SPECIAL ISSUE: AAABG2023 | RESEARCH PAPER
https://doi.org/10.1071/AN22463
ANIMAL PRODUCTION SCIENCE

# A copy number variant near *KITLG* is associated with the roan pattern in alpacas

Ishani Shah[A], Naomi Gray[A], David Groth[A], Samantha Brooks[B] and Kylie Munyard[A,*]

For full list of author affiliations and declarations see end of paper

***Correspondence to:**
Kylie Munyard
Curtin Medical School and Curtin Health Innovation Research Institute, Faculty of Health Sciences, Curtin University, Perth, WA, Australia
Email: K.Munyard@curtin.edu.au

**Handling Editor:**
Sue Hatcher

## ABSTRACT

**Context.** The alpaca roan pattern is characterised by white and coloured fibre interspersed together, with a distinctive lighter body and darker extremities, and commonly is believed to be inherited in an autosomal dominant manner. It is of interest to the alpaca fibre industry as it causes 'contamination' of coloured fibre with white fibres, but cannot be detected in white or light fawn animals. Other livestock species, such as horses, cattle, goats, and pigs, exhibit comparable phenotypes, which are associated with candidate variant(s) in either *KIT* or *KITLG*. **Aims.** To identify a region or regions of the genome that is/are causative of the roan pattern in alpacas. **Methods.** We conducted a genome-wide association study (GWAS) by using 13 roan and 14 non-roan alpacas sampled from the USA, Australia, and New Zealand. Regions of genome-wide significance were examined for variants that correlated with the roan phenotype. **Key results.** A novel candidate single-nucleotype polymorphism (SNP; Super-Scaffold_15:39 742 851T > A), located 272 kb upstream of *KITLG*, was identified in 1 of 12 regions with genome-wide significant association ($P \leq 5 \times 10^{-8}$). We identified the candidate SNP-containing region (Super-Scaffold_15:39 742 096–39 887 419) to be a 145 kb copy number variant (CNV) that is likely to be a tandem duplication. All 13 roan alpacas had one or two copies of the roan-associated *T* allele and all except three non-roans had zero copies. Furthermore, we determined the Mendelian inheritance of copy number haplotypes and their allelic composition in a roan and a non-roan family. **Conclusions.** Our data support the hypothesised autosomal incomplete dominant mode of inheritance of the roan pattern in alpacas and suggests that the effect of the *T* allele CNV version is likely to be suppressed when in *cis* with the *A* allele CNV version. However, additional verification is required to validate the finding and determine the functional effect. **Implications.** Identification of the cause, or a marker for roan pattern will allow alpaca breeders to select for or against the roan pattern, even when the phenotype is hidden, and therefore increase production output and profitability.

**Keywords:** alpaca, CNV, colour, fibre, genotyping by sequencing, GWAS, *KITLG*, pattern, roan, SNP.

**Received:** 15 December 2022
**Accepted:** 13 March 2023
**Published:** 4 April 2023

**Cite this:**
Shah I *et al.* (2023)
*Animal Production Science*, **63**(10–11), 1008–1016.
doi:10.1071/AN22463



## Introduction

Coat colours and patterns in livestock species have been artificially selected for thousands of years, giving rise to numerous colour phenotypes. Alpacas (*Vicugna pacos*) produce over 20 colours of base fibre and exhibit a range of patterns, such as classic grey, roan (aka modern grey in the USA and UK), blue-eyed white, tuxedo, piebald and appaloosa (Munyard 2013; Thangavel *et al.* 2015). The large variety of high-quality, naturally coloured fibre is gaining popularity within the 'green', environmentally conscious market and so many alpaca breeders are selecting for coloured fibres instead of the traditional white (Thangavel *et al.* 2015). For most domesticated species, selective breeding practices have significantly improved since the development and commercialisation of genetic screening (Toro 2010). Similarly, the Alpaca Coat Colour DNA test (Neogen Australasia) was developed to report the so-far known genotypes for the three loss-of-function *MC1R* alleles (no black pigment able to be produced) and three loss-of-function *ASIP* genes

(only black pigment able to be produced; Feeley and Munyard 2009; Feeley *et al.* 2011; Munyard 2013), and the presence or absence of the classic grey pattern (Jones *et al.* 2019), the only colours with known genetic causes.

The roan coat pattern in alpacas presents as depigmented fibres interspersed within the base colour, generally being concentrated to the central body and with darker extremities (Fig. 1). In the fibre industry, roan fibre is considered a contaminant to coloured fibre and undesirable for commercial processing, whereas hobby spinners and knitters often prefer such variety (Mathews *et al.* 2019). Genetic testing is required for the early and accurate detection of roan animals as the pattern is highly variable in its expressivity, distribution, and age of onset. Roan commonly appears at birth or early in life (Munyard 2013) and develops over time, whereas late-onset depigmentation is classed as age-related greying. However, in practice, distinguishing between the two traits is difficult due to the phenotypic variability of roan (Goldleaf and Windella alpacas, Dr C. Oddie pers. comm.). Detecting the pattern on white or light fawn animals is very difficult, and these so-called cryptic roans are typically responsible for the hidden dissemination of the trait within herds.

A similar autosomal dominant (or incomplete dominant) roan phenotype has also been described in other species. Genetic studies in horses (Marklund *et al.* 1999; Grilz-Seger *et al.* 2020; Voß *et al.* 2020), cattle (Charlier *et al.* 1996; Seitz *et al.* 1999), pigs (Cho *et al.* 2011; Lim *et al.* 2011; Fontanesi and Russo 2013), and goats (Talenti *et al.* 2018) have shown a shared association between roan and the candidate genes *KIT* proto-oncogene, receptor tyrosine kinase (*KIT*) or KIT ligand (*KITLG*). The biological cause of roan is predicted to be the premature depletion of melanocyte stem-cells (melanoblasts) within hair follicles, leading to gradual and permanent depigmentation of individual fibres (Hachiya *et al.* 2009; Endou *et al.* 2014; Bian *et al.* 2019; Qiu *et al.* 2019). KIT and KITLG are essential signalling proteins involved in migration, differentiation, maturation, and survival of melanoblasts (Besmer *et al.* 1993; Wehrle-Haller 2003). Apart from melanogenesis, KIT and KITLG are also vital for gametogenesis (Besmer *et al.* 1993) and haematopoiesis (Zsebo *et al.* 1990). Thus, loss-of-function mutations in either gene are associated with patterns of depigmentation (Hemesath *et al.* 1998; Cieslak *et al.* 2011) as well as pleiotropic effects such as subfertility in homozygous roan cattle (Charlier *et al.* 1996) and homozygous lethality in some roan horses (Hintz and Van Vleck 1979).

