Defining a haplotype encompassing the *LCORL-NCAPG* locus associated with increased lean growth in beef cattle

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Leif Ellis Majeres
May 2024

Acknowledgements

My warmest thanks to everyone who contributed to this project, in ways big and small. Sonia Moísa and Taylor Seay, for being mentors and helping me out with lab work; Jana Owen and Grayson Whitlock, for your time and contributions in the lab; Trevor Freeman, Maddie Henniger, and Cassidy Catrett for your assistance with code and analysis; my committee members Dr. Troy Rowan and Dr. Brynn Voy for their support and insight on this project; and to Trevor Carlsen, for listening to all the thoughts in my head, assisting with data organization and presentation, and just being there, always. All of you helped make this project possible, and I can't express my gratitude enough. This isn't just a piece of paper to me, this is a labor of love and something I truly care about, so again, to everyone who made this possible, thank you.

Beyond this project, I want to acknowledge the other animal science graduate students, particularly my officemates and the Animal Science Graduate Student Association as a whole. Thank you for providing a sense of community and camaraderie; graduate school is a struggle and I am glad we don't have to bear it alone. Beyond my life as a graduate student, I also mus0t thank my dearest friends, Étienne Tetreault and Trevor Di Fine. You've made my life lighter in so many ways, and I genuinely do not know if it would have been possible for me to be here without your support. And of course, this wouldn't be complete without thanking my mom for her support and unconditional love throughout, well, my entire life.

Finally, I have to thank my mentor, Dr. Beever. I feel so fortunate that you were willing to take me under your wing and have me here. Even through all the struggles that came with getting my degree thus far and the many more that are to come for as long as I do science as a career, I truly feel that this is the best thing that could have happened to me. So thank you for this opportunity, thank you for just letting me be here, and thank you for being a good mentor for me. Thank you for your patience and your willingness to listen and to treat me as a person. It feels like a rare thing out in the world. I feel so lucky to know you and continue to work with you.

Abstract

Numerous studies have shown genetic variation at the *LCORL-NCAPG* locus is strongly associated with growth traits in beef cattle. However, a causative molecular variant has yet to be identified. To define all possible candidate variants, 34 Charolais-sired calves were whole genome sequenced including 17 homozygous for a long-range haplotype associated with increased growth (*QQ*), and 17 homozygous for potential ancestral haplotypes for this region (*qq*). The *Q* haplotype was refined to an 814-kb region between chr6:37,199,897-38,014,080 and contained 218 variants not found in *qq* individuals. These variants include an insertion in an intron of *NCAPG*, a previously documented mutation in *NCAPG* (rs109570900), two coding sequence mutations in *LCORL* (rs109696064 and rs384548488), and 15 variants located within ATAC peaks that were predicted to affect transcription factor binding. Notably, rs384548488 is a frameshift variant likely resulting in loss of function for long isoforms of *LCORL*. To test the association of the coding sequence variants of *LCORL* with phenotype, 405 cattle from five populations were genotyped. The two variants were in complete linkage disequilibrium. Statistical analysis of the three populations that contained *QQ* animals revealed significant ($p < 0.05$) associations with genotype and birth weight, live weight, carcass weight, hip height, and average daily gain. These findings affirm the link between this locus and growth in beef cattle and describe DNA variants that define the haplotype. However, further studies will be required to define the true causative mutation.

Table of Contents

Chapter 1: Literature Review....1
Introduction....2
The *LCORL-NCAPG* locus....3
The function of *LCORL*....8
Documented impacts of *LCORL* mutations and modulations....10
Chapter 2: Defining a haplotype encompassing the *LCORL-NCAPG* locus associated with increased lean growth in beef cattle....16
Abstract....17
Introduction....18
Materials and Methods....20
Sample collection....20
Identification of haplotype-defining mutations....22
Individual variant genotyping....23
Statistical analyses....24
Results....26
Defining mutations exclusive to the *Q* haplotype....26
Discovery of structural variants....32
Genotype-phenotype relationship....35
Discussion....38
Conclusions....42
List of References....45
Vita....53

