormalities, such as pyloric atresia, urethral stenosis, hypoplastic or 'bound-down' ears and nail dystrophy.<sup>13</sup> Congenital absence of skin has been described in human cases of both dystrophic and junctional EB.<sup>14,15</sup> Recent consensus reports have recommended that eponyms be avoided in favour of descriptive terms, i.e. Bart syndrome should now be referred to as 'EB with congenital absence of skin'.<sup>11,14</sup>

The aim of this study was to describe the macroscopic, histological and ultrastructural findings of skin lesions occurring in a group of purebred churra lambs. Given that these lambs were derived from genetically related flocks, an additional aim was to determine the mode of inheritance of the skin disease.

Accepted 8 April 2015

**Sources of Funding:** This study was self-funded.

**Conflict of Interest:** No conflicts of interest have been declared.

## Materials and methods

### Sheep

The animals described in this study were purebred churra lambs derived from two different flocks (VS and MS), which were managed intensively for milk production. The flocks were located in the same region, shared breeding males and amounted to ~800 sheep when combined. Over a 4 year period (2008–2011), between 0.2 and 0.5% of lambs were born with gross skin defects (auricular cartilage malformation and/or areas devoid of epidermis). Animals with the most severe lesions were euthanized within 24 h of parturition by intravenous overdose of barbiturates. Over the 4 years, 30 affected lambs (12 from flock VS and 18 from flock MS) were submitted for postmortem examination. Three lambs from two different pure-bred churra flocks, which had no grossly apparent skin lesions and had died of infectious diseases (neonatal diarrhoea and/or bacterial pneumonia), were studied as unaffected control animals.

### Pathological examination

Samples of skin were acquired from both the periphery and the centre of the lesional areas, which were mainly located on the dorsal aspect of carpal or tarsal regions, the tongue or the buccal mucosa. At least one sample from each of these three locations was collected from every lamb studied as well as from any other macroscopic lesions found during the postmortem examination. Specimens were fixed in 10% neutral buffered formalin, processed routinely through graded alcohols and embedded in paraffin wax. Sections were mounted on glass microscope slides and stained with haematoxylin and eosin, Gomori trichrome stain and periodic acid Schiff.

Samples of grossly normal skin from the metacarpal region and tongue were collected from four of the affected lambs and from one control animal and processed for electron microscopy. Specimens were fixed in 5% glutaraldehyde and then embedded in a mixture of Solution A [Epon 812 (Poly/Bed 812; Polysciences Inc., Warrington, PA, USA) and dodecylsuccinyl anhydride at a ratio of 1.3:1, respectively] and solution B [Epon 812 and methyl-5-norbornene-2,3-dicarboxylic anhydride at a ratio of 1.13:1, respectively, with 3% benzyldimethyl amine (Polysciences, Inc.) added. Ultrathin sections (95 nm) were stained with uranyl acetate and lead citrate (Polysciences, Inc.). A transmission electron microscope (JEM-1010; JEOL Ltd, Tokyo, Japan) was used for ultrastructural examination.

### Pedigree data and analysis

Information was obtained from the Spanish churra sheep breeders' association (Asociación nacional de criadores de ganado ovino selecto de raza churra (ANCHE)) herd book for the affected animals, and Pedigree software (University of Minnesota, St Paul, MN, USA) was used to construct the pedigree.<sup>16</sup> A total of 219 animals and 145 matings were used to evaluate the genetic basis of the disease in the affected flocks. After examining the segregation, we proposed the inheritance pattern of the trait following the parsimony principle.

## Results

### Gross findings

The most striking gross lesion in the affected lambs was the absence of epidermis in several locations (Figure 1a). These presented as well-demarcated foci of variable size (ranging from a few centimetres to large areas of the body) and shape (mainly ovoid). Areas devoid of epidermis were bright red in colour and no scarring was noted. Some lesions produced serofibrinous crusts, which were more frequent on the head and anterior limbs (Table 1). The head and ears were the most common location for macroscopic lesions, where they presented two different morphologies: an absence of epidermis, which was frequently observed over the dorsal aspect at the base of

the ear (Figure 1b), and fusion with folding inward of the borders of the pinnae at their apices (Figure 1c). In seven animals, the area devoid of epithelium extended to the fused misshapen ear tip (Figure 1c). In 65% of cases with head lesions the defects were bilateral and in eight cases the auricular skin defects were linked to lesions on the muzzle (Figure 1b).

Oral cavity lesions were frequent (57% of cases) in animals that had skin defects on the head (Table 1). When there were lesions in the oral cavity the tongue was consistently affected. Sixteen animals had erosions and ulcers on the dorsal aspect of the tongue, mainly located at the torus linguae, and in several cases they were present also on the apex (Figure 1d). In two of these animals the remaining epithelium of the dorsal aspect of the tongue was easily detached. In addition to the tongue, 12 of 16 lambs with affected oral cavities also had lesions on the buccal mucosa and hard palate. Although no gross lesions were visible in the oral cavities of the remaining four lambs, the epithelium – especially that of the hard palate – was friable and sloughed easily (Figure 1e).

