Received: 13 June 2023 | Revised: 10 November 2023 | Accepted: 24 November 2023

DOI: 10.1002/biot.202300287

Biotechnology Journal

RESEARCH ARTICLE

Site specific insertion of a transgene into the murine α-casein (CSN1S1) gene results in the predictable expression of a recombinant protein in milk

Christopher Knowles[1] | Linda Petrie[1] | Claire Warren[2] | Simon G. Lillico[2] | Ailsa Carlisle[2] | C. Bruce A. Whitelaw[2] | Andreas F. Kolb[1]

[1]Nutrition, Obesity and Disease Research Theme, Rowett Institute, University of Aberdeen, Aberdeen, UK

[2]Roslin Institute, University of Edinburgh, Edinburgh, UK

Correspondence
Andreas Kolb, Nutrition, Obesity and Disease Research Theme, Rowett Institute, University of Aberdeen, Foresterhill, Aberdeen, UK.
Email: a.kolb@abdn.ac.uk

Abstract

Gene loci of highly expressed genes provide ideal sites for transgene expression. Casein genes are highly expressed in mammals leading to the synthesis of substantial amounts of casein proteins in milk. The α-casein (CSN1S1) gene has assessed as a site of transgene expression in transgenic mice and a mammary gland cell line. A transgene encoding an antibody light chain gene (A1L) was inserted into the α-casein gene using sequential homologous and site-specific recombination. Expression of the inserted transgene is directed by the α-casein promoter, is responsive to lactogenic hormone activation, leads to the synthesis of a chimeric α-casein/A1L transgene mRNA, and secretion of the recombinant A1L protein into milk. Transgene expression is highly consistent in all transgenic lines, but lower than that of the α-casein gene (4%). Recombinant A1L protein accounted for 0.5% and 1.6% of total milk protein in heterozygous and homozygous transgenic mice, respectively. The absence of the α-casein protein in homozygous A1L transgenic mice leads to a reduction of total milk protein and delayed growth of the pups nursed by these mice. Overall, the data demonstrate that the insertion of a transgene into a highly expressed endogenous gene is insufficient to guarantee its abundant expression.

KEYWORDS
cellular engineering, expression systems, genomic engineering, trangenic animals

1 | INTRODUCTION

Substantial quantities of pharmaceutically relevant proteins can be expressed in transgenic animals.[1,2] Expression in milk and eggs has received most attention and commercial funding as substantial amounts of recombinant protein can be produced in these systems.[3,4] The mammary gland is able to synthesize multi-component proteins which are correctly folded and carry the expected glycosylation and phosphorylation patterns.[1,5] One protein, alpha-antithrombin (marketed under the tradename Atryn), a blood clotting inhibitor derived from goat milk, has been in clinical use since 2009.[1] However, the generation of transgenic animals is expensive and time-consuming and the process competes with simpler expression systems like bacteria, yeast and mammalian tissue culture cells, which can all be grown at scale.[6]

Abbreviations: ER, endoplasmic reticulum; Hytk, hygromycin-phosphotransferase-thymidine kinase fusion gene; neoR, neomycin-phosphotransferase gene; RMCE, recombinase mediated cassette exchange.

One limitation which has hampered the progress of the transgenic animals as bioreactors was that microinjection of naked DNA into fertilized oocytes leads to the random integration of the transgene construct. The integration site often exerts a dominant and silencing effect on transgene expression. Hence, many transgenic lines need to be screened to identify one that expresses the transgene at sufficient levels. Approaches to overcome these limitations have included locus control regions, matrix attachment regions, lentiviral vectors, transposons and, more recently, the use of CRISPR.[7–9] Most of the limitations of animal transgenesis also apply to the expression of recombinant proteins in cultured cells.[10]

Expression of transgenes in the mammary gland can be achieved by linking the transgene of interest with a mammary gland specific promoter.[11,12] However, the expression of the recombinant protein is highly variable between transgenic lines.[13] To overcome this variability the transgene could be inserted into a genomic locus which is highly active in the lactating mammary gland.[14] One appropriate genomic location for this approach is the casein gene locus. Caseins are serine-rich phosphoproteins which are almost exclusively expressed in the lactating mammary gland.[15] The casein proteins together account for up to 80% of total milk protein and 70% of total mammary gland mRNA in cattle.[16,17] Caseins are highly phosphorylated, associate with calcium in mammary epithelial cells and are secreted into milk in the form of a casein micelle.[18] Different caseins play distinct roles in micelle formation. While inactivation of the mouse β-casein (CSN2) protein has no detrimental effects on milk production,[19] inactivation of the κ-casein (CSN3) gene is incompatible with milk production.[20] We have shown that inactivation of the α-casein gene in mice reduces milk protein synthesis significantly, but is compatible with pup survival and development.[21,22] A naturally occurring mutation of the α-casein gene in goats is compatible with normal milk production.[23] This demonstrates that caseins have different functional roles in casein micelle formation and milk protein secretion and that these roles vary across species.

We have also previously shown that a β-galactosidase gene inserted into the β-casein gene in mammary epithelial cells adopts the expression rate and the regulatory pattern of the β-casein gene that it has replaced.[14] Encouraged by our findings in cell culture, we now wanted to test whether such an approach would also be successful if a transgene were to be inserted into the mouse α-casein gene in vivo.

In the experiments outlined below we utilized the light chain of the antibody A1 (A1L) as a transgene.[11] The antibody A1 neutralizes the infection of susceptible cells with the mouse coronavirus MHV-JHM. Overexpression of the antibody in the milk of transgenic dams protects newborn mice against a lethal MHV-JHM infection.[11]

We inserted the A1L transgene into the mouse α-casein gene in transgenic mice and in cell culture. Our results demonstrate that while transgene expression is consistent between different transgenic lines, the expression of the A1L transgene does not reach the expression levels of the α-casein gene it has replaced. These results have important implications for the construction of expression systems for recombinant proteins in transgenic animals and cell culture.

2 | MATERIALS AND METHODS

2.1 | Animal experimentation

Transgenic mice were generated at Genoway (Lyon, France) on a SV129xC57BL/6 mixed background. The mice were maintained at the Roslin Institute in accordance with Home Office guidelines. This study was approved by the Roslin Institute Animal Ethics Committee and was performed under Home Office License 60/3779. Housing and maintenance were carried out as described previously.[21]

2.2 | Protein analysis

Milk was isolated at peak lactation (day 10) after cervical dislocation of the mice and processed as described previously.[21] Protein samples were stored at −80°C. Milk samples were separated on a 10% polyacrylamide gel and stained with Coomassie Blue. Western blots were carried out after semi-dry transfer of the proteins to a nitrocellulose membrane as described.[21] Protein expression was quantified on a Kodak Imaging Station using the Kodak 1D imaging software.

Additional information on materials and methods are found in the Supplementary materials.

3 | RESULTS

The objective of the experimentation was twofold. First, to test whether site directed integration of a transgene would lead to uniform expression of the transgene in different transgenic lines, and, secondly, to test whether site directed integration would phenocopy the expression rates of the gene which was replaced by the transgene.

3.1 | Generation of A1L transgenic mice

We chose to address these objectives using the α-casein gene (Figure S1) which is among the most highly expressed genes in eukaryotic cells[17] and may therefore provide an excellent platform for the expression of recombinant proteins in milk.