Although mammalian pigmentation is highly conserved, genetic heterogeneity exists for many interspecies colour phenotypes (Cieslak *et al.* 2011; Brancalion *et al.* 2022). Over 688 pigmentation-associated genes are known (Baxter *et al.* 2019). We hypothesised that an autosomal incompletely dominant regulatory variant in or around a gene involved in pigmentation was responsible for the roan coat pattern in alpacas. The aim was to identify a candidate gene and associated variant(s) through use of a genome-wide association study (GWAS), and thereby improve our knowledge of the genetic basis of the roan coat pattern in alpacas.

## Materials and methods

### Samples

Whole-genome sequence data (WGS) were generated from blood samples of 46 white, fawn, bay, black, white spotted, classic grey, blue eyed white and roan (modern grey) alpacas. Animals were a mix of closely and distantly related individuals, and were sampled from Australian, New Zealand and American herds. Sequences were generated using Illumina Novaseq 150 bp paired-end reads with approximately 30× depth-of-coverage per sample. Phenotype data, in the form of photographs, and pedigree information were available for 35 of the sampled animals.

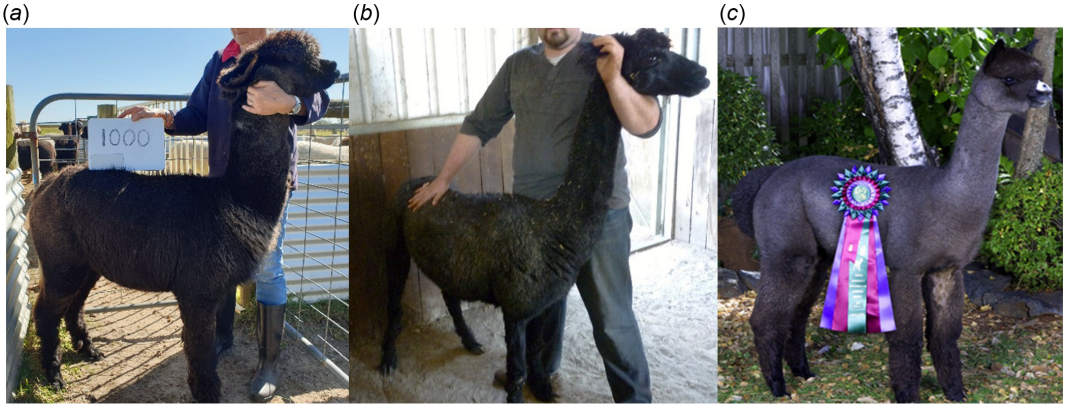


**Fig. 1.** Alpacas displaying variable expressivity of the roan coat pattern. (*a*–*c*) Extent and distribution of the pattern increases from alpaca to alpaca (as shown in *a*–*c* respectively). (*c*) White spotting, like on the nose of this alpaca may or may not be present.

## Genotyping by sequencing

All fastq files were processed through an in-house genotyping by sequencing (GBS) bioinformatics workflow, by using bash script on cloud computing infrastructure (Nimbus Pawsey), providing two instances with eight cores and 32 Gb of RAM. The GBS workflow was based on the best practices for variant-discovery analysis guide by the Broad institute (Van der Auwera and O'Connor 2020) and used open-source software. Raw FASTQ format reads were trimmed (polyG minimum length = 5) and quality filtered via Fastp v.0.23.2 (Chen *et al.* 2018) by using default parameters. Quality-controlled reads were mapped (with local alignment) to the draft VicPac4 reference genome (Brian Davis, Texas A&M, pers. comm.) via Bowtie2 v2.3.5.1, by using default parameters (Langmead and Salzberg 2012). Mapped reads were sorted and indexed via Samtools v1.10 (Danecek *et al.* 2021), followed by formatting (AddOrReplaceReadGroups) and quality control (MarkDuplicates) using Picard v2.27.1 (Broad Institute 2019). Individual BAM (binary alignment map) files were processed with GATK v4.2.6.1 HaplotypeCaller (Poplin *et al.* 2018) to produce GVCF (genomic-variant call format) files, which were subsequently merged using GATK GenomicsDBImport. Joint genotyping was performed with GATK GenotypeGVCFs by using the GenomicsDBImport workspace as input. Single-nucleotide polymorphisms (SNP) were extracted from the VCF file via GATK SelectVariants.

## Genome-wide association analysis

The unfiltered VCF file containing only SNP data was processed using PLINK v1.9 (Purcell *et al.* 2007). SNPs with a minor allele frequency of <0.05 and those missing a genotype for >5% of animals were removed. Animals missing genotypes for >5% of SNPs were also excluded. Dark roan alpacas ($n$ = 13) were selected as cases and dark non-roans ($n$ = 14) as controls. The white and light fawn animals ($n$ = 19) were excluded from the GWAS due to the ambiguity of their observed phenotype. Quality-controlled genome-wide SNPs were included in a case-control association analysis using the Fisher's exact test model in PLINK (Purcell *et al.* 2007). A genome-wide significance threshold of $P = 5 \times 10^{-8}$ was applied, and Manhattan plots were generated using SNPEVG v3.2 (Wang *et al.* 2012).

