Insertion of an endogenous Jaagsiekte Sheep Retrovirus element into the *BCO2* - gene abolish its function and leads to yellow discoloration of adipose tissue in Norwegian Spælsau (*Ovis aries*).

Matthew Peter Kent[1*], Michel Moser[1*], Inger Anne Boman[2], Kristine Lindtveit[1], Mariann Árnyasi[1], Kristil Kindem Sundsaasen[1], Dag Inge Våge[1#]

[1]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway.
[2]The Norwegian Association of Sheep and Goat Breeders, PO Box 104, NO-1431 Ås, Norway.

*These authors contributed equally to this work.
Corresponding author.

Abstract

In sheep, the accumulation of carotenoids in adipose tissue leading to yellow fat is a heritable, recessive trait, that can be attributed to a nonsense mutation in the *beta-carotene oxygenase 2 (BCO2)* gene. However, the fact that other sheep breeds affected by yellow fat do not suffer from the reported nonsense mutation suggests that other functional mechanisms must exist. We investigated one such breed, the Norwegian spælsau, and detected an aberration in *BCO2* mRNA. Nanopore sequencing of genomic DNA revealed the insertion of a 7,9 kb endogenous Jaagsiekte Sheep Retrovirus (enJSRV) sequence in the first intron of the *BCO2* gene. Closer examination of cDNA revealed that the *BCO2* first exon was spliced directly into position 415 of the enJSRV-sequence, immediately downstream of a potential -AG splice acceptor site. The hybrid protein product consists of 29 amino acids coded by the *BCO2* exon 1, followed by 29 arbitrary amino acid coded by the enJSRV-sequence, before a translation stop codon is reached. Considering that the functional BCO2 protein consists of 575 amino acids, it is unlikely that the 59 amino acid long BCO2/enJSRV hybrid protein can display any enzyme function. The found null-allele of BCO2 presents an alternative functional mechanism for the inactivity of BCO2 and is a perfect example for the potential of screening for structural variants using long-read sequencing data.

Introduction

Yellow coloration of adipose tissue in sheep is known to be a heritable trait in different sheep breeds, caused by accumulation of carotenoids (Hill 1962; Crane & Clare 1975; Kirton *et al.* 1975; Baker *et al.* 1985). Mainly due to consumer preferences, yellow fat is considered as a quality failure in sheep meat production and therefore not wanted.

In animals, two enzymes displaying different cleavage effects on carotenoids have been identified; *BCO1* (*beta-carotene oxygenase 1*) symmetrically cleaves β-carotene into two retinal molecules, while *BCO2* (*beta-carotene oxygenase 2*) asymmetrically cleaves β-carotene into β-apo-10'-carotenal (C27) and β-ionone (C13) (von Lintig & Vogt 2000; Wyss *et al.* 2000; Kiefer *et al.* 2001; Redmond *et al.* 2001).

Naturally occurring mutations influencing either gene expression or protein coding sequence have been reported for both genes in several organisms, changes in *BCO2* especially seem to result in tissue accumulation (Eriksson *et al.* 2008; Berry *et al.* 2009). The reason for this might be ascribed to the broader substrate specificity for BCO2 (Amengual *et al.* 2011; Mein *et al.* 2011).

We have earlier reported that a homozygous nonsense mutation in the coding sequence of ovine *BCO2* causes yellow fat in Norwegian White Sheep breed (Våge & Boman 2010). However, in the native Norwegian breed Spælsau, the yellow fat trait is found to segregate in the absence of this mutation, suggesting that alternative functional mechanism is influencing this trait. In the present study we searched for novel *BCO2* mutations in the Spælsau breed that could explain the yellow fat phenotype.

Material and methods

Animals

Fat colour in sheep carcasses is graded as “normal” or “yellow” in Norwegian abattoirs and recorded at the individual level within the national Sheep Recording System. For initial analysis we obtained access to 4 carcasses at slaughter from a farm with a history yellow fat lambs (sample set 1). Liver samples (approximately 0.125 cm3) from a Spælsau ram carcass (individual 70346), showing yellow fat, were collected and stored in RNAlater™ (QIAGEN, Hilden, Germany). Additional samples (two males; 50289,70203, and one female; 20025) from carcasses without visible yellow fat were also sampled. These three individuals could potential be carriers of the genetic variant causing yellow fat, based on their relationship to known “yellow fat animals”, but most likely not homozygous for this variant.