The coding region of the light chain of antibody A1 (A1L)[11] was inserted into the second exon of the α-casein gene using sequential homologous and site-specific recombination (Figure 1). The expected gene structure after these modifications is shown in Figure 1A and Figures S2 and S3; the expected transcripts generated from the endogenous and the modified allele are shown in Figure 2A. Note that the first exon and parts of exon 2 of the α-casein gene are part of the chimeric transgene mRNA. However, the endogenous translational start codon of the α-casein gene was mutated (Figure S2). The transgene contains its own translational start signal, its own signal peptide, and its own polyadenylation site. Therefore α-casein exons 3 to 34 are not part of the transgene mRNA. The correct

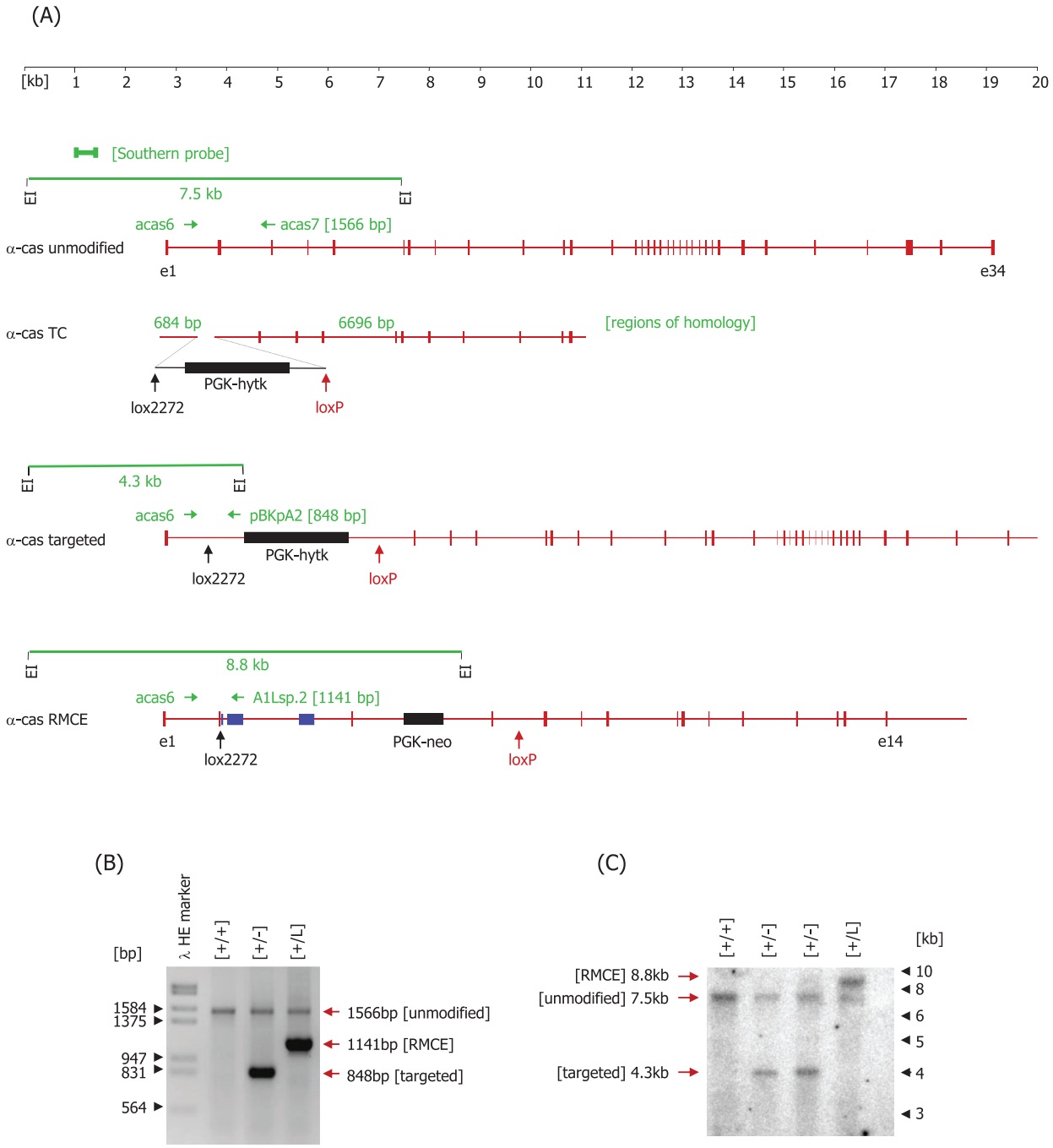

FIGURE 1 Panel A: Schematic representation of the unmodified mouse α-casein gene (α-cas unmodified), the α-casein targeting construct (α-cas TC), the α-casein gene modified by homologous recombination (α-cas targeted), and the α-casein gene modified by homologous and site-specific recombination (α-cas RMCE). The exon-intron structure (exons as red boxes, introns as red lines) from exon 1 (e1) to exon 34 (e34) is indicated. The hytk and the neoR selectable marker genes are represented as black boxes. The exons of the A1L transgene encoding the light chain of antibody A1 are represented as blue boxes. The position of recombinase target sites are indicated as vertical arrows (lox2272, loxP). The position of primer binding sites used for PCR genotyping are indicated as green arrows. The expected PCR product sizes are indicated. The relative positions of the EcoRI restriction sites (EI) and the Southern blot probe (Southern probe) used for Southern blot genotyping (green lines) are indicated, as are the sizes of the expected hybridising DNA fragments. Panel B: PCR analysis of genomic DNA isolated from the three representative ES cell clones (wild-type [+/+], gene targeted [+/−], and modified by RMCE [+/L]) using the primer combination acas6, acas7, pBKpA2 and A1Lsp.2. A 1566 bp band representing the unmodified α-casein allele [unmodified] is detected in all samples. An α-casein allele modified by homologous recombination is indicated by the 848 bp PCR product [targeted]. An allele modified by additional site-specific recombination (RMCE) is indicated by a 1141 bp PCR product [RMCE]. Phage λ digested with HindIII and EcoRI is used as molecular weight marker. Panel C: Southern blot analysis of EcoRI digested DNA derived from four representative ES cell clones. The probe indicated in panel A detects a 7.5 kb DNA fragment in the unmodified α-casein allele [unmodified], a 4.3 kb band in the targeted α-casein allele [targeted], and an 8.8 kb band in the RMCE modified α-casein allele [RMCE].