All but one affected lamb showed skin defects on at least one of the four limbs; two limbs were commonly affected (Table 1). Lesions were present more frequently on the forelimbs and involved mainly the metacarpal region and the dorsal and palmar aspects of the phalanges (Figure 1f). On the forelimbs, the skin defect consistently extended proximally to the carpus or the antebrachium in the more severe cases (Figure 1f). Isolated patches of epidermis and wool were surrounded by hyperaemic dermis. Similar but less frequent lesions were found on the hindlimbs, where the phalanges were the most common location for lesions. Eight of the lambs with lesions on any limb showed incomplete or complete lack of development of at least one hoof, and in the other six lambs the hooves were partly detached at the time of postmortem examination. Lesions were also present on the thoracic (three lambs) and lumbar regions (seven lambs), where they appeared as large, red areas devoid of epithelium, extending onto the abdomen in two individuals. One lamb, which was devoid of epithelium on several areas of the head (ears and muzzle), limbs (anterior and posterior) and rump also suffered from atresia of the penile urethra. Oedema, congestion and accumulation of urine in the subcutaneous tissue of the ventral abdominal wall were present.

### Histopathology

The microscopic examination of all areas where epithelium was absent showed a sharply demarcated loss of epidermis, glands and hair follicles in the transition from normal to affected skin (Figure 2). The affected dermis was severely congested and trichrome-stained sections showed sparsely disarranged collagen fibres surrounded by oedema. A serofibrinous exudate with an infiltration of mixed inflammatory cells (neutrophils, macrophages, lymphocytes and plasma cells) associated with bacteria was present in a number of sections, especially those from the limbs. In some instances there was marked epidermal hyperplasia and orthokeratotic hyperkeratosis at the border of the lesions (Figure 2).



**Figure 1.** Macroscopic lesions in affected lambs. (a) Affected lamb showing multiple skin lesions of the ears, limbs and caudal half of the body. (b) Epidermal defects resulting in dermal exposure of the ears, muzzle and occipital area of the head. The transition between affected and unaffected areas was very well demarcated. (c) Affected lamb showing two different types of macroscopic lesions on the ear: lack of epithelium on the pinna and malformation of the apex. (d) Multifocal to coalescent ulceration of the dorsal surface of the tongue. The torus linguae was most severely affected. (e) Hard palate: single ulcer and sloughing of the epithelium. (f) Limbs: marked congestion of the dermis, partly covered by a fibrinonecrotic exudate. Note the detachment of the hoof in the most severely affected limb.

Specimens from all lambs showed subepithelial splitting of variable severity in at least one of the samples examined. The roof of the microscopic vesicle was formed by full-thickness epidermis, with all the cell layers, including intact basal cells. In areas with greater separation between the epidermis and dermis, vesicles were filled with clear, eosinophilic fluid and erythrocytes. Detachment of hair follicles from the dermis was also a common feature. Periodic acid Schiff-stained sections showed that the basement membrane was located on the floor of the cleft (Figure 3). Whether macroscopic

lesions were present, or not, subepithelial splitting was apparent in the oral mucosa and the skin.

#### Transmission electron microscopy

Subepithelial cleft formation was present in all skin and tongue samples from affected animals examined by transmission electron microscopy. The lamina densa was found at the base of the cleft adjacent to the dermis, while no evidence of lamina lucida was seen (Figure 4). While desmosomes were normal, there was an evident reduction in the number of hemidesmosomes and sub-



**Table 1.** Location of macroscopic skin/mucosal defects in the studied animals

| Region         | n (%*)  | Location       | n (%†)  |       |        |
|----------------|---------|----------------|---------|-------|--------|
| Head           | 28 (93) | Tongue         | 16 (57) |       |        |
|                |         | Gingiva        | 1 (4)   |       |        |
|                |         | Buccal mucosa  | 2 (7)   |       |        |
|                |         | Palate         | 12 (43) |       |        |
|                |         | Muzzle         | 8 (29)  |       |        |
|                |         | Ear, right     | 24 (86) |       |        |
|                |         | Ear, left      | 21 (70) |       |        |
|                |         | Orbital area   | 4 (13)  |       |        |
|                |         | Occipital area | 2 (7)   |       |        |
|                |         | Thorax         | 3 (11)  | Neck  | 1 (33) |
|                |         |                |         | Other | 2 (67) |
| Abdomen        | 7 (23)  | Rump           | 5 (71)  |       |        |
|                |         | Other          | 2 (29)  |       |        |
| Right forelimb | 25 (83) | Humerus        | 1 (3)   |       |        |
|                |         | Radius         | 0 (0)   |       |        |
|                |         | Metacarpus     | 13 (52) |       |        |
|                |         | Phalanges      | 18 (72) |       |        |
| Left forelimb  | 22 (73) | Hoof           | 10 (40) |       |        |
|                |         | Humerus        | 1 (5)   |       |        |
|                |         | Radius         | 1 (5)   |       |        |
|                |         | Metacarpus     | 10 (45) |       |        |
| Right hindlimb | 13 (43) | Phalanges      | 17 (77) |       |        |
|                |         | Hoof           | 10 (45) |       |        |
|                |         | Femur          | 1 (8)   |       |        |
|                |         | Tibia          | 2 (15)  |       |        |
| Left hindlimb  | 11 (37) | Metatarsus     | 5 (38)  |       |        |
|                |         | Phalanges      | 7 (54)  |       |        |
|                |         | Hoof           | 8 (62)  |       |        |
|                |         | Femur          | 1 (9)   |       |        |
| Pelvis         | 1 (3)   | Tibia          | 1 (9)   |       |        |
|                |         | Metatarsus     | 4 (36)  |       |        |
|                |         | Phalanges      | 7 (64)  |       |        |
|                |         | Hoof           | 7 (64)  |       |        |
|                |         | Urethra        | 1 (100) |       |        |