To validate findings, a larger sample set consisting 26 lambs were sampled from two separate farms in Norway (sample set 2). These lambs constituted two half-sib groups, descending from two rams known to produce yellow fat offspring. The mothers of these lambs had also an increased probability of carrying at least one yellow fat allele, based on relationship to known carriers.

RNA extraction and cDNA synthesis

RNA was extracted using the RNeasy®Plus Universal Mini Kit from QIAGEN according to the manufacturer's instructions. The concentration and quality of the RNA was measured using a NanoDrop 8000 (Thermo Scientific), and the integrity of RNA was measured using a Bioanalyzer 2100 (Agilent). All samples had an RNA integrity number value (RIN) of at least 8.1. cDNA was produced using a SuperScript™ II Reverse Transcriptase kit from Invitrogen according to manufacturer's instructions.

PCR amplification of BCO2 cDNA

Four pairs of previously tested primers (Våge & Boman 2010) were used to amplify 4 overlapping regions of the BCO2 cDNA (Våge & Boman 2010). The thermocycler program included 10 min at 95°C, and 40 cycles of 95°C for 30 sec, 57°C for 30 sec and 72°C for 1.5 min. AmpliTaq Gold® Polymerase (Applied Biosystems) was used.

Nanopore sequencing

High-molecular-weight DNA from a white fat (20025) and yellow fat (70346) individual was extracted from liver tissue using Genomic-tips (G/100) from Qiagen. The Short Read Eliminator kit from Circulomics was used to reduce the presence of fragments smaller than 25Kb. Libraries were prepared using the 1D Native barcoding DNA Nanopore Protocol (LSK109, Oxford Nanopore), with native barcoding (EXP-NBD104, Oxford Nanopore). Equal masses of both libaries were combined and 300ng was loaded into a single PromethION flowcell (FLO-PRO002). After 23hours the run was stopped (36Gb raw data) and the flow-cell regenerated using Oxford Nanopores nuclease flush protocol, 230ng of the remaining pooled library was added to the same flow-cell which was run for an additional 22.2 hours (22Gb raw data).

Base-calling and quality filtering

Nanopore reads were base-called using Guppy, version 3.0.3 (Oxford Nanopore, MinKNOW v.19.05.1) using the 'High accuracy' flip-flop model. Reads from the two sequencing runs were merged and demultiplexed with QCAT (https://github.com/nanoporetech/qcat).

Demultiplexed reads from each sample were trimmed for 50 bp from the start and quality-filtered for average PHRED quality > 7 and minimum length of 4000 bp with fastp (v0.19.5) options " –disable trim_poly_g --disable_adapter_trimming -q 7 -l 4000 -f 50 " (Chen *et al.* 2018).

Structural variant (SV) detection

Quality-filtered reads were mapped to the sheep reference genome (RefSeq accession GCF_002742125.1) using minimap2 (v2.16-r922) (Li 2018). Structural variants were called with the SV-caller SVIM (v0.5.0) using default parameters (Heller & Vingron 2019). Called SVs within 100 Kbp of the candidate gene *BCO2* (NM_001159278.1) were inspected manually.

As a complementary approach and to spot larger rearrangements between the sheep reference genome and the yellow fat individual, quality-filtered reads of sheep_70346 were assembled using the software Flye (v2.4.2) with options "--genome-size 2.8g -m 10000" (Kolmogorov *et al.* 2019). Contigs from the sheep_70346 genome assembly were mapped to the sheep reference genome using minimap2 (v2.16-r922) (Li 2018). Candidate contigs spanning *BCO2*-region were aligned and dot-plotted against the sheep reference genome using Gepard (v1.40) for visual inspection (Krumsiek *et al.* 2007).

Identification of intronic insertion in BCO2

Candidate SVs within the *BCO2* gene region were filtered for presence only in sheep_70346 or following expected allele pattern among the two individuals (homozygous in sheep_70346 and heterozygous in sheep_20025). On-line blastn (Coordinators 2013) was used to identify the inserted sequence in *BCO2* intronic region (INS_chr15_25022547). Insertion boundaries were determined by aligning contigs of sheep_70346 genome assembly against the Oar_rambouillet_v1.0 reference genome.