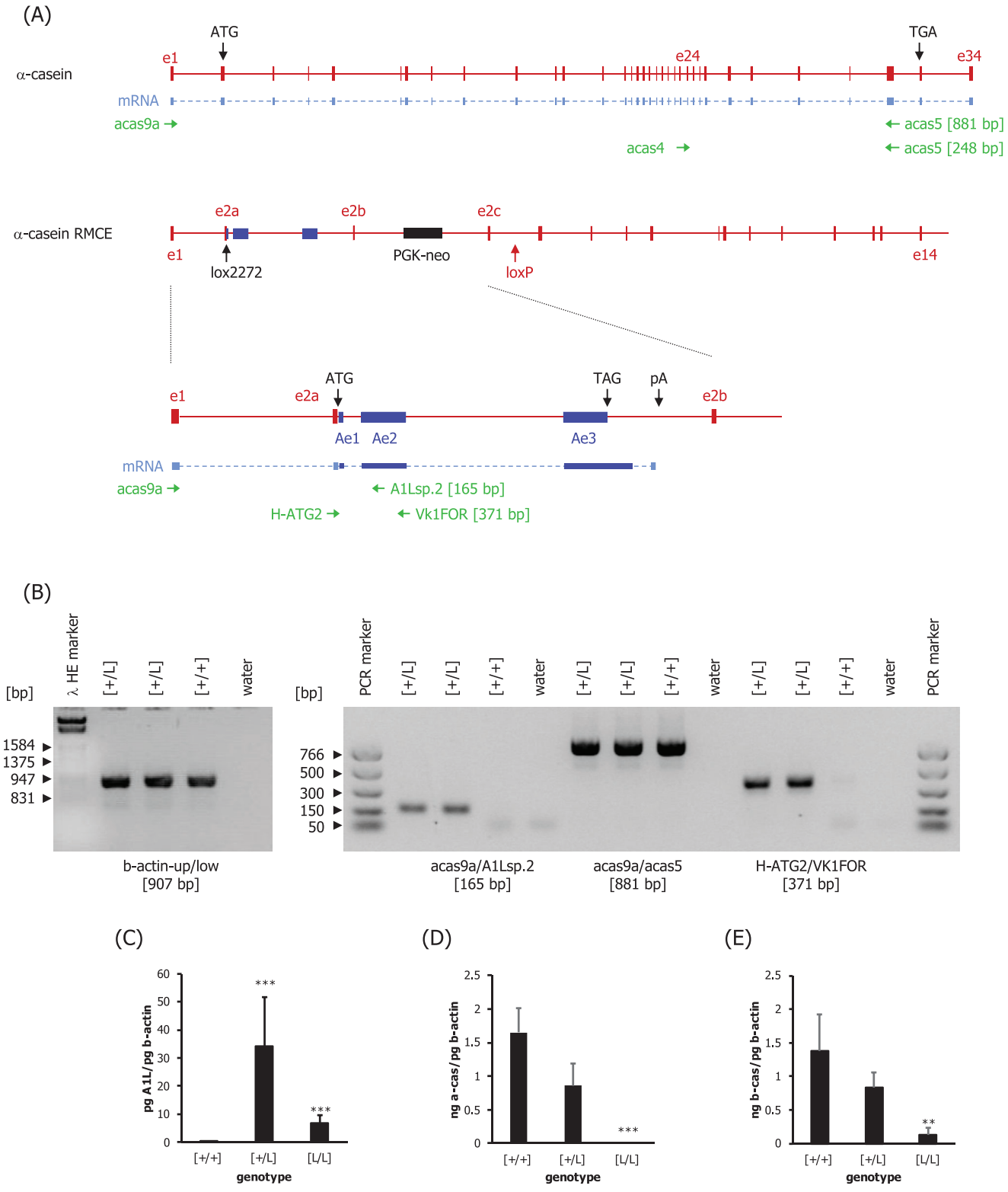

FIGURE 2 A1L transgene mRNA expression. Panel A: Schematic representation of the unmodified and RMCE modified mouse α-casein genes and the mRNAs derived from it. The α-casein exon-intron structure (exons as boxes, introns as lines) from exon 1 (e1) to exon 34 (e34) is shown. The exons of the A1L antibody light chain gene are indicated as blue boxes (Ae1, Ae2, Ae3). The position of the translational start (ATG) and termination (TAG) sites and the polyadenylation site (pA) are marked with a vertical arrow. The predicted structure of the mRNAs derived from the unmodified and the RMCE modified alleles are indicated as light blue boxes. The position of the primer binding sites for the oligonucleotides used for RT-PCR and qPCR are indicated by green arrows. The expected PCR product sizes are indicated. Insertion of the A1L gene into the α-casein locus is expected to generate a chimeric mRNA incorporating exons 1 and 2 (partially) of the α-casein gene and exons 1 through 3 of the A1L gene. The translational start site of the α-casein gene (located in exon 2) has been mutated (from ATG to TAG). The ATG of the A1L light chain gene is the

modification of the α-casein gene was assessed by PCR (primer sequences are shown in supplementary table S1) and confirmed by Southern blotting (Figure 1B,C). The successfully modified ES cells were used to generate transgenic mice originally on a C56B/6-SV129/OLA background. For long term maintenance the mice were crossed onto a CD1 background.

3.2 | Transgene mRNA expression is consistent in A1L transgenic mouse lines

Expression of the endogenous casein genes and the RMCE modified allele were analyzed by RT-PCR and qPCR of RNA derived from mammary gland tissue at peak lactation (day 10) (RT-PCR and qPCR primer sequences, amplicon sizes and annealing temperatures are shown in supplementary table S2). Figure 2A indicates the primer pairs which were used for the RT-PCR analysis. The RMCE event is expected to link the first exon of the α-casein gene with the 5′ end of the second exon of the α-casein gene which in turn is fused to the first exon of the IgG light chain gene A1L (encoding the IgG signal peptide) (Figure S3). The 3′ end of the IgG gene carries an extra SV40 intron and an SV40 polyadenylation site (Figure 2A).

Expression of the α-casein gene and the β-actin gene was detected in heterozygous [+/L] and wild-type [+/+] animals (Figure 2B). The cDNAs indicative of the RMCE modified α-casein-A1L allele (acas9a/A1Lsp.2 and H-ATG2/VK1FOR) were only detected in the heterozygous animals [+/L] (Figure 2B). The DNA sequence of the PCR products was confirmed by Sanger sequencing (Figure S3B).

Expression of the α-casein, β-casein and A1L genes was quantified in mammary gland tissue by qPCR. As expected, the expression of α-casein is undetectable in α-casein deficient/homozygous A1L transgenic mice [L/L] and reduced to around 50% in heterozygous A1L transgenic mice [+/L] (Figure 2D). The expression of β-casein is also significantly reduced in homozygous α-casein deficient/A1L transgenic mice (Figure 2E). The reduction in β-casein expression is most likely due to endoplasmic reticulum stress in mammary epithelial cells elicited by the absence of α-casein. Alpha casein has a critical functional role in the formation of casein micelles. Its absence leads to a reduction in overall milk protein content by more than 50%.[21] A1L mRNA expression is not detected in wild-type [+/+] animals, but is readily detected in heterozygous [+/L] and homozygous A1L [L/L] transgenic animals (Figure 2C). Expression of the A1L transgene under the control of the endogenous α-casein promoter is reduced in homozygous [L/L] mice (Figure 2C) relative to heterozygous [+/L] mice despite both α-casein alleles carrying the transgene. Comparison of the expression rates of the A1L transgene and the α-casein gene with gene specific standard curves demonstrates that expression of the chimeric α-casein-A1L mRNA in heterozygous mice [+/L] is around 4% of that of the endogenous α-casein gene (35pg A1L/pg β-actin and 800pg α-casein/pg β-actin) (Figure S4).

Expression of the α-casein, β-casein and A1L genes was also analyzed in kidney, liver and spleen tissue (Figure S5). None of the three genes is expressed at detectable amounts in tissues other than the mammary gland. The neomycin resistance gene (neo), which is under transcriptional control of the ubiquitous phosphoglycerol-kinase (PGK) promoter, shows low levels of expression in kidney and liver tissue, but no detectable expression in spleen (Figure S5). Intriguingly the expression of the neo gene is substantially (around 40-fold) higher in mammary tissue compared to kidney or liver tissue suggesting that the genomic context has a dominant effect on PGK promoter activity. This effect must be mediated by the chromatin structure surrounding the selection marker gene, as the neo gene itself is not linked to a mammary gland specific promoter (Figure 1A). The PGK1 gene itself (from which the PGK promoter is derived) is expressed at similar levels in mammary gland and liver, at higher levels in the kidney, and only at low levels in the spleen (Figure S5E).