## Candidate variants

All genomic regions with genome-wide significant association ($P \leq 5 \times 10^{-8}$) and with underlying support observed on the Manhattan plot, were manually assessed for potential candidate variants. Both SNP and indel genotypes, within the complete VCF file, were considered and viewed on Integrated Genomic Viewer (IGV) v2.12.3 (Robinson *et al.* 2011). Variant search was additionally extended to the core pigment genes *KIT, MITF, ASIP, MC1R*, and those significantly downregulated in prematurely greying human hair follicles (Bian *et al.* 2019), namely, *SOX10, TYR, TYRP1, MLANA, PMEL, MATP* and *GPR143*. Gene sequences were retrieved from the VicPac3.1 gene database on NCBI (https://www.ncbi.nlm.nih.gov/gene/) and aligned to the candidate VicPac4 (draft) regions via the BLASTn v2.9.0+ command line tool (Camacho *et al.* 2009) to identify genes in those regions. Where possible, pedigree data were checked for Mendelian inheritance of variant genotype and all animals were checked for potential misclassification of phenotype through examination of available photographs and personal communication with breeders.

## PCR and sequencing of the candidate SNP

To validate the candidate SNP (Super-Scaffold_15:39 742 851T > A), 10 additional dark roans plus four animals from the GWAS were genotyped through targeted polymerase chain reaction (PCR) and subsequent DNA sequencing. Genomic DNA was extracted from blood samples following the PureLink Genomic DNA Mini Kit (Invitrogen) protocol. Forward (5′-TCATGCTCCCCTGAACAACA-3′) and reverse (5′-TCTCGTTGATGATGGTGGGAG-3′) PCR primers were designed using Primer3Plus (https://www.primer3plus.com/), having a predicted product size of 541 bp. PCR was performed using a T100 Thermal Cycler (Kyratec) in a total volume of 10 μL, containing ~100 ng genomic DNA, 2.5 U MyTaq HS DNA polymerase (Bioline), 2 μL MyTaq reaction buffer (5×) (Bioline) and 3 pmol of each primer. The amplification conditions were as follows: 1 min at 95°C, 30 cycles of 15 s at 95°C, 15 s at 60°C, 30 s at 72°C, then 5 min at 72°C. For each sample, the above PCR was performed in triplicate and the pooled products were visualised on a 2% w/v SB buffer agarose gel, followed by purification by using a FavorPrep PCR Clean-Up Mini Kit (Favorgen). Two separate sequencing reactions (forward and reverse) were performed by the Australian Genome Research Facility, by using the respective PCR primers, and following the ABI Prism BigDye Terminator Cycle Sequencing protocol. Sequence data were visualised and manually analysed using FinchTV v1.4.0 (Team Giospiza 2004).

## Copy-number variant analysis

Analysis of copy number was limited to the scaffold that contained the candidate variant and was conducted using CNVpytor v1.2.1 (Suvakov *et al.* 2021), following the developers' guide. Briefly, depth-of-coverage of mapped reads was corrected for GC bias and normalised to a diploid copy number for each sample, complemented by B-allele (referring to the non-reference or variant allele) frequency analysis of SNPs and small indels. Read-pair orientation of mapped reads was assessed in the candidate region using IGV (Robinson *et al.* 2011) to view BAM files. The proportion of roan and non-roan associated allele at the candidate SNP, and variants in visual linkage disequilibrium (LD) with it, were considered

when manually determining the roan:non-roan copy-number ratio for each sample.

## Results

In total, 28 296 476 SNPs and 4 191 807 indels were detected. Following SNP extraction and quality control, 16 863 747 SNPs were retained, and all samples had a genotyping rate of >95%. The GWAS showed 12 regions with association greater than $P \leq 5 \times 10^{-8}$ (Fig. 2*a*). However, only one region contained variant genotypes that were concordant with phenotypes (Supplementary material Fig. S1).

A candidate SNP (Super-Scaffold_15:39 742 851*T* > A) was identified 272 kb upstream of *KITLG* (Fig. 2*b*). All 13 roan alpacas were genotyped as either homozygous or heterozygous for the reference allele (*T*), and all but 3 of 14 non-roans were homozygous for the variant allele (*A*). It is interesting to note that the alpaca used to generate the reference genome is light fawn in colour. Almost half of the white and light fawn animals ($n = 19$) were heterozygous for the roan-associated allele, and one was homozygous (Table 1). Mendelian inheritance of the SNP genotype was concordant for the 12 parent-offspring trios included in the GWAS (data not shown). Additionally, pedigree analyses of all animals supported the dominant mode of inheritance of roan (or cryptic roan in the white and light fawn animals; data not shown).

On inspection of raw read depth at the candidate SNP, an allelic imbalance was observed for all except five heterozygotes ($n = 20$). Sanger sequencing of the four re-sequenced exemplar samples (*T/T*, *A/A*, *T/A* balanced, *T/A* unbalanced) was consistent with the WGS genotypes. Three of ten additional (i.e. Sanger sequence only) roan animals were genotyped *A/A* and the remaining seven had an unreliable *T/A* genotype, which supported a hypothesis of allelic imbalance at this locus.

In the Super-Scaffold_15:39 742 096–39 887 419 region, which contained the Super-Scaffold_15:39 742 851*T* > *A* candidate, an abrupt increase in depth-of-coverage was observed in all except 8 of the 46 of the BAM files, suggestive of a CNV. Subsequently, the presence of a CNV 127–272 kb upstream of *KITLG* was validated (Fig. 3*a*, *b*). The copy number ratio of the roan (*T*):non-roan (*A*) associated allele was estimated for all samples (Table 1). All eight US roan alpacas were closely related, and each had two copies of *T* and zero or one copies of *A*. Of the five Australian roan alpacas, only two were related (second generation) and all had one or two copies of *T* and one to four copies of *A*. All except three (one Australian and two New Zealand) non-roan animals ($n = 14$) had no copies of *T* and three to seven copies of *A*.

Read-pair orientation analysis of individual BAM files demonstrated a trend of right–left oriented discordant paired-end reads mapped at the boundaries of the CNV, indicative of a tandem duplication (Fig. 3*c*). A smaller 40 kb CNV region (Super-Scaffold_15:39 763 502–39 804 057) was predicted to occur in tandem within the 145 kb (Super-Scaffold_15:39 742 096–39 887 419) tandem duplication.