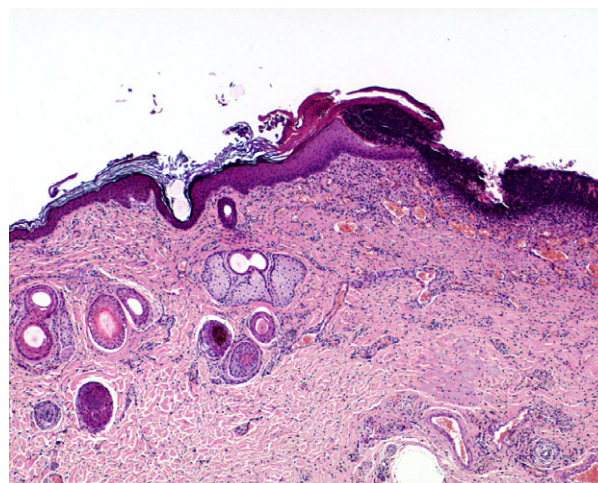
\*Percentage of total studied animals showing gross lesions in this region.

†Percentage of animals showing lesions in this location.

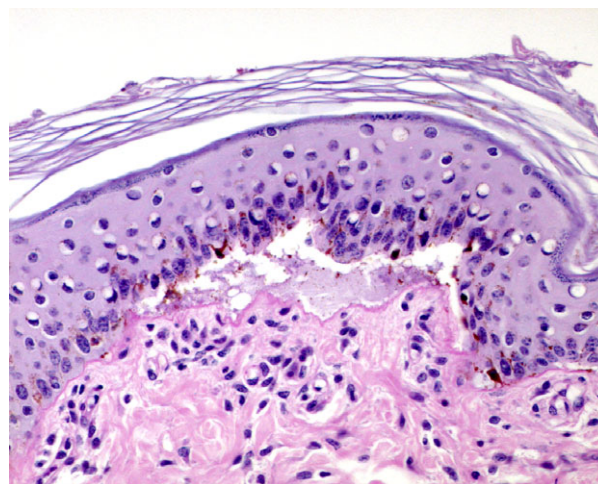
basal dense plaques in the affected animals (Figure 5a). Collagen fibres of affected areas were scarce and disarranged (Figure 5a). The ultrastructure of the skin from the control lambs was normal (Figure 5b).

### Pedigree analysis

The breeding history of the affected flocks was consistent with a monogenic autosomal recessive pattern of inheritance such that the affected individuals must be homozygous for the causal mutation (aa) and the parents of affected animals must be classified as carriers (Aa), as seen in Supporting information Figure S1. Analysis of pedigrees revealed that the breeding practices in the flocks were based on mating closely related individuals. A detailed analysis of pedigrees showed that all affected animals could be traced, on both paternal and maternal sides, to a common ancestor designated as 'VS94.11'. The relationship of VS94.11 with affected individuals from the 'VS' flock was extensive (Figure 6a). A ram (designated MSVS03085) descended from VS94.11 was introduced into the breeding stock of the MS flock. Inbreedings between the offspring of this ram were subsequently responsible for the development of EB in that flock (Figure 6b).



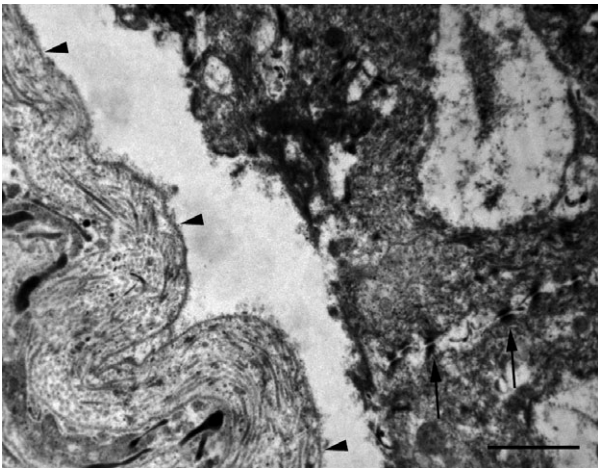
**Figure 2.** Histopathological findings in affected skin. Congenital absence of skin denoted by a sharp transition between unaffected and affected (right) areas where the epidermis and cutaneous adnexa are absent. Fibrinopurulent dermatitis involving the exposed dermis, and focal epidermal hyperplasia and orthokeratotic hyperkeratosis at the border of the lesion. Haematoxylin and eosin stain.



**Figure 3.** Histopathological findings in affected skin. Subepidermal cleft formation beneath the basal cells and serofibrinous exudate in the blister. Note staining of the basement membrane on the floor of the bulla. Periodic acid Schiff stain.