3.3 | Transgene protein expression is consistent in A1L transgenic mouse lines

Milk samples derived from 12 transgenic lines were assessed for A1L protein expression (Figure 3). Coomassie Blue staining of milk samples separated on a 10% polyacrylamide gel demonstrates that expression of the α-casein protein is visibly reduced in the milk of heterozygous [+/L] A1L transgenic animals and is undetectable in the milk of homozygous [L/L] A1L transgenic mice (Figure 3A). The concentration of the other casein species (β-casein, γ-casein and κ-casein) is also drastically reduced in the absence of α-casein [L/L]. This confirms the findings previously made in α-casein deficient mice.[21] Expression of the A1L protein is visible in stained protein gels (Figure 3A, L-chain) but is much lower than that of the α-casein protein. The A1L protein is visible in the milk of heterozygous A1L transgenic mice [+/L], however, in milk of homozygous A1L transgenic mice the A1L protein is obscured by the appearance of two proteolytic breakdown products of β-casein (which have been identified previously).[21] The abundance of the A1L protein was also assessed by Western blotting alongside a human IgG control protein of known concentration (Figure 3B). Analysis of milk proteins fractionated into whey and casein fractions clearly demonstrates that the A1L protein is found exclusively in the whey

first translational start site in the chimeric mRNA. Panel B: RT-PCR analysis of the mRNA expressed from the α-casein alleles in wildtype [+/+] heterozygous [+/L], and homozygous [L/L] A1L transgenic mice. RNA was isolated from lactating mammary gland tissue at lactation day 10. cDNA was PCR amplified using primer pairs specific for β-actin (b-actin-up/low), α-casein (acas9a/acas5) and the A1L transgene (acas9a/A1Lsp.2 and ATG2/VK1FOR). Panel C: qPCR analysis of A1L gene expression in wildtype [+/+] ($n = 3$), heterozygous [+/L] ($n = 6$), and homozygous [L/L] ($n = 3$) mice using the primer combination acas9a/A1Lsp.2. Panel D: qPCR analysis of α-casein gene expression using the primer combination acas4/acas5. Panel E: qPCR analysis of β-casein gene expression using the primer combination bcas21/bcas22. Error bars indicate standard deviation. For comparisons against wild-type mice in a one-way ANOVA $p < 0.01$ is indicated by **, $p < 0.001$ by ***.

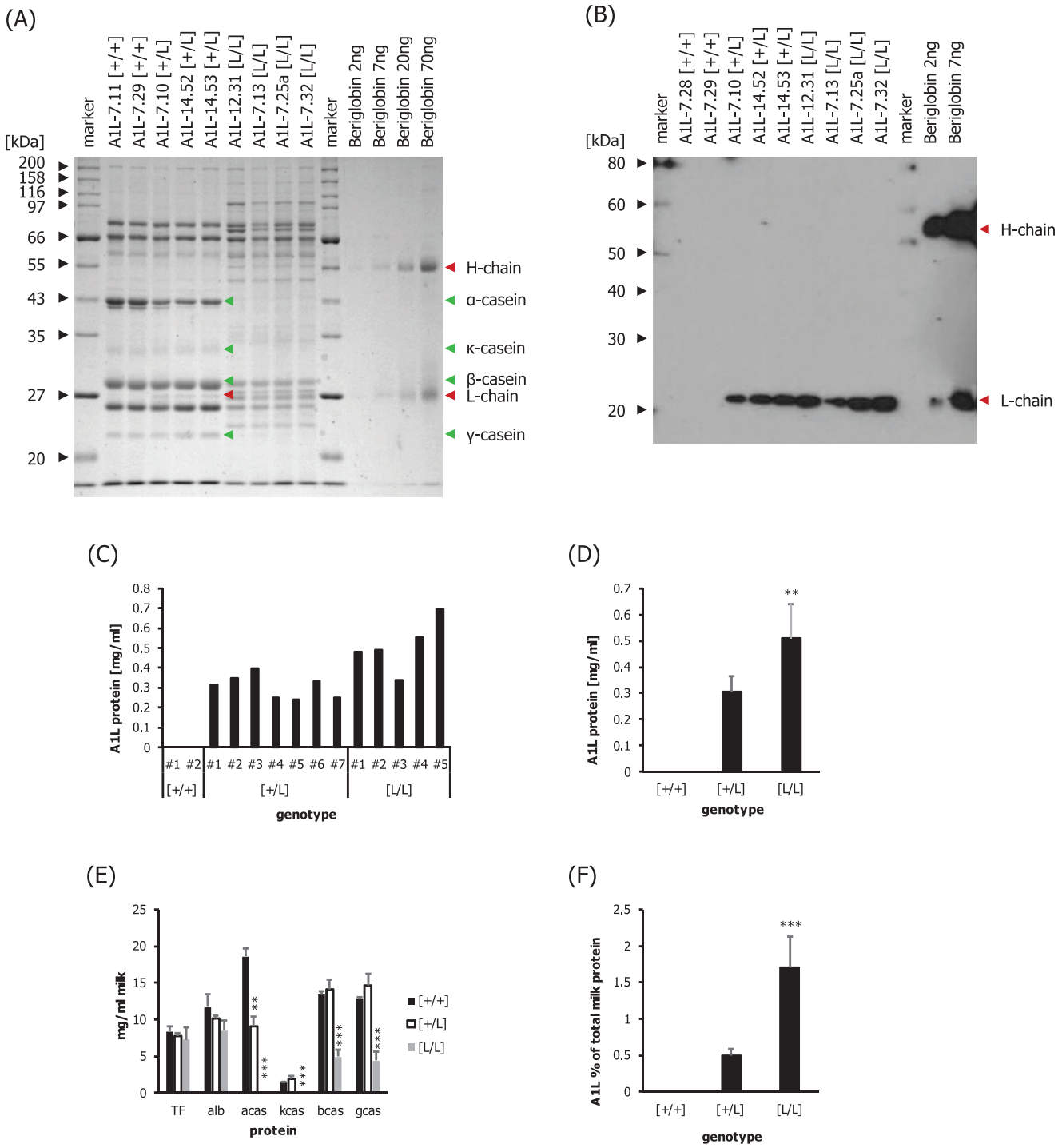

FIGURE 3 A1L transgene protein expression. Panel A: SDS-polyacrylamide gel analysis of milk derived from wild-type [+/+] ($n = 2$), heterozygous [+/L] ($n = 3$) and homozygous [L/L] ($n = 4$) α-casein gene modified mice. Defatted whole milk samples and human control IgG was separated on a 10% gel and stained with Coomassie Blue. The sizes of the protein molecular weight markers (New England Biolabs, broad range protein marker) the positions of the α-casein, κ-casein, β-casein, γ-casein (CSN1S2), IgG heavy chain and IgG light chain proteins are indicated. Panel B: Western blot analysis of the sample set shown in Panel A. Samples were separated on a 15% protein gel and blotted to nitrocellulose. Human IgG was detected using a horse-radish peroxidase-coupled goat-anti-human IgG antiserum (Jackson Laboratories) at a 1:5000 dilution. The positions of the IgG heavy and light chain proteins are indicated. Panel C: Quantification of the A1L protein in milk samples. TIF files of the Western blots were scanned using the KODAK 1d software. Signals from the IgG standard were used to develop a standard curve against which the signals from the milk samples were quantified. Protein concentrations in milk are shown as mg/mL. Panel D: Average expression of A1L IgG protein in the milk of wildtype [+/+] ($n = 5$), heterozygous [+/L] ($n = 7$) and homozygous [L/L] ($n = 5$) A1L transgenic mice. A1L concentrations in the milk of homozygous A1L transgenic mice [L/L] are significantly higher than in heterozygous [+/L] mice ($p < 0.01$, one-way ANOVA). Panel E: Concentrations of major milk proteins in the milk of wildtype [+/+] ($n = 2$), heterozygous [+/L] ($n = 3$) and homozygous [L/L] ($n = 4$) A1L transgenic