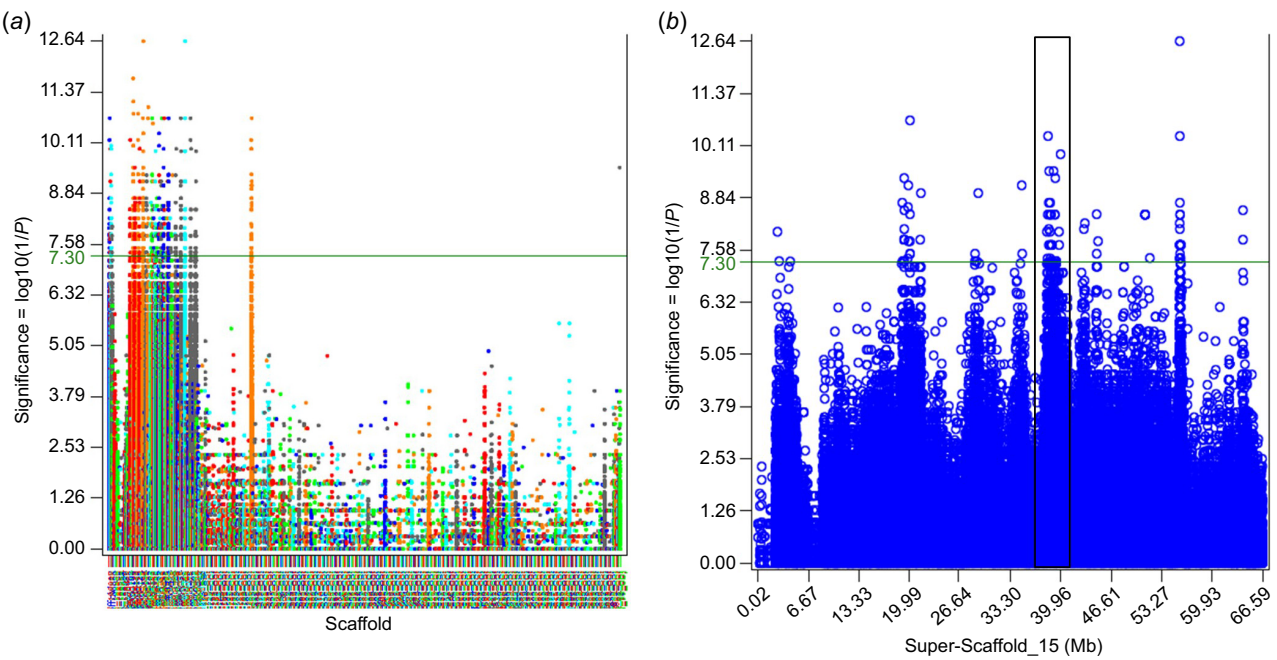


**Fig. 2.** GWAS comparing roan ($n = 13$) and non-roan ($n = 14$) alpacas. (*a*) Manhattan plot. Of the 790 scaffolds, 39 peaks reached the genome-wide significance ($P \leq 5 \times 10^{-8}$) threshold (green line). (*b*) Manhattan plot of Super-Scaffold_15 only; the candidate region is highlighted with a box.

**Table 1.** Genotypes at the candidate SNP Super-Scaffold_15:39 742 851*T* > *A* (derived from SNPs genotyped using GATK) and diploid copy number (CN) of each allele estimated relative to the CNVpytor normalised copy-number read depth of the candidate region Super-Scaffold_15:39 742 096–39 887 419.

| Phenotype (origin) | *N* | Occurrence of SNP genotype | | | Occurrence of CNV genotype | | |
|---|---|---|---|---|---|---|---|
| | | *T/T* | *T/A* | *A/A* | CN *T* = 2[A] CN *A* ≥ 0 | CN *T* = 1[A] CN *A* ≥ 1 | CN *T* = 0[A] CN *A* ≥ 2 |
| Roan alpacas | | | | | | | |
| Black (Australia) | 5 | 0 | 5 | 0 | 2 | 3 | 0 |
| Black (USA) | 3 | 2 | 1 | 0 | 3 | 0 | 0 |
| Chestnut (USA) | 2 | 1 | 1 | 0 | 2 | 0 | 0 |
| Black or brown roan + white spotting (USA) | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| Bay (USA) | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| Total | 13 | 6 | 7 | 0 | 10 | 3 | 0 |
| Non-roan alpacas | | | | | | | |
| Black (Australia) | 6 | 0 | 1 | 5 | 0 | 1 | 5 |
| Black (New Zealand) | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Bay (New Zealand) | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Black or bay + white spotting (Australia) | 2 | 0 | 0 | 2 | 0 | 0 | 2 |
| Total | 14 | 0 | 3 | 11 | 0 | 3 | 11 |
| White and light fawn alpacas | | | | | | | |
| Blue eyed white (Australia) | 2 | 0 | 0 | 2 | 0 | 0 | 2 |
| Classic grey (Australia) | 5 | 0 | 1 | 4 | 0 | 1 | 4 |
| Classic grey (New Zealand) | 3 | 0 | 2 | 1 | 0 | 2 | 1 |
| Fawn + white spotting (Australia) | 2 | 1 | 1 | 0 | 1 | 1 | 0 |
| White/fawn (New Zealand) | 7 | 0 | 6 | 1 | 0 | 6 | 1 |
| Total | 19 | 1 | 10 | 8 | 1 | 10 | 8 |

[A]The CN of the roan-associated allele (*T*) detected in a single animal ranged from 0 to 2, whereas the CN of the non-roan-associated allele (*A*) ranged from 0 to 7.

All progeny *T:A* counts were concordant with those expected through Mendelian inheritance. We could unambiguously infer the CNV haplotypes within two nuclear families by using Mendelian inheritance principles (Fig. 4). These were the only families in which at least one individual had a normal diploid copy count (CN = 2) in the homozygous state. All US roan animals within a single family had at least one haplotype with just the *T* allele version. The other family of four consisted of two non-roan and two classic grey animals, and all except the dam (non-roan) had only haplotypes composed of the *A* allele copies. Possible misclassification of the dam's phenotype was checked, and she did not roan at a later stage in life.