Detection of BCO2 exon 1 and BCO2- JSRV hybrid cDNA in the yellow fat animal

A 112 bp exon 1 fragment of *BCO2* from the yellow fat individual were amplified using primers 7550/7551 (Table 1). A hybrid fragment consisting of sheep *BCO2* exon 1 sequence and JSRV sequence were amplified from the same individual using primers 7550/7555. In both cases cDNA was used as template with the following thermo cycler program: 10 min at 95°C and 40 cycles at 95°C for 30 sec, 57°C for 30 sec and 72°C for 30 sec using AmpliTaq Gold®

Polymerase (Applied Biosystems). The 7550-7555 product was DNA sequenced by Sanger technology at Eurofins Scientific, Luxembourg, using the same two primers for the sequencing reaction.

PCR to detect the presence or absence of the endogenous virus sequence

To detect the presence or absence of the endogenous virus sequence, three pairs of primers were designed (Table 1): one pair (7552/7553) to amplify the insertion point without any insertion, one pair (7554/7555) to amplify the junction between the *BCO2* intron sequence and the upstream virus region and one (7556/7557) to amplify the junction between the *BCO2* intron sequence and the downstream virus region. Expected fragments lengths are 221 bp, 684 bp and 768 bp, respectively. DNA was denatured for 10 min at 95°C and PCR run for 40 cycles at 95°C for 30 sec, 58°C for 30 sec and 72°C for 1.5 min using using AmpliTaq Gold® Polymerase (Applied Biosystems).

Results/Discussion

Our initial experiments to understand the mechanism(s) responsible for yellow fat included performing PCR on cDNA from normal and affected animals to detect the expression of *BCO2*. Four primer pairs amplifying *BCO2* produced 4 fragments of expected size in the three individuals showing no yellow fat, while no fragments were amplified from the individual with yellow fat (Figure 1). Our provisional conclusion was that the yellow fat phenotype was likely due to downregulated expression of *BCO2* mRNA, but results did not offer insight into the mechanism underlying this.

To investigate if novel genomic rearrangements in the *BCO2* region could account for the lack of detectable mRNA, genomic DNA from the single yellow fat individual (70346) and one additional individual without the yellow fat phenotype (20025) were sequenced using nanopore technology. Over the course of two runs (using one flowcell) the PromethION yielded a total of 6908462 reads summing to approximatey 73 Gb raw data. Barcode adapters were detected in 78% of reads (n = 5,387,322) and facilitated the demultiplexing and assignment of 21.05 Gbp

(1330100 reads) to sample 20025, and 27.16 Gbp (2512544 reads) to sample 70346. After quality-filtering (Q>7 and Length > 4000 bp), a total of 20.2 Gbases (7 x coverage) and 25.8 Gbases (9 x coverage) with median and N50 read lengths of 14,792 and 23,141 bp, and 10,567 and 13,570 bp for 20025 and 70346 remained for alignment and assembly.

Initially, all nanopore reads were mapped to the Oar_rambouillet_v1.0 reference and examined for structural variants. Within the genomic sequence of *BCO2* (location NC_040266.1: 25021687-25091194), five structural variants were detected. Four of them, found in both individuals, may represent breed specific differences and were discarded as candidate SVs explaining the yellow fat phenotype. The single remaining SV consisted of a 7,938 bp insertion (Figure 2) at reference genome position NC_040266.1 : 25,022,547 which is 730 bp downstream from the end of *BCO2*'s first exon. The inserted sequence showed high similarity (99.04 % identity) to an endogenous Jaagsiekte sheep retrovirus (accession no. MF175071.1). The enJSRV insertion was verified with PCR using primer pairs 7554-7555 and 7556-7557 in the two sequenced animals (20025 and 70346), which gave products (of expected size) spanning the upstream and downstream sheep-virus junctions, respectively. Under the PCR conditions we used, the primer pair spanning the insertions site 7552-7553 did not amplify any product in the yellow fat animal, while it did in the heterozygous animal.

After realising that the enJSRV-sequence was inserted between exon 1 and exon 2 in the *BCO2* gene, we tested if exon 1 (located upstream of the insertion point) could still be expressed by amplifying a subregion of exon 1 (primers 7550-7551) from cDNA. A band of expected size (112 bp) was produced, indicating that *BCO2* was being transcribed individual 70346. The next step was then to test the hypothesis that the enJSRV - sequence could have interrupted the splicing process, since using PCR primers located downstream of the insertion point did not yield any products from cDNA (see Figure 1 for details). An initial test was done, using the only enJSRV reverse primer at hand (7555) and combined this primer with the exon 1 forward primer (7550). This primer pair yielded a product of 149 bp from cDNA of the yellow fat individual.