fraction (Figure S6). Quantification of the Western blots relative to a human IgG standard using densitometric scanning of Western blots demonstrates consistency of expression within the heterozygous (0.25 to 0.35 mg mL^{-1} of milk) and homozygous (0.5 to 0.7 mg mL^{-1} of milk) transgenic lines (Figure 3C). Average expression of the A1L protein in heterozygous mice is 0.3 and 0.55 mg mL^{-1} in homozygous mice (corresponding to 0.5% and 1.6% of total milk protein, respectively). A1L protein expression in homozygous mice is significantly increased relative to heterozygous mice ($p < 0.01$) (Figure 3D). Intriguingly, the increase in A1L protein abundance in the milk of homozygous [L/L] relative to heterozygous [+/L] A1L transgenic mice occurs while the A1L specific transcript appears to be reduced (Figure 2C). This finding is consistent with published data.[24] Densitometric scanning of Coomassie stained gels confirmed α-casein as the most abundant mouse milk protein at a concentration of 20 mg mL^{-1}.[21] This represents around 20% of total milk protein.[19] The expression and synthesis of casein proteins is significantly reduced in A1L homozygous [L/L] (and therefore α-casein deficient) mice (Figure 3E) due to ER stress. As a consequence, the concentration of A1L as a fraction of total milk protein is increased by around 3-fold in homozygous A1L transgenic mice relative to homozygous transgenic mice (Figure 3F).

3.4 | α-casein deficiency reduces total milk protein concentration

The development of pup weights in response to the insertion of the A1L transgene into the α-casein gene was analyzed (Figure 4A). Weight gain of cross-fostered wildtype offspring was not affected when nursed by heterozygous [+/L] dams relative to pups nursed by wildtype [+/+] dams (Figure 4A). In contrast, pup weight gain was significantly reduced in wildtype pups nursed by homozygous A1L transgenic dams [L/L] (Figure 4A,B). The delay in weight gain was similar to that observed in pups nursed by α-casein deficient dams (Figure 4C).[21] These data suggest that the expression of the recombinant A1L protein is unable to substitute for the loss of the α-casein protein. This conclusion is also supported by the fact that a significant upregulation of the endoplasmic reticulum stress proteins grp78 (BIP) and grp94 is observed in the lactating mammary gland of both, α-casein deficient mice [−/−] and homozygous A1L transgenic mice [L/L] (Figure 4D). Analysis of liver, kidney and brain weight in pups at day 21 of lactation (the point of weaning) demonstrates that liver weight (relative to total body weight) is significantly reduced ($p < 0.001$) in pups nursed by homozygous [L/L] mice, while brain weight is significantly increased ($p < 0.001$) in the same pups (Figure 4E,F). In contrast, relative kidney weight is not affected by the dam's milk composition. This indicates that a lack of protein supply after birth affects different organs differently.

3.5 | A1L transgene regulation in cell culture

Our previous results had indicated that a transgene which is inserted into the β-casein gene adopts the level of expression of the integration site.[14] To further analyze the regulation of the α-casein-A1L allele, the genetic modification of the transgenic animals was replicated in the mammary epithelial cell line HC11 using a combination of homologous recombination, RMCE and somatic cell fusion (Figure S7 and methods in the Supplementary materials). HC11 cells are tetraploid.[14] The integration of the A1L transgene is highly likely to have occurred only in one of the 4 available casein alleles. The targeting frequencies for the modified cell lines are shown in supplementary table S3.

Expression of the casein genes and the A1L transgene was analyzed using RT-PCR, qPCR and Western blotting in HC11 fusion cell pools in the presence and absence of lactogenic hormones. RT-PCR analysis demonstrates that both, the unmodified α-casein alleles and the transgene-containing α-casein allele are expressed only in response to lactogenic hormone treatment (Figure S8B), while no expression can be detected in the absence of lactogenic hormones. Expression of the α-casein, β-casein, γ-casein and A1L genes was strongly induced at the RNA level as quantified by qPCR (Figure S8C). Western blot analysis of cytoplasmic extracts derived from three HC11 fusion cell pools confirmed a strong induction of β-casein protein synthesis in response to lactogenic hormone treatment and a less pronounced activation of α-casein protein synthesis (Figure S9B,C). No A1L protein was detectable in the extracts derived from the HC11 fusion cell clones (Figure S9D). Collectively the results in HC11 cells demonstrate that, like in the transgenic animals, the A1L transgene is expressed at a lower level than the unmodified α-casein gene.

4 | DISCUSSION

The objective of the experimentation was to determine whether insertion of a transgene into a defined gene locus [a] leads to transgene expression at the level of the replaced gene, and [b] whether the transgene expression is consistent among transgenic lines.

The two objectives were addressed using an antibody transgene inserted site-specifically into the highly expressed α-casein gene. The results demonstrate that consistent expression of the transgene is indeed achieved. In heterozygous mice the expression levels of the recombinant protein (the light chain of the anti-viral antibody A1, A1L) vary between 0.25 and 0.45 mg mL^{-1} (a maximum variation of 80%). This contrasts with observations in a transgenic mouse expressing the same antibody gene under the control of the ovine β-lactoglobulin promoter[11] in which expression levels varied between 0.04 and 2.9 mg mL^{-1} (a 72-fold, or 7200% variation) (Figure S10). The

mice. Protein expression was assessed using densitometric scanning of Coomassie Blue stained protein gels relative to protein standards of known concentration. Panel F: Percentage of recombinant A1L protein relative to total milk protein. The values were calculated by comparing the Western blot derived A1L protein concentrations with the total protein amounts derived from the 6 main milk proteins analysed by Coomassie Blue protein staining. Error bars indicate standard deviation. For comparisons against wild-type mice in a one-way ANOVA $p < 0.05$ is indicated by *, $p < 0.01$ by **, $p < 0.001$ by ***.