## Discussion

These data suggest an association of the *T* allele version of a 145 kb tandem CNV upstream of *KITLG* with the roan phenotype in alpacas. It is also tempting to speculate that the roaning effect of *T* is suppressed when in *cis* with the non-roan associated version *A*. In support of this concept, one non-roan animal had a haplotype composed of one *T* and two *A* copies (cn = *AAT*; Fig. 4). Similarly, two roan animals, each with cn = *TA* and cn = *T* haplotypes, expressed roan to a lesser extent than did their roan relatives with only the cn = *T* haplotype (Fig. 4).

The 145 kb candidate CNV ends 127 kb upstream of *KITLG*. HapMap data have shown that 90% of CNVs have two or more allelic states (McCarroll *et al.* 2008; Handsaker *et al.* 2015) and that 66.7% of haplotypes consisting of copy-number gains are composed of alternative allelic copies (Palta *et al.* 2015), which reflects what we have discovered here. The association of the genes *KITLG* or its receptor *KIT* with the roan phenotype has been well established in other species (Charlier *et al.* 1996; Marklund *et al.* 1999; Cho *et al.* 2011; Lim *et al.* 2011; Fontanesi and Russo 2013; Talenti *et al.* 2018; Grilz-Seger *et al.* 2020; Voß *et al.* 2020). More compelling evidence that this CNV could be associated with the roan trait in alpacas is that *KITLG* expression in mice is regulated by a ~200 kb region 100 kb upstream of the gene (Bedell *et al.* 1995), and mutations in this region cause depigmentation,

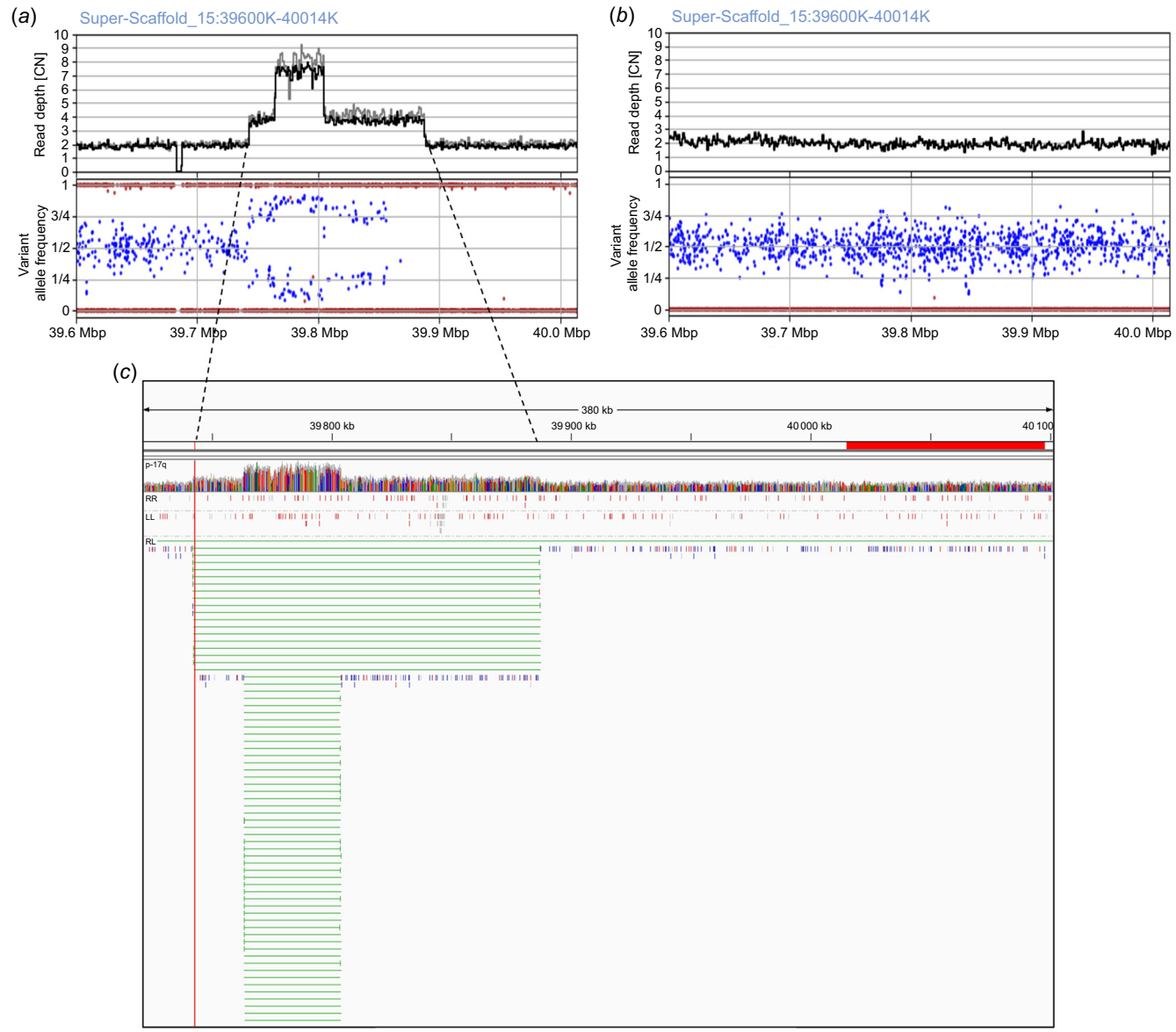


**Fig. 3.** CNV and read-pair orientation analysis. (*a*, *b*) CNVpytor output of BAM-file analysis on two roan animals. The top panel shows normalised read depth to a diploid copy number (CN = 2) and the bottom panel is the corresponding B-allele (variant allele) frequency analysed using the VCF file. (*a*) CNVpytor output depicting a roan animal with a copy number gain and B-allele frequency deviating from 0.5 in the same region, representing allelic imbalance. (*b*) CNVpytor output depicting a roan animal with a normal diploid copy number. (*c*) IGV interface of Illumina paired-end reads (green) mapped in the right–left orientation at the extremities of two regions (Super-Scaffold_15:39 742 096–39 887 419 and Super-Scaffold_15:39 763 502–39 804 057) with a conspicuous increase in depth-of-coverage, indicating a tandem duplication. The position of the candidate SNP is marked with a vertical red line and *KITLG* is highlighted in red (top right).

anaemia and embryonic death (Bedell *et al.* 1996). Similarly, a study comparing the parallel evolution of colour variation in human and stickleback fish has attributed variation in *KITLG* expression to *cis*-regulatory changes in a large region upstream of the gene (Miller *et al.* 2007). Therefore, a potential mechanism of action for roan in alpacas is disruption to a *cis*-regulatory element involved in *KITLG* regulation, such as an enhancer domain, caused by the *T*-allele version, which could be rescued when the functional *A*-allele version is in *cis*.