### Discussion

The histopathological changes identified in the cases reported here, which were characterized by subepithelial splitting of the epithelium of both oral mucosae and skin, are consistent with a diagnosis of EB.<sup>2,10</sup> Ultrastructural evaluation of these lesions enabled further classification of the condition as JEB due to the presence of dermo-epithelial clefts within the lamina lucida of the basement membrane.<sup>1,2,5</sup> The gross skin lesions identified (i.e. extensive areas devoid of epidermis or oral mucosal epithelium) merit classification of this inherited disease as 'generalized severe JEB with congenital absence of skin'.<sup>11</sup> Previous descriptions of this condition in human medicine have referred to the congenital absence of skin as 'aplasia cutis'.<sup>13</sup>

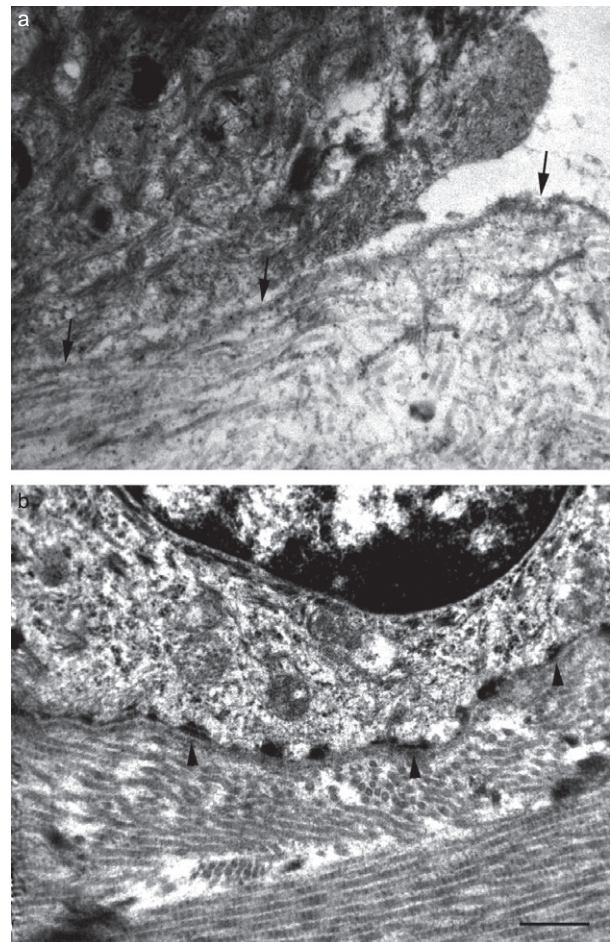


**Figure 4.** Ultrastructural findings in skin with junctional epidermolysis bullosa. Cleft formation between the basal cell and the lamina densa (arrowheads) located at the base of the bulla. Note the normal desmosomes between basal cells (arrows). Electron microscopy  $\times 15,000$  magnification. Scale bar represents  $1 \mu\text{m}$ .

The concurrence of EB and congenital absence of skin is rare in humans and, to the best of the authors' knowledge, there are no confirmed reports in domestic animals. However, in veterinary medicine the term 'epitheliogenesis imperfecta' has been used to describe cases of inherited EB with macroscopic absence of epidermis,<sup>17</sup> and this has resulted in confusion.<sup>18</sup> The nature of the lesions (i.e. absence of epithelium) in the our cases and their predilection for the head and extremities are similar to those previously described in lambs and calves.<sup>18–20</sup> It is possible that prior reports were describing the same condition, but the lack of detailed histological descriptions of grossly nonaffected skin in previously reported cases prevents the retrospective confirmation of whether JEB was present or not.

While the pathogenesis of hereditary EB is well known (i.e. structural defects of basement membrane proteins involved in anchoring the epidermis to the dermis), the pathogenesis of epidermal absence is not.<sup>1</sup> In human medicine, the occurrence of congenital absence of epidermis – as a unique lesion referred to as aplasia cutis congenita – is typically sporadic, with an estimated incidence of three in 10,000 births.<sup>21</sup> Very few cases have been described as inherited and several theories have been proposed to explain sporadic cases, including intra-uterine trauma, vascular compromise, infectious agents and medications.<sup>21</sup> Conversely, congenital absence of epidermis linked to EB is thought to be a clinical manifestation of the EB rather than a separate entity, probably as a consequence of *in utero* rupture of bullae and trauma.<sup>22,23</sup> Like the lambs described in this study, areas of missing epidermis in human newborns often include the legs, arms and head,<sup>22–25</sup> suggesting a shared pathogenic mechanism between the two species.

In addition to loss of the epidermis, the affected skin of lambs reported here lacked the dermal adnexae (hair follicles and glands). It has been shown that expression of the transcriptional factor p63 in the epidermis during embryogenesis is essential for skin appendage development.<sup>26</sup> Therefore, loss of epidermis at this stage due to