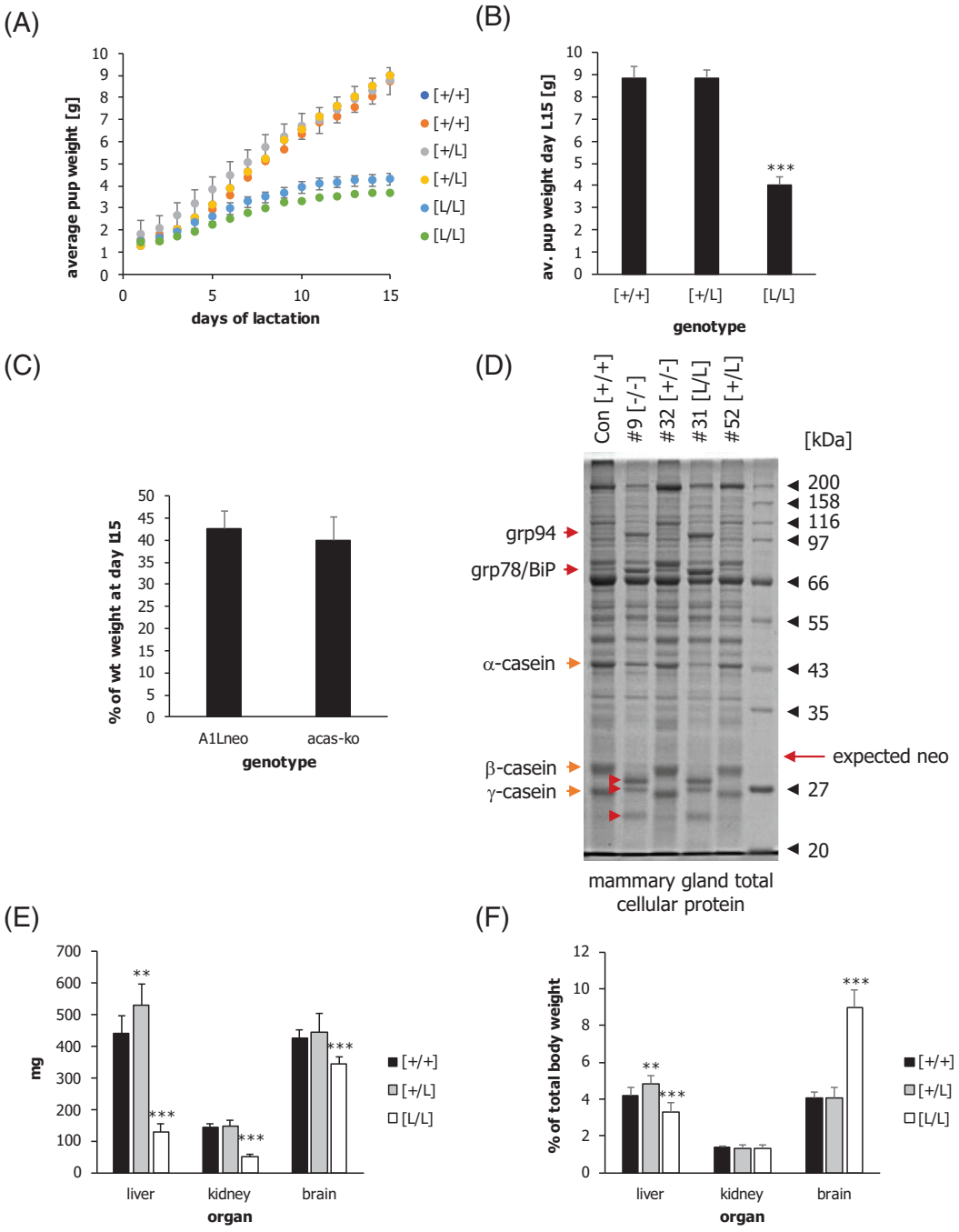

FIGURE 4 Weight development of pups (adjusted to 10 pups per litter) nursed by wildtype [+/+] ($n = 2$ litters), heterozygous [+/L] ($n = 2$) and homozygous [L/L] ($n = 2$) mice from day 1 to day 15 of lactation. All pups were wildtype C57B/6 pups cross fostered by transgenic or control dams on a CD1 background. Panel A: Average weight of pups (in grams) weighed every day from day 1 of lactation to day 15 of lactation. Panel B: Average pup weight on day 15 of lactation. Panel C: Percentage of weight of pups nursed by α-casein deficient dams relative to pups nursed by control dams in the experiments described in this paper and those observed in Kolb et al. 2011 (at day 15 of lactation). Panel D: Protein expression in lactating mammary tissue. Total protein extracts prepared from mammary tissue of wildtype [+/+], homozygous [−/−] and heterozygous [+/−] α-casein knock-out mice and homozygous [L/L] and heterozygous [L/+] A1L transgenic mice were separated on a 10% polyacrylamide gel and stained with Coomassie Blue. The positions of the α-casein, β-casein, γ-casein, grp78/BIP and grp94 proteins are indicated. The red arrowheads indicate the breakdown products of β-casein and γ-casein that are generated in the mammary gland of α-casein deficient mice.[21] Panel E: Weights of liver, kidney and brain in pups nursed by wildtype dams [+/+], heterozygous dams [+/L] and homozygous [L/L] A1L transgenic dams at day 21 of lactation. Panel F: Percentage of liver, kidney and brain weights relative to total body weight in pups nursed by wildtype [+/+], heterozygous [+/L] and homozygous [L/L] dams at day 21 of lactation. Significance values were determined by one-way ANOVA followed by a Bonferroni post-hoc test in Graph-Pad Prism: $^{*}p < 0.05$, $^{***}p < 0.001$.

β-lactoglobulin-A1L transgenic mice were generated by DNA microinjection into fertilized oocytes. The transgene is therefore inserted at random genomic sites with a wide variation in transgene copy number.[11] However, the expression rate per transgene copy is higher in the α-casein-A1L transgenic mice than in the β-lactoglobulin-A1L transgenic mice (Figure S10B), suggesting that the α-casein promoter is better at directing transgene expression than the b-lactoglobulin promoter.

However, the experiments in α-casein-A1L transgenic mice demonstrated that the A1L transgene was expressed at a lower level than the α-casein gene it had replaced. RNA levels of the transgene (as judged by qPCR) were found to be around 4% of the expression of the corresponding endogenous α-casein allele in heterozygous mice. This translates into a similar ratio of protein expression (a maximum concentration of 0.45 mg mL^{-1} A1L protein compared to 10 mg mL^{-1} α-casein protein for one allele, i.e., 4.5%; and 0.7 mg mL^{-1} A1L protein compared to 20 mg mL^{-1} for 2 alleles, i.e., 3.5%).

Numerous mechanisms can lead to the silencing of transgenes.[10] These include, but are not limited to, 1. the insertion of the transgene at an unfavorable genomic site which does not support high levels of gene expression, 2. the presence of transgene repeats which lead to epigenetic silencing, 3. the presence of excess genetic elements which may interfere with gene regulation (e.g. selectable marker genes and their promoter elements), or 4. aberrant splicing. In the context of transgenic animals the dominant effects of the transgene integration site and repeat sequences are well established.[25–28] The site-directed insertion of a single copy transgene of interest into the defined genomic location should overcome some of these limitations.[26] However, the results in this manuscript demonstrate that this approach does not inevitably lead to transgene expression levels matching those of the gene which was replaced.

Three other published experiments are relevant to the interpretation of the data obtained in the α-casein-A1L transgenic mice (Table S4). In all four experimental approaches a single copy transgene was inserted at a highly expressed milk protein gene locus.

Firstly, the insertion of a β-galactosidase reporter gene into the β-casein gene (together with a selectable marker gene) faithfully reproduces the expression and induction pattern of the β-casein gene in cell culture.[14] There are several differences between those experiments and the experiments described here. These include a different integration site (β-casein vs. α-casein), a different transgene (β-galactosidase vs. the A1L antibody transgene) and a different transgene origin (a bacterial cDNA vs. a mammalian genomic version of the gene). However, our data here demonstrate that the A1L transgene is correctly spliced and no unexpected splice variants are formed. Moreover, the immediate splice signals are unaffected by the genetic modification of the α-casein allele. However, the effects of sequences in introns on gene expression is complex[29] and the modifications of the α-casein gene may have impaired splicing efficiency of the modified allele.