List of Tables

Table 2.1 The populations of cattle used for this project 21
Table 2.2 Primers, enzymes, and expected fragment sizes for the PCR-RFLP genotyping assay 25
Table 2.3 Primers and probes for the 5'-3' exonuclease genotyping assay 25
Table 2.4 The 22 variants detected in coding sequence or ATAC peaks 29
Table 2.5 Consensus ATAC peaks with variants, associated tissues, and signal scores 30
Table 2.6 Transcription factors predicted to bind to the ATAC regions containing variants 31
Table 2.7 Genotype frequencies by population 36
Table 2.8 Phenotype and the statistical significance of genotype in the association model 36
Table 2.9 Statistical model for each phenotype 37

Chapter 1

Literature Review

Introduction

Beef is a widely demanded animal product, being the third most consumed meat globally [1]. In 2020 alone, over 70 megatons (155 billion pounds) of beef were produced and consumed worldwide [1]. However, this demand comes with considerable environmental costs. It has been estimated that for every kilogram of beef consumed in the United States, the equivalent of 48.4 kilograms of CO_2 is emitted, mostly in the form of CH_4 from enteric fermentation and N_2O released from manure [2]. The beef industry's water footprint is also comparatively high, at an average of 15,415 cubic meters of water per ton of beef produced, over three times that of chicken [3]. In terms of land use, beef demands 30 m^2 of land per kg of meat produced, three times that of pork or chicken, though this comes with the important caveat that cattle are able to digest less nutrient-dense feed and thus use lower quality land, which may otherwise be unsuitable for agriculture [4].

Yet despite the burden placed on the environment, demand for beef is continuing to grow, driven by increases in global wealth and population [1, 5]. To keep pace, livestock production must grow sustainably, prioritizing increasing efficiency and decreasing the environmental impacts of the industry. Genetic improvement holds a great deal of potential in this regard, and livestock species have already demonstrated its merits. In the past 50 years, dairy cattle have more than doubled in milk productivity, with half of that change being attributable to genetic selection [6]. Similarly, modern broiler chickens grow at a rate 400% faster than typical for birds 50 years ago, even under otherwise identical conditions [7]. Beef cattle production has improved at a steady pace as well, with the average age at harvest falling from 609 days in 1977 down to 485 in 2007 [8]. However, with the aforementioned environmental costs and growing demand, continued improvement in efficiency is imperative.

Based on the above, selection for improved production traits certainly has merit, and thus, many studies have aimed to identify quantitative trait loci (QTL) with the potential for

selection. For meat animals, body size tends to be commonly investigated, due to its association with growth and the ultimate product, the carcass. Cattle that can reach a target slaughter weight faster spend less time on feed, saving producers money and potentially reducing greenhouse gas emissions and waste over the animal's lifespan. There is a balance between selection for overall body size and against high birth weight, due to the increased risk of dystocia and calf mortality posed by larger calves [9]. Generally, birth weight and weaning weight are highly correlated traits, having an estimated genetic correlation of around 0.6 [10]. However, the existence of 'curve bender' sires, whose calves are born with below average birth weight, but grow to above average weaning weights and beyond, challenges the idea that selection for increased growth and body size inherently poses a greater risk of dystocia. Despite the aforementioned caveat, growth and size are certainly desirable traits to select upon and are among the most commonly recorded traits within the beef industry. Thus, many of the early efforts to identify QTL involved these traits. Over the past two decades, these studies have consistently identified several major QTL, most prominently among them being *PLAG1* and the *LCORL-NCAPG* locus [11-17].

The *LCORL-NCAPG* locus

The *LCORL-NCAPG* locus in particular has been associated with stature in cattle since at least 2000 [18], and subsequent studies have linked this region on bovine chromosome 6 (BTA6) with numerous important lean growth traits for beef cattle, including increased birth weight [15, 17, 19-29], body weight throughout life [15-17, 19, 28-32], average daily gain (ADG) [16, 20, 33-35], feed intake [16, 33], carcass weight [12, 14, 19, 31-33, 36-38], bone weight [23, 24, 37, 39, 40], and ribeye area [19, 33] as well as reduced adiposity [14, 23, 25, 30, 33]. While the causative mutation behind this change in phenotype has remained elusive, it is thought to have been narrowed to a 591-kb region located on bovine chromosome 6 (BTA6), including

LCORL, *NCAPG*, *DCAF16*, and *FAM184B*, as well as a distal area lacking any annotated features [38]. This region is highly conserved among mammals [41, 42], and its association with body size and growth velocity has been observed in many species, including humans [43-54], dogs [42, 55-57], horses [58-63], sheep [41, 64, 65], goats [66-68], pigs [69], chickens [70-73], rabbits [74], and reindeer [75].