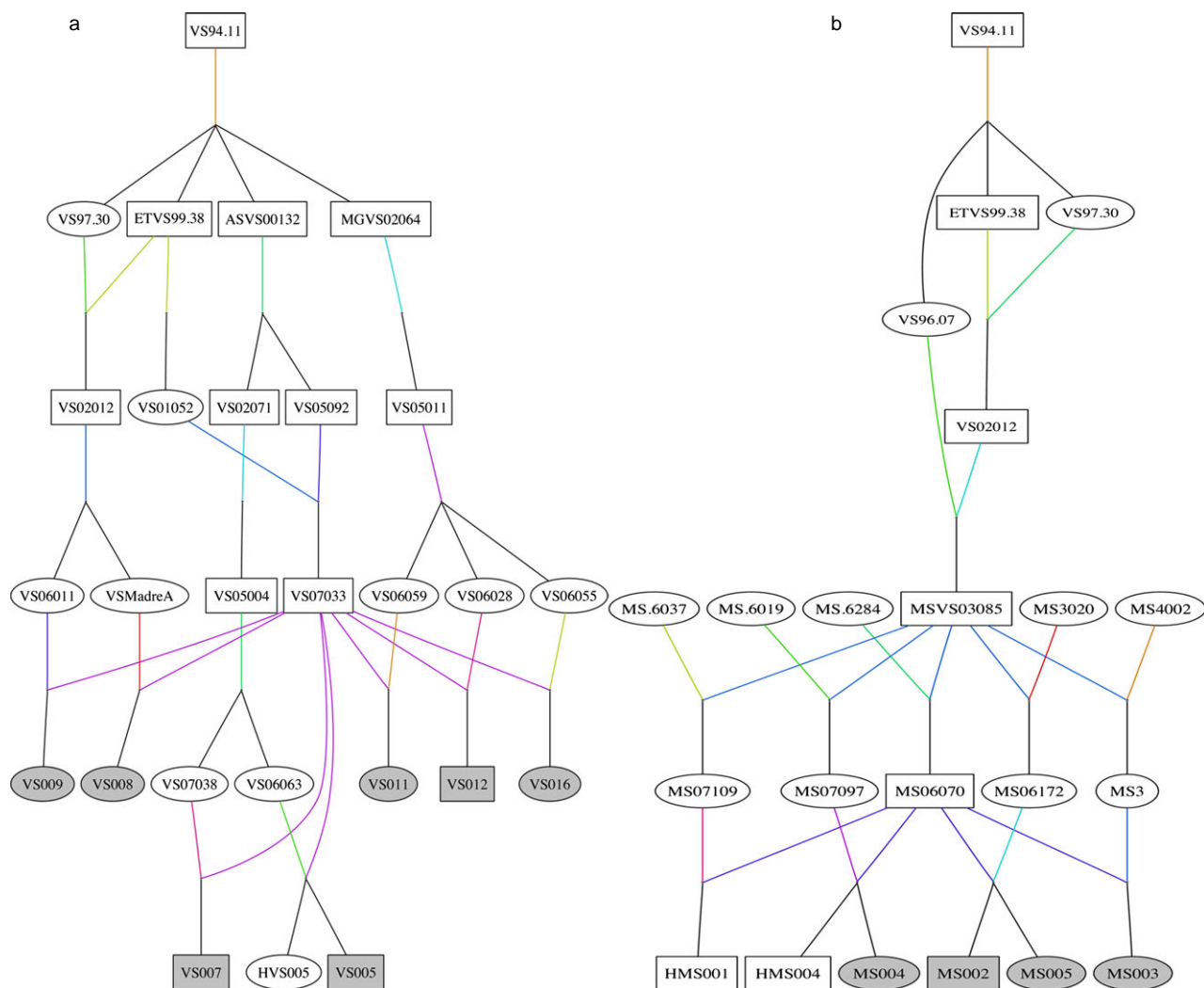


**Figure 5.** Ultrastructural findings in skin with junctional epidermolysis bullosa. (a) Junctional epidermolysis bullosa in an affected lamb: absence of hemidesmosomes. Note the scarcity and disarrangement of collagen fibres in the dermis, below the lamina densa (arrows). (b) Skin of control lamb. Normal skin with hemidesmosomes (arrowheads). Electron microscopy  $\times 30,000$  magnification. Scale bar represents  $0.5 \mu\text{m}$ .

*in utero* formation of bullae and subsequent sloughing of the epithelium may deprive the skin of the stimulus and cell repertoire necessary for adnexal development. Another possibility is that the *in utero* epithelial detachment could evolve into a scarring process involving the subjacent dermis, thus interfering with normal adnexal development. In our cases trichrome stain showed that scar tissue was histologically absent. However, mammalian fetuses can heal cutaneous lesions without the formation of scar tissue,<sup>27</sup> so this hypothesis cannot be addressed in the circumstances of our study. In any case, the close relationship between the congenital absence of epithelium and adnexae with JEB, found together in all studied lambs, would fit with the proposed hypothesis that the congenital absence of skin is a clinical manifestation of EB that occurs *in utero*.

Loss of epithelium has been reported to occur with all three major types of EB in people<sup>28</sup> but is most commonly related to dominant dystrophic EB in Bart syndrome<sup>29</sup> (recently renamed 'EB with congenital absence of skin').<sup>11</sup> The three major types of EB have also been described in veterinary medicine and, more specifically,





**Figure 6.** Pedigree analysis of the animals affected by epidermolysis bullosa. Filled symbols represent affected animals in the two flocks. Squares represent males; ovals represent females. Animal identification is indicated within the symbols. The occurrence of the affected lambs is consistent with an autosomal recessive trait; all can be traced back on both paternal and maternal sides to a common ancestor 'VS94.11' for both the VS flock (a) and the MS flock (b).

JEB has been observed in horses, cattle and sheep. Although congenital absence of skin has not been reported in any of these cases,<sup>2</sup> the causative mutation of JEB, identified in horses, sheep and dogs, affects the genes *LAMC2* or *LAMA3*, which are responsible for expression of laminin 332, a protein of the hemidesmosomes.<sup>2</sup> The fact that every lamb from the present study showed congenital absence of skin linked to JEB suggests that this syndrome may have been caused by single or multiple mutations affecting genes different from *LAMC2* or *LAMA3*.