Secondly, a recent publication by Smirnov and colleagues used a similar approach to insert the coding region of the human granulocyte macrophage stimulating factor (hGMCSF) into the mouse α-casein gene utilizing CRISPR-Cas9 stimulated homologous recombination.[24] The hGMCSF gene was inserted into the second exon (as in our experiments) and, as an alternative, into the entire genomic space of α-casein exons 1 through 33. The authors showed low mRNA expression for the insertion into the 2nd exon, but high mRNA expression (analogous to the endogenous α-casein gene) when the entire α-casein gene was replaced. However, the reported protein concentrations (0.01 mg mL^{-1} for the 2nd exon insertion and 0.5 mg mL^{-1} for the whole gene replacement approach) were similar to the protein expression reported here. The authors suggested that the low expression of the hGMCSF after insertion onto the 2nd exon is due to the absence of a polyadenylation signal associated with the transgene (the authors relied on the polyadenylation site of the endogenous α-casein gene). But the data reported here seem to contradict this explanation. The A1L transgene in our present study was equipped with its own polyadenylation signal and was successfully expressed in NS0, CRFK and HEK293 cells.[30,31] But its expression in transgenic animals did not match that of the endogenous α-casein gene.

Third, a site-directed insertion of a human lactoferrin transgene into the goat β-lactoglobulin gene using TALENs and nuclear transfer was also reported.[32] Protein expression levels between 2 and 3 mg mL^{-1} of human lactoferrin were achieved. This represents the majority of the whey protein typically present in goat milk (3 to 7 mg mL^{-1}).[33] The human lactoferrin transgene consisted of a cDNA with an associated polyadenylation signal. The insertion was targeted to the 1st exon which contains the β-lactoglobulin ATG codon. The modification also inserted a PGK-neo selection marker cassette into the β-lactoglobulin gene.

The available datasets allow some preliminary conclusions which can be verified in future experiments. 1. The presence of a selection marker cassette did not interfere with the expression of a transgene inserted into the goat β-lactoglobulin gene[32] or the mouse β-casein gene.[14] Furthermore, transgene expression from within the α-casein gene (reported by Smirnov and colleagues) was low despite the absence of the selection marker.[24] These data suggest that the presence of a selection marker gene is not a likely reason for the reduced A1L transgene expression from within the mouse α-casein gene. 2. The insertion of a transgene into an allele which contains multiple exons appears to be less conducive to transgene expression than integration into genes with fewer exons (β-casein: 9 exons, or β-lactoglobulin: 7 exons). The reduction of the number of exons by removing additional splice sites supports transgene expression[24] (Table S4). 3. Equipping the transgene with its own polyadenylation site (rather than utilizing the polyA site present in the genomic target site) is insufficient to generate high expression levels.

The data described in here demonstrate the advantages and disadvantages of using the α-casein gene as a site for transgene expression. 1. Mice with a homozygous deletion of the α-casein gene show a reduction in milk protein production (heterozygous mice are not affected by this limitation). Therefore α-casein might be less suitable for large scale protein production (mice carrying a similar modification of the β-casein gene are likely to provide a superior expression platform). 2. The large number of exons in the α-casein gene may be disadvantageous for transgene expression and splicing. 3. However, the reduction in overall

casein content of milk increases the relative concentration and hence the purity of the recombinant protein.

Overall, our data demonstrate that, 1. site-directed integration of the A1L transgene into the mouse α-casein gene leads to a consistent level of expression in different transgenic lines; 2. the inserted transgene does not replicate the expression levels of the gene it has replaced; 3. the co-inserted PGKneo selection marker gene adopts the regulation of the α-casein gene locus without being linked to an endogenous promoter; 4. the A1L transgenic mice faithfully copy the biological phenotype of the α-casein knock-out mice with regards to milk composition, mammary gland biology, and pup development.

AUTHOR CONTRIBUTIONS

Andreas Kolb designed and oversaw the study, carried out experimental work, and wrote the manuscript. Christopher Knowles, Claire Warren, Simon Lillico, Ailsa Carlisle, and Linda Petrie carried out experimental work. Simon Lillico and Bruce Whitelaw co-wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the BBSRC (Gene Technologies Underpinning Healthcare: #12599, CASE studentship BB/J012343/1), the Scottish Government (RESAS transition funding to A. Kolb), and the Genomia Trust Fund (protein expression in protein poor milk).

CONFLICT OF INTEREST STATEMENT

The authors declare no financial or commercial conflict of interest.

DATA AVAILABILITY STATEMENT

Data available in article supplementary material.