An allele dosage effect may also contribute to the expression of roan. The observed copy-number counts varied both among and between roan and non-roan animals. The variation in copy number was mostly attributed to the *A*-allele version (Table 1). This not only reflects the strong (artificial) selection for non-roan alpacas but also suggests that the dosage of *A* may be associated with the extent or onset of roan. A study in dogs identified a 6 kb CNV located 152 kb upstream of *KITLG* that was associated with

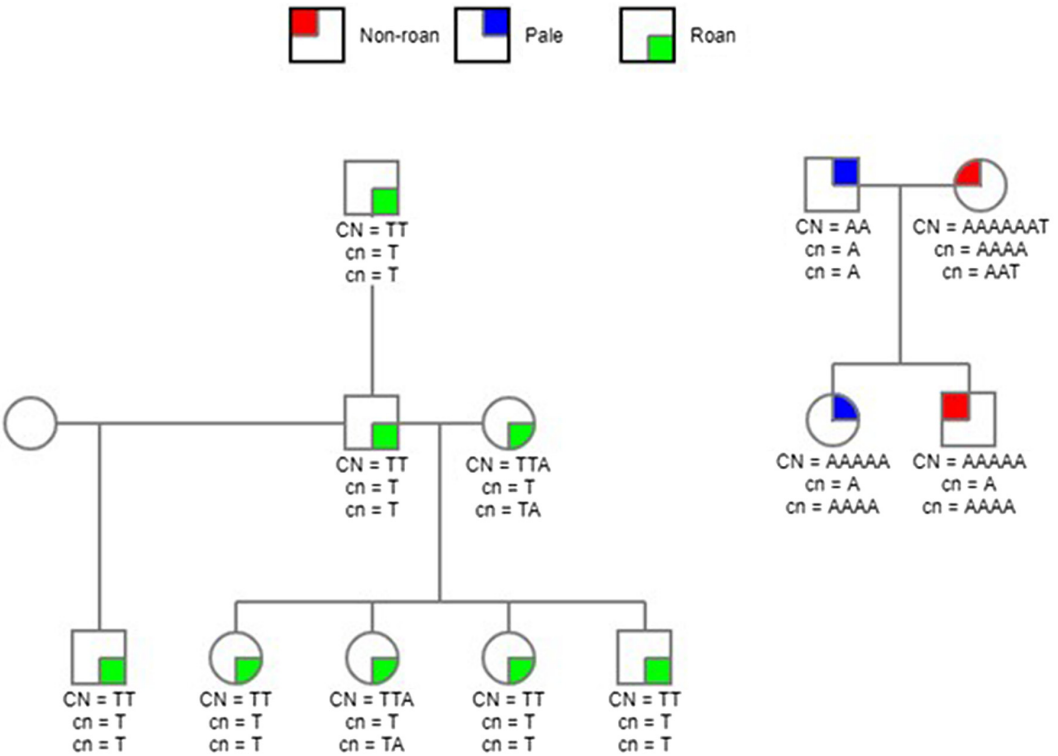


**Fig. 4.** Pedigree information of a US (left) and Australian (right) alpaca family and the respective roan-associated (*T*) and non-roan-associated (*A*) allele copy-number composition of inherited haplotypes, generated using Progeny Genetics online tool (https://pedigree.progenygenetics.com/). 'CN' refers to the diploid copy number and 'cn' refers to the phased haploid copy number of *T* and/or *A*. All animals with a phenotype label were those originally used in the WGS.

pigment intensity, with a higher copy number correlating to higher intensity (Weich *et al.* 2020). Since the early depletion of hair follicle melanoblasts, thus causing gradual and permanent depigmentation, is the probable cause of roan (Hachiya *et al.* 2009; Endou *et al.* 2014; Bian *et al.* 2019; Qiu *et al.* 2019), it is plausible that a related mechanism may also contribute to pigment intensity. Although we did not account for variation in pigment intensity in this work, the observed *A* copy-version counts were higher among the Australian (unrelated) roans than the US roans (related). Unlike the Australian roans, the US roans were selectively bred for the roan pattern. This information, in addition to the fact that the US alpacas displayed the pattern earlier than did the Australian roans, suggests that a higher copy number of the *A* version may reduce, or delay, the expression of roan.

Interestingly, the alpaca that is the source of the reference genome also had a cn = *T* haplotype. Considering her light fawn base colour and darker extremities, it is plausible that she was a cryptic roan. Almost half of the sampled white or light fawn animals had one copy of *T*, and although unphased, this observation is consistent with the knowledge that cryptic roan animals often go undetected until mated with dark base-colour animals. All but three dark non-roan animals had varying copy numbers of *A* alone (Table 1), indicating homozygosity at the putative functionally important site. An autosomal incomplete dominant inheritance pattern can be inferred from these data, whereby a single haplotype consisting solely of the *T* allele copy version likely confers the expression of roan. There is no evidence of homozygous lethality as evidenced by six roan alpacas (USA) that are homozygous for the *T* allele version of the CNV (Table 1).

Breed specific allelic heterogeneity owing to potential founder effect is known in roan horses (Grilz-Seger *et al.* 2020; Voß *et al.* 2020) and pigs (Cho *et al.* 2011; Lim *et al.* 2011). Therefore, we cannot exclude the possibility of multiple causative variants for roan in alpacas or that the candidate SNP may be in LD with the true casual variant(s). This may explain why 3 of the 10 additional roan animals genotyped *A/A* at the candidate SNP. However, all 10 animals were sampled from the same farm and the three non-concordant animals developed the roan pattern at a much later age than did the rest (Windella and Goldleaf alpacas, Dr C. Oddie and Mr B. Fallon pers. comm.). It is possible that the late onset of pattern development was age-related, which highlights the importance of recording the onset and progression of the roan phenotype.