The specific association between congenital absence of skin and JEB of humans commonly co-occurs with pyloric atresia (PA) and other extracutaneous abnormalities, such as genitourinary malformations or rudimentary ears.<sup>11,23</sup> Known as 'EB with pyloric atresia' (EB-PA),<sup>14,30</sup> this disease is caused by mutations in the hemidesmosomes,<sup>31</sup> specifically in genes encoding for  $\alpha 6\beta 4$  integrin.<sup>32,33</sup> While PA was not noted in the lambs analysed in this study, incomplete PA could be hard to detect during postmortem examination of a neonate. Malformation

of the ears was a common feature in affected lambs and one animal also exhibited urethral stenosis. Therefore, future studies could target  $\alpha 6\beta 4$  integrin genes as reasonable candidates for the causative mutation in these flocks.

The genealogical analysis of 241 ancestors of affected lambs allowed the inference that the mode of inheritance of the disease follows a pattern consistent with an autosomal recessive allele. Pedigree analysis showed that inbreeding was a common practice in these flocks and that systematic use of carrier rams produced a founder-like effect in subsequent generations. This was evident with the predominance of ram VS80.76 in the population's pedigree. A similar event was observed several generations later with ram VS94.11, which was the most recent common ancestor of animals affected by EB. If either of these rams was a carrier of the mutation responsible for the disease, the probability of finding homozygous animals with the mutation (identity-by-descent) would be highly favoured by inbreeding. Likewise, we can track the introduction of EB into the MS flock follow-



ing the introduction of ram MSVS03085, a direct descendant of VS94.11 and probably a heterozygous carrier of the mutated allele.

In conclusion, the macroscopic, histological and ultrastructural features reported here suggest that the condition suffered by these lambs should be classified as 'generalized severe JEB with congenital absence of skin' according to recent guidelines.<sup>11</sup> Comparisons with prior descriptions of JEB in humans and animals suggest that the genetic cause of this condition may be different from those previously described in animals. Further studies are required to determine the precise protein(s) and genetic mutation(s) responsible for the lesions.

## Acknowledgements

The authors are grateful to the farmers of the two flocks for their generous collaboration and to Mark Dagleish, Moredun Research Institute for proofreading and constructive criticism of the manuscript.