Sequencing this 149 bp fragment revealed that where the sheep *BCO2* exon 1 ended, the cDNA continued directly with enJSRV sequence, from position 415 according to MF175071.1-sequence (Figure 2). This shows that the wild-type ovine BCO2 protein product is completely disrupted after amino acid position 29 by being first chimerized with 30 amino acids coded by

the enJSRV-sequence and then terminated by a stop codon (accession no. MT024238). Although untested, knowing that the full-length protein consists of 575 amino acids, we theorize that it is highly unlikely that these 59 peptides can function as a BCO2 enzyme in animals homozygous for the enJSRV insertion.

The preliminary conclusion drawn from a very small number of individuals was strengthened when we included the analysis of sample set 2, consisting of two half-sib families with an expected high frequency of the yellow fat allele. Out of 26 genotyped lambs, 10 were heterozygous for the insertion while 4 was homozygous for the insertion. The remaining 12 lambs did not carry the insertion. In the abattoir-report the 4 homozygous animals were reported to have yellow fat, while the remaining group had white fat (Table 2).

We think this paper clearly illustrate the strength of using long-read sequencing technology to search for structural variation and it also exemplifies how such variation can directly influence gene function.

Acknowledgements

This project has received financial support from the Norwegian Association of Sheep and Goat Breeders (NSG). We are grateful to the sheep breeder and the personnel at the Nortura abattoir that made this study possible. We also wish to thank Øystein W Milvang for genotyping sample set 2.

Authors' contributions

IAB conceived the study together with DIV and MPK, identified suitable animals to sample and coordinated the sample collection. KL purified RNA/DNA and did the PCR-based experiments. MPK designed the Nanopore sequencing experiment and contributed in the overall planning of the project. MA did the Nanopore sequencing together with MPK. MM analysed the Nanopore sequencing data. KSS planned and coordinated the labwork. DIV assembled the results and made a first draft of the paper. All authors discussed and interpreted results and contributed to writing the paper. All authors have read and approved the final manuscript.

Ethics declaration

The animals described herein are not to be considered as experimental animals as defined in EU directive 2010/63 and in the Regulation of Animal Experiments in Norway. Consequently, we did not seek ethical review or approval of this study regarding the use of experimental animals. Tissue samples used for mRNA isolation were collected from ordinary production animals in a Norwegian abattoir after the animals were dead (sample set 1). Tissue samples used for DNA isolation in the two half-sib groups were collected as part of the routine ear-tagging used on commercial farms in Norway (sample set 2).

Competing interests

The authors declare no competing interests.

Data Availability

Raw long-range sequencing data have been submitted to the European Nucleotide Archive under accession number ERR3522460 and ERR3522461. The hybrid mRNA sequence is available under accession number MT024238. Other data are available upon reasonable requests.

References

Amengual J., Lobo G.P., Golczak M., Li H.N., Klimova T., Hoppel C.L., Wyss A., Palczewski K. & von Lintig J. (2011) A mitochondrial enzyme degrades carotenoids and protects against oxidative stress. *FASEB J* **25**, 948-59.

Baker R.L., Steine T., Vabeno A.W. & Breines D. (1985) The inheritance and incidence of yellow fat in Norwegian sheep. *Acta Agriculturae Scandinavica* **35**, 389-97.

Berry S.D., Davis S.R., Beattie E.M., Thomas N.L., Burrett A.K., Ward H.E., Stanfield A.M., Biswas M., Ankersmit A.E., Oxley P.E., Barnett J.L., Pearson J.F., van der Does Y., Macgibbon A.H., Spelman R.J., Lehnert K. & Snell R.G. (2009) Mutation in Bovine Beta-Carotene Oxygenase 2 Affects Milk Color. *Genetics* **182**, 923-6.

Chen S., Zhou Y., Chen Y. & Gu J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i90.

Coordinators N.R. (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **41**, D8-D20.

Crane B. & Clare N.T. (1975) Nature of carotenoid pigments in yellow fat of sheep. *New Zealand Journal of Agricultural Research* **18**, 273-5.

Eriksson J., Larson G., Gunnarsson U., Bed'hom B., Tixier-Boichard M., Stromstedt L., Wright D., Jungerius A., Vereijken A., Randi E., Jensen P. & Andersson L. (2008) Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken. *Plos Genetics* **4**.

Heller D. & Vingron M. (2019) SVIM: Structural Variant Identification using Mapped Long Reads. *Bioinformatics*.