We hypothesise that a *cis*-regulatory mechanism alters the expression of *KITLG* and causes premature depigmentation of

individual fibres, which may be further influenced by allelic dosage effects. This study has provided novel insights into the genetic basis of the roan coat pattern in alpacas and suggests that the *KITLG* CNV influences both pheomelanic and eumelanic base-colour phenotypes.

## Supplementary material

Supplementary material is available online.

## References


Baxter LL, Watkins-Chow DE, Pavan WJ, Loftus SK (2019) A curated gene list for expanding the horizons of pigmentation biology. *Pigment Cell & Melanoma Research* **32**, 348–358. doi:10.1111/pcmr.12743

Bedell MA, Brannan CI, Evans EP, Copeland NG, Jenkins NA, Donovan PJ (1995) DNA rearrangements located over 100 kb 5' of the Steel (Sl)-coding region in Steel-panda and Steel-contrasted mice deregulate Sl expression and cause female sterility by disrupting ovarian follicle development. *Genes & Development* **9**, 455–470. doi:10.1101/gad.9.4.455

Bedell MA, Cleveland LS, O'Sullivan TN, Copeland NG, Jenkins NA (1996) Deletion and interallelic complementation analysis of *Steel* mutant mice. *Genetics* **142**, 935–944. doi:10.1093/genetics/142.3.935

Besmer P, Manova K, Duttlinger R, Huang EJ, Packer A, Gyssler C, Bachvarova RF (1993) The *kit*-ligand (steel factor) and its receptor *c-kit*/W: pleiotropic roles in gametogenesis and melanogenesis. *Development* **119**(Suppl.), 125–137. doi:10.1242/dev.119.Supplement.125

Bian Y, Wei G, Song X, Yuan L, Chen H, Ni T, Lu D (2019) Global downregulation of pigmentation-associated genes in human premature hair graying. *Experimental and Therapeutic Medicine* **18**, 1155–1163. doi:10.3892/etm.2019.7663

Brancalion L, Haase B, Wade CM (2022) Canine coat pigmentation genetics: a review. *Animal Genetics* **53**, 3–34. doi:10.1111/age.13154

Broad Institute (2019) 'Picard tools.' Broad Institute, GitHub repository. http://broadinstitute.github.io/picard/

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421. doi:10.1186/1471-2105-10-421

Charlier C, Denys B, Belanche JI, Coppieters W, Grobet L, Mni M, Womack J, Hanset R, Georges M (1996) Microsatellite mapping of the bovine roan locus: a major determinant of White Heifer Disease. *Mammalian Genome* **7**, 138–142. doi:10.1007/s003359900034

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890. doi:10.1093/bioinformatics/bty560

Cho I-C, Zhong T, Seo B-Y, Jung E-J, Yoo C-K, Kim J-H, Lee J-B, Lim H-T, Kim B-W, Lee J-H, Ko M-S, Jeon J-T (2011) Whole-genome association study for the roan coat color in an intercrossed pig population between Landrace and Korean native pig. *Genes & Genomics* **33**, 17–23. doi:10.1007/s13258-010-0108-4

Cieslak M, Reissmann M, Hofreiter M, Ludwig A (2011) Colours of domestication. *Biological Reviews* **86**, 885–899. doi:10.1111/j.1469-185X.2011.00177.x

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi:10.1093/gigascience/giab008

Endou M, Aoki H, Kobayashi T, Kunisada T (2014) Prevention of hair graying by factors that promote the growth and differentiation of melanocytes. *The Journal of Dermatology* **41**, 716–723. doi:10.1111/1346-8138.12570

Feeley NL, Munyard KA (2009) Characterisation of the melanocortin-1 receptor gene in alpaca and identification of possible markers associated with phenotypic variations in colour. *Animal Production Science* **49**, 675–681. doi:10.1071/AN09005

Feeley NL, Bottomley S, Munyard KA (2011) Three novel mutations in *ASIP* associated with black fibre in alpacas (*Vicugna pacos*). *The Journal of Agricultural Science* **149**, 529–538. doi:10.1017/S0021859610001231

Fontanesi L, Russo V (2013) Molecular genetics of coat colour in pigs. *Acta Agriculturae Slovenica* 4, 16.

Grilz-Seger G, Reiter S, Neuditschko M, Wallner B, Rieder S, Leeb T, Jagannathan V, Mesarič M, Cotman M, Pausch H, Lindgren G, Velie B, Horna M, Brem G, Druml T (2020) A genome-wide association analysis in Noriker horses identifies a SNP associated with roan coat color. *Journal of Equine Veterinary Science* **88**, 102950. doi:10.1016/j.jevs.2020.102950

Hachiya A, Sriwiriyanont P, Kobayashi T, Nagasawa A, Yoshida H, Ohuchi A, Kitahara T, Visscher MO, Takema Y, Tsuboi R, Boissy RE (2009) Stem cell factor-KIT signalling plays a pivotal role in regulating pigmentation in mammalian hair. *The Journal of Pathology* **218**, 30–39. doi:10.1002/path.2503

Handsaker RE, Van Doren V, Berman JR, Genovese G, Kashin S, Boettger LM, McCarroll SA (2015) Large multiallelic copy number variations in humans. *Nature Genetics* **47**, 296–303. doi:10.1038/ng.3200

Hemesath TJ, Price ER, Takemoto C, Badalian T, Fisher DE (1998) MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes. *Nature* **391**, 298–301. doi:10.1038/34681

Hintz HF, Van Vleck LD (1979) Lethal dominant roan in horses. *Journal of Heredity* **70**, 145–146. doi:10.1093/oxfordjournals.jhered.a109213