## References

- Hargis AM, Ginn PE. The integument. In: McGavin MD, Zachary JF, eds. *Pathologic Basis of Veterinary Disease*. 4th edition. St Louis: Mosby Elsevier 2007; 1252–1253.
- Medeiros GX, Riet-Correa F. Epidermolysis bullosa in animals: a review. *Vet Dermatol* 2015; 26: 3–13, e1–e2.
- Sawamura D, Nakano H, Matsuzaki Y. Overview of epidermolysis bullosa. *J Dermatol* 2010; 37: 214–219.
- Bruckner-Tuderman L, McGrath JA, Robinson EC *et al*. Animal models of epidermolysis bullosa: update 2010. *J Invest Dermatol* 2010; 130: 1485–1488.
- Ostmeier M, Kerkmann A, Frase R *et al*. Inherited junctional epidermolysis bullosa (Herlitz type) in German black-headed mutton sheep. *J Comp Pathol* 2012; 146: 338–347.
- Alley MR, O'Hara PJ, Middelberg A. An epidermolysis bullosa of sheep. *N Z Vet J* 1974; 22: 55–59.
- Bruckner-Tuderman L, Guscetti F, Ehrensperger F. Animal model for dermolytic mechanobullous disease: sheep with recessive dystrophic epidermolysis bullosa lack collagen VII. *J Invest Dermatol* 1991; 96: 452–458.
- Henderson DC. Foot lameness. In: Henderson DC, ed. *The Veterinary Book for Sheep Farmers*. Ipswich: Farming Press. 1990; 587–616.
- McTaggart HS, Ritchie JS, Copland AN. Red foot disease of lambs. *Vet Rec* 1974; 94: 153–159.
- Perez V, Benavides J, Delgado L *et al*. Dystrophic epidermolysis bullosa in Assaf lambs. *J Comp Pathol* 2011; 145: 226–230.
- Fine JD, Bruckner-Tuderman L, Eady RA *et al*. Inherited epidermolysis bullosa: updated recommendations on diagnosis and classification. *J Am Acad Dermatol* 2014; 70: 1103–1126.
- Bart BJ. Epidermolysis bullosa and congenital localized absence of skin. *Arch Dermatol* 1970; 101: 78–81.
- Frieden IJ. Aplasia cutis congenita: a clinical review and proposal for classification. *J Am Acad Dermatol* 1986; 14: 646–660.
- Fine JD, Eady RA, Bauer EA *et al*. The classification of inherited epidermolysis bullosa (EB): report of the Third International Consensus Meeting on Diagnosis and Classification of EB. *J Am Acad Dermatol* 2008; 58: 931–950.
- Rosmaninho A, Machado S, Selores M. Bart syndrome. *Int J Dermatol* 2014; 53: e54–e55.
- Garbe JR, Da Y. Pedigree: a software tool for the graphical visualization of large and complex populations. *Yi Chuan Xue Bao* 2003; 30: 1193–1195.
- Benoit-Biancamano MO, Drolet R, D'Allaire S. Aplasia cutis congenita (epitheliogenesis imperfecta) in swine: observations from a large breeding herd. *J Vet Diagn Invest* 2006; 18: 573–579.
- Maestrale C, Masia M, Marruchella G *et al*. Pathology in practice. Epitheliogenesis imperfecta in a lamb. *J Am Vet Med Assoc* 2013; 242: 179–181.
- Leipold HW, Mills JH, Huston K. Epitheliogenesis imperfecta in Holstein-Friesian calves. *Hastings Vet J* 1973; 14: 114–118.
- Munday BL. Epitheliogenesis imperfecta in lambs and kittens. *Br Vet J* 1970; 126: 47.
- Browning JC. Aplasia cutis congenita: approach to evaluation and management. *Dermatol Ther* 2013; 26: 439–444.
- Dang N, Klingberg S, Rubin AI *et al*. Differential expression of pyloric atresia in junctional epidermolysis bullosa with *ITGB4* mutations suggests that pyloric atresia is due to factors other than the mutations and not predictive of a poor outcome: three novel mutations and a review of the literature. *Acta Derm Venereol* 2008; 88: 438–448.
- Puvabanditsin S, Garrow E, Samransamruajit R *et al*. Epidermolysis bullosa associated with congenital localized absence of skin, fetal abdominal mass and pyloric atresia. *Pediatr Dermatol* 1997; 14: 359–362.
- Zelickson B, Matsumura K, Kist D *et al*. Bart's syndrome. Ultrastructure and genetic linkage. *Arch Dermatol* 1995; 131:663–668.
- Maman E, Maor E, Kachko L *et al*. Epidermolysis bullosa, pyloric atresia, aplasia cutis congenita: histopathological delineation of an autosomal recessive disease. *Am J Med Genet* 1998; 78:127–133.
- Mikkola ML. p63 in skin appendage development. *Cell Cycle* 2007; 6 285–290.
- Gawronska-Kozak B, Grabowska A, Kopcewicz M *et al*. Animal models of skin regeneration. *Reprod Biol* 2014; 14: 61–67.
- Chiaverini C, Charlesworth A, Fernandez A *et al*. Aplasia cutis congenita with dystrophic epidermolysis bullosa: clinical and mutational study. *Br J Dermatol* 2014; 170: 901–906.
- Casanova JM, Martí RM, Baradad M *et al*. Bart syndrome associated to lethal junctional epidermolysis bullosa (Herlitz form). *Actas Dermosifiliogr* 2006; 97: 658–661.
- Wallerstein R, Klein ML, Genieser N *et al*. Epidermolysis bullosa, pyloric atresia, and obstructive uropathy: a report of two case reports with molecular correlation and clinical management. *Pediatr Dermatol* 2000; 17: 286–289.
- Pulkkinen L, Uitto J. Hemidesmosomal variants of epidermolysis bullosa. Mutations in the alpha6beta4 integrin and the 180-kD bullous pemphigoid antigen/type XVII collagen genes. *Exp Dermatol* 1998; 7: 46–64.
- Brown TA, Gil SG, Sybert VP *et al*. Defective integrin alpha 6 beta 4 expression in the skin of patients with junctional epidermolysis bullosa and pyloric atresia. *J Invest Dermatol* 1996; 107: 384–391.
- Gil SG, Brown TA, Ryan MC *et al*. Junctional epidermolysis bullosis: defects in expression of epiligrin/nicein/kalinin and integrin beta 4 that inhibit hemidesmosome formation. *J Invest Dermatol* 1994; 5 (Suppl): 31S–38S.

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Familial relationships between the 18 animals affected by junctional epidermolysis bullosa and congenital absence of skin. Filled symbols represent affected lambs. Occurrence is consistent with an autosomal recessive trait. All cases can be traced to the common ancestor 'VS94.11' for both the VS flock and the MS flock.

## Résumé

**Contexte** – Jusqu'à 0.5% des agneaux de race churra de deux troupeaux génétiquement apparentés ont montré des lésions cutanées congénitales de sévérité variable, mettant en danger la vie des agneaux dans les cas les plus graves.

**Hypothèses/Objectifs** – Le premier objectif de cette étude était de déterminer le type de la maladie congénitale épithéliale de ces animaux en fonction des lésions macroscopiques cutanées, des modifications histologiques et ultrastructurales et de la nature héréditaire de la maladie.

**Sujets** – Trente agneaux nouveaux nés atteints de deux troupeaux génétiquement apparentés ont été étudiés. Trois agneaux supplémentaires issus de deux autres troupeaux, sans lésion cutanée apparente, morts de maladie infectieuse, ont été inclus dans l'étude en tant que contrôles.

**Méthodes** – Les examens histologiques et ultrastructuraux de la peau et de la muqueuse orale ont réalisés. Les analyses de pédigrées ont permis d'étudier les relations généalogiques.

**Résultats** – Une épidermolyse bulleuse jonctionnelle généralisée sévère avec absence congénitale de peau a été décrite chez tous les agneaux étudiés et un mode de transmission autosomal récessif a été identifié.

**Conclusions et importance clinique** – Les données pathologiques et le mode de transmission chez ces agneaux sont identiques à un sous type d'épidermolyse bulleuse de l'homme, ce qui n'avait pas encore été décrit en médecine vétérinaire.