Hill F. (1962) Xanthophyll pigmentation in sheep fat. *Nature* **194**, 865-6.

Kiefer C., Hessel S., Lampert J.M., Vogt K., Lederer M.O., Breithaupt D.E. & von Lintig J. (2001) Identification and characterization of a mammalian enzyme catalyzing the asymmetric oxidative cleavage of provitamin A. *Journal of Biological Chemistry* **276**, 14110-6.

Kirton A.H., Crane B., Paterson D.J. & Clare N.T. (1975) Yellow fat in lambs caused by carotenoid pigmentation. *New Zealand Journal of Agricultural Research* **18**, 267-72.

Kolmogorov M., Yuan J., Lin Y. & Pevzner P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540-6.

Krumsiek J., Arnold R. & Rattei T. (2007) Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**, 1026-8.

Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-100.

Mein J.R., Dolnikowski G.G., Ernst H., Russell R.M. & Wang X.D. (2011) Enzymatic formation of apo-carotenoids from the xanthophyll carotenoids lutein, zeaxanthin and beta-cryptoxanthin by ferret carotene-9',10'-monooxygenase. *Arch Biochem Biophys* **506**, 109-21.

Redmond T.M., Gentleman S., Duncan T., Yu S., Wiggert B., Gantt E. & Cunningham F.X. (2001) Identification, expression, and substrate specificity of a mammalian beta-carotene 15,15 '-dioxygenase. *Journal of Biological Chemistry* **276**, 6560-5.

von Lintig J. & Vogt K. (2000) Filling the gap in vitamin A research - Molecular identification of an enzyme cleaving beta-carotene to retinal. *Journal of Biological Chemistry* **275**, 11915-20.

Våge D.I. & Boman I.A. (2010) A nonsense mutation in the beta-carotene oxygenase 2 (BCO2) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (Ovis aries). *BMC Genet* **11**, 10.

Wyss A., Wirtz G., Woggon W.D., Brugger R., Wyss M., Friedlein A., Bachmann H. & Hunziker W. (2000) Cloning and expression of beta,beta-carotene 15,15 '-dioxygenase. *Biochemical and Biophysical Research Communications* **271**, 334-6.

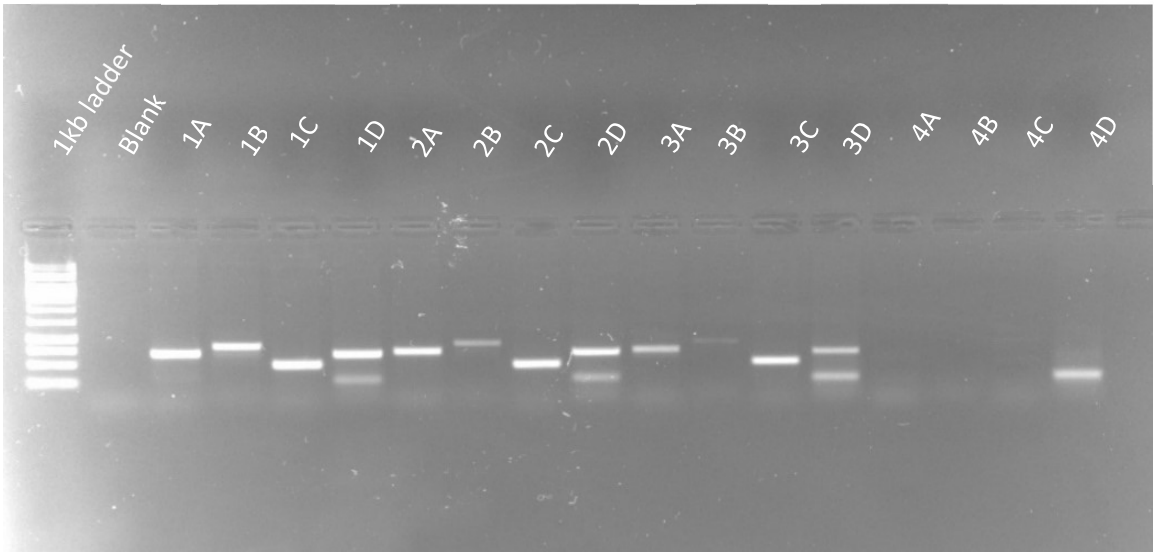

Figure 1

Figure is showing 4 different sets of primers (A: 7541-7542, B: 7543-7544, C: 7545-7546, D: 7547-7548) amplifying different subregions of the ovine BCO2 cDNA sequence in 4 different individuals (20025, 50289, 70203 and 70346, respectively). No fragments are amplified in the yellow fat individual (70346), the band in 4D is an artefact which also is visible in the 3 other individuals (in addition to the band of expected size found in the other 3).