ORCID

Andreas F. Kolb https://orcid.org/0000-0001-9056-5504

REFERENCES

1. Hunter, P. (2019). The prospects for recombinant proteins from transgenic animals. *EMBO Reports*, *20*(8). https://doi.org/10.15252/EMBR.201948757
2. Wang, Y., Zhao, S., Bai, L., Fan, J., & Liu, E. (2013). Expression systems and species used for transgenic animal bioreactors. *BioMed Research International*, *2013*, 580463. https://doi.org/10.1155/2013/580463
3. Shepelev, M. V., Kalinichenko, S. V., Deykin, A. V., & Korobko, I. V. (2018). Production of recombinant proteins in the milk of transgenic animals: Current state and prospects. *Acta Naturae*, *10*(3), 40. https://doi.org/10.32607/20758251-2018-10-3-40-47
4. Mukae, T., Okumura, S., Watanobe, T., Yoshii, K., Tagami, T., & Oishi, I. (2021). Production of recombinant monoclonal antibodies in the egg white of gene-targeted transgenic chickens. *Genes*, *12*(1), 1–11. https://doi.org/10.3390/genes12010038
5. Whitelaw, C. B. A., Joshi, A., Kumar, S., Lillico, S. G., & Proudfoot, C. (2016). Genetically engineering milk. *Journal of Dairy Research*, *83*(1), 3–11. https://doi.org/10.1017/S0022029916000017
6. Tripathi, N. K., & Shrivastava, A. (2019). Recent developments in bioprocessing of recombinant proteins: Expression hosts and process development. *Frontiers in Bioengineering and Biotechnology*, *7*, 420. https://doi.org/10.3389/fbioe.2019.00420
7. Hay, E. A., Knowles, C., Kolb, A., & MacKenzie, A. (2017). Using the CRISPR/Cas9 system to understand neuropeptide biology and regulation. *Neuropeptides*, *64*, 19–25. https://doi.org/10.1016/j.npep.2016.11.010
8. Monzani, P. S., Adona, P. R., Ohashi, O. M., Meirelles, F. V., & Wheeler, M. B. (2016). Transgenic bovine as bioreactors: Challenges and perspectives. *Bioengineered*, *7*(3), 123–131. https://doi.org/10.1080/21655979.2016.1171429
9. Yang, L., Church, G., Zhao, H. Y., Huang, L., Gao, Y., Wei, H. J., & Yang, G. (2021). Porcine germline genome engineering. *Proceedings of the National Academy of Sciences USA*, *118*(22). https://doi.org/10.1073/pnas.2004836117
10. Cabrera, A., Edelstein, H. I., Glykofrydis, F., Love, K. S., Palacios, S., Tycko, J., Zhang, M., Lensch, S., Shields, C. E., Livingston, M., Weiss, R., Zhao, H., Haynes, K. A., Morsut, L., Chen, Y. Y., Khalil, A. S., Wong, W. W., Collins, J. J., Rosser, S. J., … Deans, T. L. (2022). The sound of silence: Transgene silencing in mammalian cell engineering. *Cell Systems*. *13*(12), 950–973. https://doi.org/10.1016/j.cels.2022.11.005
11. Kolb, A. F., Pewe, L., Webster, J., Perlman, S., Whitelaw, C. B. A., & Siddell, S. G. (2001). Virus-neutralizing monoclonal antibody expressed in milk of transgenic mice provides full protection against virus-induced encephalitis. *Journal of Virology*, *75*(6), 2803–2809. https://doi.org/10.1128/JVI.75.6.2803-2809.2001
12. Hua, R., Liu, J., Li, Y., Fan, Y., Zeng, B., Geng, G., & Li, Q. (2021). Novel functional recombinant human follicle-stimulating hormone acquired from goat milk. *Journal of Agricultural and Food Chemistry*, *69*(9), 2793–2804. https://doi.org/10.1021/acs.jafc.0c07208
13. Sorrell, D. A., & Kolb, A. F. (2005). Targeted modification of mammalian genomes. *Biotechnology Advances*, *23*(7-8), 431–469. https://doi.org/10.1016/j.biotechadv.2005.03.003
14. Robinson, C., & Kolb, A. F. (2009). Analysis of mammary specific gene locus regulation in differentiated cells derived by somatic cell fusion. *Experimental Cell Research*, *315*(3), 508–522. https://doi.org/10.1016/j.yexcr.2008.10.029
15. Vilotte, J. L., Chanat, E., Le Provost, F., Whitelaw, C. B. A., Kolb, A. F., & Shennan, D. B. (2013). Genetics and biosynthesis of milk proteins. *In advanced dairy chemistry*, 1A, 431–461. Springer US. https://doi.org/10.1007/978-1-4614-4714-6_14
16. Davies, D. T., & Law, A. J. R. (1980). The content and composition of protein in creamery milks in south-west Scotland. *Journal of Dairy Research*, *47*(1), 83–90. https://doi.org/10.1017/S0022029900020902
17. Brodhagen, J., Weikard, R., Thom, U., Heimes, A., Günther, J., Hadlich, F., Zerbe, H., Petzl, W., Meyerholz, M. M., Hoedemaker, M., Schuberth, H. J., Engelmann, S., & Kühn, C. (2019). Development and evaluation of a milk protein transcript depletion method for differential transcriptome analysis in mammary gland tissue. *BMC Genomics*, *20*(1), 1–19. https://doi.org/10.1186/S12864-019-5781-3/TABLES/8
18. Holt, C., Carver, J. A., Ecroyd, H., & Thorn, D. C. (2013). Invited review: Caseins and the casein micelle: Their biological functions, structures, and behavior in foods. *Journal of Dairy Science*, *96*(10), 6127–6146. https://doi.org/10.3168/jds.2013-6831
19. Kumar, S., Clarke, A. R., Hooper, M. L., Horne, D. S., Law, A. J., Leaver, J., Springbett, A., Stevenson, E., & Simons, J. P. (1994). Milk composition and lactation of beta-casein-deficient mice. *Proceedings of the National Academy of Sciences USA*, *91*(13), 6138–6142. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=44153&tool=pmcentrez&rendertype=abstract
20. Shekar, P. C., Goel, S., Rani, S. D. S., Sarathi, D. P., Alex, J. L., Singh, S., & Kumar, S. (2006). κ-Casein-deficient mice fail to lactate. *Proceedings of the National Academy of Sciences USA*, *103*(21), 8000–8005. https://doi.org/10.1073/pnas.0601611103
21. Kolb, A. F., Huber, R. C., Lillico, S. G., Carlisle, A., Robinson, C. J., Neil, C., Petrie, L., Sorensen, D. B., Olsson, I. A. S., & Whitelaw, C. B. A.

(2011). Milk lacking α-casein leads to permanent reduction in body size in mice. *PLoS One*, *6*(7), e21775. https://doi.org/10.1371/journal.pone.0021775

22. Huber, R. C., Kolb, A. F., Lillico, S. G., Carlisle, A., Sandøe, P., Sørensen, D. B., Remuge, L., Whitelaw, C. B. A., & Olsson, I. A. (2013). Behaviour of postnatally growth-impaired mice during malnutrition and after partial weight recovery. *Nutritional Neuroscience*, *16*(3), 125–134. https://doi.org/10.1179/1476830512Y.0000000038
23. Chanat, E., Martin, P., & Ollivier-Bousquet, M. (1999). αS1-casein is required for the efficient transport of β- and κ-casein from the endoplasmic reticulum to the Golgi apparatus of mammary epithelial cells. *Journal of Cell Science*, *112*(19), 3399–3412. https://doi.org/10.1242/JCS.112.19.3399
24. Smirnov, A. V, Kontsevaya, G. V, Shnaider, T. A., Yunusova, A. M., Feofanova, N. A., Gerlinskaya, L. A., Serova, I. A., Serov, O. L., & Battulin, N. R. (2022). Evaluation of the α-casein (CSN1S1) locus as a potential target for a site-specific transgene integration. *Scientific Reports*, *12*(1), 7983. https://doi.org/10.1038/s41598-022-12071-1
25. Clark, A. J., Bissinger, P., Bullock, D. W., Damak, S., Wallace, R., Whitelaw, C. B. A., & Yull, F. (1994). Chromosomal position effects and the modulation of transgene expression. *Reproduction, Fertility and Development*, *6*(5), 589–594. https://doi.org/10.1071/RD9940589
26. Strathdee, D., Ibbotson, H., & Grant, S. G. N. N. (2006). Expression of transgenes targeted to the Gt(ROSA)26Sor locus is orientation dependent. *PLoS One*, *1*(1), e4. https://doi.org/10.1371/journal.pone.0000004
27. Garrick, D., Fiering, S., Martin, D. I., & Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. *Nature Genetics*, *18*(1), 56–59.
28. Mitsuda, S. H., & Shimizu, N. (2016). Epigenetic repeat-induced gene silencing in the chromosomal and extrachromosomal contexts in human cells. *PLoS One*, *11*(8), e0161288. https://doi.org/10.1371/journal.pone.0161288
29. Shaul, O. (2017). How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology*, *91*(Pt B), 145–155. https://doi.org/10.1016/j.biocel.2017.06.016
30. Kolb, A. F., & Siddell, S. G. (1997). Expression of a recombinant monoclonal antibody from a bicistronic mRNA. *Hybridoma*, *16*(5), 421–426. http://www.ncbi.nlm.nih.gov/pubmed/9388025
31. Kolb, A. F., & Whitelaw, C. B. A. (2006). A virus-neutralising antibody is not cytotoxic in vitro. *Molecular Immunology*, *43*(6), 677–689. https://doi.org/10.1016/j.molimm.2005.04.008
32. Cui, C., Song, Y., Liu, J., Ge, H., Li, Q., Huang, H., Hu, L., Zhu, H., Jin, Y., & Zhang, Y. (2015). Gene targeting by TALEN-induced homologous recombination in goats directs production of β-lactoglobulin-free, high-human lactoferrin milk. *Scientific Reports*, *5*(1), 1–11. https://doi.org/10.1038/srep10482
33. Roy, D., Ye, A., Moughan, P. J., & Singh, H. (2020). Composition, structure, and digestive dynamics of milk from different species – a review. *Frontiers in Nutrition*, *7*, 195. https://doi.org/10.3389/FNUT.2020.577759/BIBTEX

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Knowles, C., Petrie, L., Warren, C., Lillico, S. G., Carlisle, A., Whitelaw, C. B. A., & Kolb, A. F. (2024). Site specific insertion of a transgene into the murine α-casein (CSN1S1) gene results in the predictable expression of a recombinant protein in milk. *Biotechnology Journal*, *19*, e2300287. https://doi.org/10.1002/biot.202300287