Human studies have recognized height to be a trait that is both highly heritable [76] and highly polygenic, controlled by thousands of variants, most with a miniscule effect on phenotype. It has been recognized from these studies that variants not reaching genome-wide significance are needed to explain the gap between the calculated heritability of height compared to the variance explained in height GWAS. To explain just 29% of phenotypic variation in human height, Wood et al. [53] found they needed 9,500 SNPs. The *LCORL-NCAPG* locus was among these, with an allele observed in their study having the eighth largest effect size and second most statistically significant effect. Their reported effect size was 0.101 standard deviations above mean when adjusted for sex and age, which would equate to roughly 6 to 8 mm in mature height.

While in humans, it takes thousands of small variants to explain a fraction of the variance for height, the story is very different in domesticated animals. In dogs, 14 markers can explain 95% of the variation in weight across dog breeds [42]. Horses, too, have had an estimated 83% of their body size variation attributed to only four loci [58]. Likewise in cattle, one study in the Japanese Black breed estimated that about a third of variance for carcass weight could be explained by three loci [12]. However, in other breeds of cattle, the genetic architecture of stature appears more complex. A within-breed analysis in 17 different populations in cattle found that the top 163 variants could only explain 13.8% of the variance for stature phenotype, seeming more like height variation in humans [13]. Irrespective of the population structure, the *LCORL-NCAPG* locus stands tall as one of the most prominent regions controlling stature,

appearing in the top lists for all of the aforementioned studies and many others. Its effect size, even in populations where the individual effects of loci are comparatively small, is substantial enough to be consistently observed.

Given this region's large impact on such an important production trait, there historically must have been selective pressure on this region. Evidence of selective sweeps, where strong selection for a novel variant results in regions of high LD, can be observed at the *LCORL-NCAPG* locus in the domesticated species discussed. Its strong effect on body size has resulted in both positive and negative selection, depending on the breeding objective of the population under selection [13, 42, 58, 68, 69, 74, 77, 78]. This phenomenon is clearly present in cattle as well; the 600-kb region surrounding *LCORL-NCAPG* has been shown to be practically fixed for the same haplotype in Brown Swiss, Gelbvieh, and Montbeliard cattle. The same study also found this haplotype to be very common in Charolais and Fleckvieh cattle, though not to the same degree of fixation [13]. Many studies observe this haplotype segregating in Charolais populations [13, 22, 24, 25, 27, 31, 79], but it has been observed in other breeds, including Japanese Black and Brown [12, 32, 38, 80], Simmental [25, 34, 37, 39], and curiously, American Black and Red Angus populations [17, 81], despite this haplotype's apparent absence in Irish and Australian Black Angus populations [13]. In studies where this haplotype is found, it is consistently associated with the higher growth phenotype. For the remainder of this review, we will refer to this derived haplotype as '*Q*', and to ancestral chromosomes as '*q*'.

The LD surrounding this locus has long confounded attempts to identify the true causative mutation. The close proximity of genes in this region has resulted in studies at different times pointing to *LCORL* and *NCAPG* as the gene mediating these changes. Setoguchi et al. [38] attempted to refine the location of this QTL in 2009, using microsatellite markers to identify an identical-by-state region in half-sib families of Japanese Brown and Black cattle correlated with increased carcass weight. This original identical-by-state region spanned an

815-kb area, including not only *LCORL*, *NCAPG*, *DCAF16*, and *FAM184B*, but also *MED28* and *LAP3.* Resequencing of this region identified SNPs that were used for later genotyping of individuals. This resulted in the discovery of the shared 591-kb haplotype noted previously, ruling out *MED28* and *LAP3*. The most strongly correlated SNP with phenotype was *NCAPGc.1326T>G*, a missense mutation found in exon 9 of *NCAPG* resulting in the substitution of Met-442 for Ile. While it was acknowledged that the region was in significant LD and that the three neighboring genes could not be ruled out, these investigators considered *NCAPG* to be the most probable candidate gene underlying the phenotype due to this mutation.