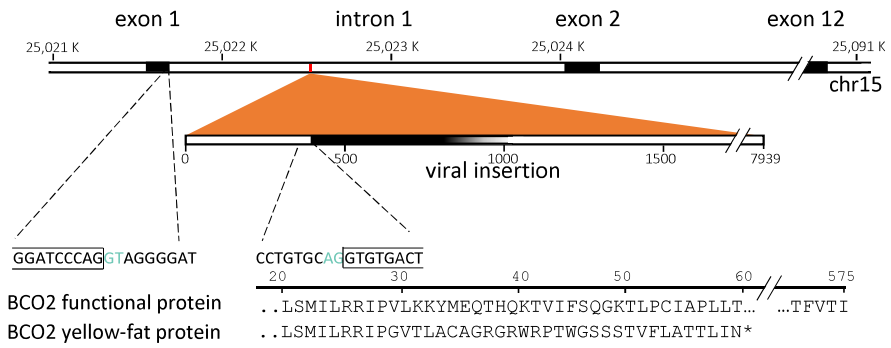

Figure 2. Diagram of the viral insertion mutation in BCO2 gene (NC_040266.1:25021687-25091194) of the Spaelsau breed

Table 1

Primers used in addition to those described in Våge & Boman 2010.

Primer	Sequence 5’ - 3’	Position	Direction	NCBI reference seq
7550	CTGCTGCTGCAGAACTCAAC	18-37	F	XM_012095240.3
7551	GGGATCCGGCGTAATATCAT	110-129	R	XM_012095240.3
7552	AATCCCATGGACGGAGAAG	25022430-25022448	F	NC_040266.1
7553	CTCCCTTTAAGAGGCAGCAT	25022631-25022650	R	NC_040266.1
7554	CAGGAAACCTGGGTTCGAT	25022347-25022365	F	NC_040266.1
7555	GGCGAGGAAAACTGTCGAG	470-488	R	MF175071.1
7556	CACCAACATCTTATGGAGCTTTT	7454-7476	F	MF175071.1
7557	GGGTTCTTTGCCATTAGCAC	25022801-25022820	R	NC_040266.1

Table 2: Fat colour grading and genotypes of 26 lambs descending from two different rams (76245736 and 89012427) known to produce yellow fat offspring. The ewes also have an increased probability of carrying the yellow fat allele, due to relationship to known carriers. Genotypes are scored as wild type (wt) or carrier of the endogenous Jaagsiekte Sheep Retrovirus (ins) based on the PCR-test described in materials and methods.

Lamb ID	**Ram**	**Ewe**	**Sex**	**Phenotype**	**Genotype**
90027	76245736	37200184	male	white fat	wt/ins
90029	76245736	37200184	male	white fat	wt/wt
90041	76245736	89740216	male	white fat	wt/ins
90059	76245736	37198929	female	white fat	wt/ins
90060	76245736	37198929	male	white fat	wt/wt
90072	76245736	89740213	female	white fat	wt/ins
90073	76245736	89740213	female	white fat	wt/wt
90001	89012427	82424293	female	**yellow fat**	**ins/ins**
90002	89012427	82424293	female	**yellow fat**	**ins/ins**
90015	89012427	76245888	female	white fat	wt/ins
90019	89012427	88844575	female	**yellow fat**	**ins/ins**
90075	89012427	76382061	female	white fat	wt/ins
90077	89012427	76382061	female	white fat	wt/wt
90083	89012427	37199669	female	white fat	wt/wt
90104	89012427	70456589	female	white fat	wt/ins
90115	89012427	82991541	female	white fat	wt/wt
90201	89012427	88842546	male	white fat	wt/wt
90202	89012427	88842546	male	white fat	wt/wt
90214	89012427	76245888	male	white fat	wt/wt
90228	89012427	76245885	male	white fat	wt/ins
90230	89012427	76245885	male	white fat	wt/wt
90297	89012427	89400436	male	**yellow fat**	**ins/ins**
90298	89012427	89400436	male	white fat	wt/wt
90307	89012427	37199669	male	white fat	wt/ins
90330	89012427	70456589	male	white fat	wt/wt
90336	89012427	82991541	male	white fat	wt/ins