Jones M, Sergeant C, Richardson M, Groth D, Brooks S, Munyard K (2019) A non-synonymous SNP in exon 3 of the *KIT* gene is responsible for the classic grey phenotype in alpacas (*Vicugna pacos*). *Animal Genetics* **50**, 493–500. doi:10.1111/age.12814

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359. doi:10.1038/nmeth.1923

Lim HT, Zhong T, Cho IC, Seo BY, Kim JH, Lee SS, Ko MS, Park HB, Kim BW, Lee JH, Jeon JT (2011) Novel alternative splicing by exon skipping in KIT associated with whole-body roan in an intercrossed population of Landrace and Korean Native pigs. *Animal Genetics* **42**, 451–455. doi:10.1111/j.1365-2052.2011.02225.x

Marklund S, Moller M, Sandberg K, Andersson L (1999) Close association between sequence polymorphism in the *KIT* gene and the roan coat color in horses. *Mammalian Genome* **10**, 283–288. doi:10.1007/s003359900987

Mathews M, Mathews T, Kostiakos ECK (2019) Gorgeous Greys. Camelid Connections Magazine – Specialty Publications. Candelo New South Wales, Australia. https://camelidconnections.com.au

McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PIW, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics* **40**, 1166–1174. doi:10.1038/ng.238

Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, Kingsley DM (2007) *cis*-regulatory changes in kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**, 1179–1189. doi:10.1016/j.cell.2007.10.055

Munyard K (2013) 'Inheritance of white colour in alpacas: identifying the genes involved.' (Rural Industries Research and Development Corporation: Wagga Wagga, NSW, Australia)

Palta P, Kaplinski L, Nagirnaja L, Veidenberg A, Möls M, Nelis M, Esko T, Metspalu A, Laan M, Remm M (2015) Haplotype phasing and inheritance of copy number variants in nuclear families. *PLoS ONE* **10**, e0122713. doi:10.1371/journal.pone.0122713

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E (2018) Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv: The Preprint Server for Biology* 201178. doi:10.1101/201178

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575. doi:10.1086/519795

Qiu W, Chuong C-M, Lei M (2019) Regulation of melanocyte stem cells in the pigmentation of skin and its appendages: biological patterning and therapeutic potentials. *Experimental Dermatology* **28**, 395–405. doi:10.1111/exd.13856

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nature Biotechnology* **29**, 24–26. doi:10.1038/nbt.1754

Seitz JJ, Schmutz SM, Thue TD, Buchanan FC (1999) A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mammalian Genome* **10**, 710–712. doi:10.1007/s003359901076

Suvakov M, Panda A, Diesh C, Holmes I, Abyzov A (2021) CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *GigaScience* 10, giab074. doi:10.1093/gigascience/giab074

Talenti A, Bertolini F, Williams J, Moaeen-ud-Din M, Frattini S, Coizet B, Pagnacco G, Reecy J, Rothschild MF, Crepaldi P, Italian Goat Consortium (2018) Genomic analysis suggests *KITLG* is responsible for a roan pattern in two Pakistani goat breeds. *Journal of Heredity* **109**, 315–319. doi:10.1093/jhered/esx093

Team Geospiza (2004) 'FinchTV 1.4.0.' (Geospiza, Inc.: Seattle, WA, USA)

Thangavel K, Rathinamoorthy R, Ganesan P (2015) Sustainable luxury natural fibers – production, properties, and prospects. In 'Handbook of sustainable luxury textiles and fashion. Vol. 1'. (Eds MA Gardetti, SM Subramanian) pp. 59–98. (Springer: Singapore)

Toro MA (2010) Future trends in animal breeding due to new genetic technologies. *Advances in Animal Biosciences* **1**, 546–557. doi:10.1017/S2040470010005431

Van der Auwera GA, O'Connor BD (2020) 'Genomics in the cloud: using Docker, GATK, and WDL in terra.' 1st edn. (O'Reilly Media: Sebastopol, CA, USA)

Voß K, Tetens J, Thaller G, Becker D (2020) Coat color roan shows association with KIT variants and no evidence of lethality in Icelandic horses. *Genes* **11**, 680. doi:10.3390/genes11060680

Wang S, Dvorkin D, Da Y (2012) SNPEVG: a graphical tool for GWAS graphing with mouse clicks. *BMC Bioinformatics* **13**, 319. doi:10.1186/1471-2105-13-319

Wehrle-Haller B (2003) The role of kit-ligand in melanocyte development and epidermal homeostasis. *Pigment Cell Research* **16**, 287–296. doi:10.1034/j.1600-0749.2003.00055.x

Weich K, Affolter V, York D, Rebhun R, Grahn R, Kallenberg A, Bannasch D (2020) Pigment intensity in dogs is associated with a copy number variant upstream of *KITLG*. *Genes* **11**, 75. doi:10.3390/genes11010075

Zsebo KM, Williams DA, Geissler EN, Broudy VC, Martin FH, Atkins HL, Hsu R-Y, Birkett NC, Okino KH, Murdock DC, *et al.* (1990) Stem cell factor is encoded at the *SI* locus of the mouse and is the ligand for the c-*kit* tyrosine kinase receptor. *Cell* **63**, 213–224. doi:10.1016/0092-8674(90)90302-U

**Data availability.** The data that support this study will be shared upon reasonable request to the corresponding author.

**Conflicts of interest.** The authors declare no conflicts of interest.

**Declaration of funding.** This research was funded by the Morris Animal Foundation (Grant # D18LA-002).

**Acknowledgements.** This research was supported by funding from the Morris Animal Foundation (D18LA-002) and the Curtin Medical School. This work was also supported by resources provided by the Pawsey Supercomputing Research Centre with funding from the Australian Government and the Government of Western Australia. Thanks go to all of the alpaca owners and breeders who provided access to their animals, with special mention to Dr Carolyn Oddie (Windella alpacas), Brett Fallon and Tyler Bennie (Goldleaf alpacas) for their help with sample collection and wisdom on alpaca breeding practices.

**Author affiliations**

[A]Curtin Medical School and Curtin Health Innovation Research Institute, Faculty of Health Sciences, Curtin University, Perth, WA, Australia.

[B]Animal Sciences, University of Florida, Gainesville, FL, USA.