## Resumen

**Introducción** – hasta un 0,5% de los corderos de raza Churra de dos rebaños genéticamente relacionados mostraron defectos congénitos de la piel de severidad variable, poniendo en peligro la vida de los corderos en los casos más severos

**Hipótesis/Objetivos** – el objetivo primario de este estudio fue clasificar el tipo de enfermedad epitelial congénita sufrido por estos animales basado en la descripción de los defectos de la piel macroscópicos, los cambios histológicos y ultraestructurales, y la naturaleza hereditaria de la condición.

**Animales** – 30 corderos recién nacidos afectados de dos rebaños genéticamente relacionados fueron objeto del estudio. Se adquirieron como animales control no afectados tres corderos adicionales de otros dos diferentes rebaños, sin que presentaran lesiones de la piel aparentes y que habían muerto de enfermedades infecciosas.

**Métodos** – se desarrolló el examen histológico y ultraestructural de la piel y la mucosa oral. El análisis de pedigree se utilizó para investigar las relaciones genealógicas

**Resultados** – se describe epidermolisis bullosa de la unión generalizada severa con ausencia congénita de la piel en todos los corderos estudiados y un modo de herencia recesivo autosómico.

**Conclusiones e importancia clínica** – los hallazgos patológicos y el modo de herencia en estos corderos son similares a un subtipo de epidermis bullosa hereditaria en humanos, enfermedad que no ha sido previamente descrita en medicina veterinaria.

## Zusammenfassung

**Hintergrund** – Bis zu 0,5% von Churra Lämmern aus zwei genetisch verwandten Herden zeigten angeborene Hauterkrankungen von unterschiedlichem Schweregrad, was bei den schwersten Fällen das Leben der Lämmer gefährdete.

**Hypothese/Ziele** – Das Primärziel dieser Studie war eine Klassifizierung der Arten der angeborenen epithelialen Erkrankungen, an denen diese Tiere litten. Das erfolgte basierend auf der Beschreibung der makroskopischen Hautdefekte, der histologischen und ultrastrukturellen Veränderungen und der angeborenen Natur dieses Zustandes.

**Ergebnisse** – Es wurde eine generalisierte schwere junctionale Epidermolysis bullosa mit einem angeborenen Fehlen der Haut bei allen untersuchten Lämmern beschrieben und ein autosomal rezessiver Vererbungsmodus identifiziert.

**Schlussfolgerungen und klinische Bedeutung** – Die pathologischen Befunde und die Art der Vererbung bei diesen Lämmern war ähnlich wie ein vererbter Subtyp der Epidermolysis bullosa des Menschen, der in der Veterinärmedizin bisher noch nicht beschrieben worden war.

## 摘要

**背景** – 两群基因学相关的楚拉羊中,大于0.5%羊羔表现出不同程度的先天性皮肤病变,大多数严重病例会危及生命

**假设/目的** – 主要目的是将患病动物的先天性上皮疾病类型进行分类,分类基于对病变的宏观描述、组织学、超微结构的变化和遗传特点。

**动物** – 两群基因学相关的新生羊羔参与实验,其中有30只患病羊羔。另外3只从其他两群中获得,没有明显皮肤病变,但死于传染病,作为无病对照组。

**方法** – 对皮肤样本和口腔黏膜样本进行组织病理学和超微检查。用系谱分析评估基因学关系。

**結果** — 该实验的所有羊羔的先天性皮肤病变,呈现全身性严重的交界处大疱性表皮松懈症,伴有先天性皮肤缺失,可确认为常染色体隐形遗传。

**总结与临床意义** — 这些羊羔的病理学结果和遗传方式,与人类遗传性大疱性表皮松懈症亚型相同,之前从未在兽医学中被报道过。

#### 要約

**背景** — 遺伝的に関連がある2つの群から構成される0.5%のチュルラ羊は様々な重症度で、もっとも重症な症例では羊の命が脅かされる先天性の皮膚病変を示す。

**仮説/目的** — この研究の主な目的は肉眼的な皮膚異常、組織学、微細構造的変化ならびに遺伝の状態に関する記述に基づいて、これらの動物が罹患した先天性上皮疾患を分類すること。

**供与動物** — 2つの遺伝的に関連した群の31頭の罹患した新生羊で調査した。肉眼的な皮膚病変がなく、感染症で死亡した2つの群から入手した3頭の追加の羊を罹患していないコントロール動物として調査した。

**方法** — 皮膚および口腔粘膜材料の組織学および超微細構造学的な検査を行った。系統関係を調べるために血統解析を用いた。

**結果** — 調査したすべての羊から皮膚に先天的欠損を示す全身に重度の接合型表皮水泡症が説明され、常染色体劣性遺伝様式が特定された。

**結論および臨床的な重要性** — これらの羊の病理学的な所見および遺伝形式はヒトの遺伝性の表皮水泡症の亜型と類似しており、獣医学では今まで報告されていない。

## Supporting Information

**Figure S1.** Familial relationships between the 18 animals affected by junctional epidermolysis bullosa and congenital absence of skin. Filled symbols represent affected lambs. Occurrence is consistent with an autosomal recessive trait. All cases can be traced to the common ancestor 'VS94.11' for both the VS flock and the MS flock.