NCAPG (non-SMC condensin I complex subunit G), encodes CAP-G, a subunit of the condensin I protein complex [82]. This condensin complex plays a key role in chromosome condensation and segregation during cell division. The CAP-G subunit in particular has been has been shown to aid in localizing the complex to chromosomal DNA during mitosis [83]. It has been hypothesized that based on the apparent role of *NCAPG* in cell division, that perhaps its activity may regulate cellular proliferation and growth in some way [38].

Eberlein et al. [22] followed up on the discovery of the *NCAPGc.1326T>G* mutation by testing if there was a link between genotypes at this locus, *NCAPG* expression, and phenotype. This study utilized multiple ovulation and embryo transfer to produce Charolais x German Holstein F2 calves for phenotypic data collection, which allowed fetal and maternal genetic factors contributing to embryonic growth to be analyzed independently. They genotyped 139 calves for 23 SNPs in a 10-Mb interval, including the *NCAPGc.1326T>G* variant. Analyses of F2 genotypes with phenotypes found *NCAPGc.1326T>G* to be the marker most significantly associated with fetal growth. Expression levels of *NCAPG* mRNA were quantified, demonstrating ubiquitous *NCAPG* expression across bovine fetal tissue types, with suggestive (p=0.09) evidence that animals bearing a *NCAPGc.1326G* allele had reduced *NCAPG* expression. While this study did demonstrate that the effect of this allele is a result of fetal

genetics alone, not maternal genetics, there is no definitive proof whether *NCAPG* or *LCORL* is the causative gene based on the evidence presented. The refined haplotype of “at least 193-kb” that they identified still encompassed both genes and may have extended beyond them as well.

However, other investigations into this locus suggest the narrative is incomplete when only considering the *NCAPGc.1326T>G* SNP as the causative mutation. Gutiérrez-Gil et al. attempted to refine the location of this QTL with a Charolais x Holstein population in the UK, sharing their choice of breed composition and breeding scheme with the population used by Eberlein et al. in Germany. Briefly, Charolais sires were crossed with Holstein dams to create an F1 generation. The F1 population was then intercrossed to create an F2 population, and with the parent breeds to create Charolais and Holstein backcross populations. As with the previous studies, the *NCAPG-LCORL* QTL was significantly associated with growth. The F1 sires of this population should be heterozygous (*Qq)*; receiving the *Q* haplotype that should contain the true causative mutation from their Charolais sire, and the *q* haplotypes from their Holstein dam. However, among the genotyped F1 sires, the *NCAPGc.1326G* allele was not present on every chromosome bearing the associated *Q* haplotype, and in fact, some sires were homozygous *NCAPGc.1326G*. These findings directly conflict with the conclusions of the previous two studies, that suggested the *NCAPGc.1326G* allele was the causative mutation within the *Q* haplotype [24]. Gutiérrez-Gil’s finding is further supported by previous QTL analyses conducted by our lab that observed similar incongruity with the putative causative mutation and *q* haplotype among 82 Simmental and Simmental-Angus sires. It would be expected that animals heterozygous for the causative haplotype (*Qq*) or homozygous for the wild-type, reduced growth haplotype (*qq*) would correspondingly be heterozygous for *NCAPGc.1326G/T* or homozygous *NCAPGc.1326T/T* respectively. Yet, several *Qq* animals were instead observed to be homozygous for *NCAPGc.1326G*, and several *qq* sires were also found to be heterozygous [25].

We suggest that the gene whose action may be responsible for the change in phenotype at this locus is *LCORL*. Several studies, including the data from our lab, have pointed to the region upstream of *LCORL* as bearing the most significantly associated SNPs with this phenotype. While not definitive due to the high LD in this region, these findings suggest changes in the regulation of *LCORL* could be causing increased growth [17, 25, 27, 31].

The function of *LCORL*

Despite the large body of evidence for the substantial influence on growth associated with this locus, the function of *LCORL* itself is not well understood. In cattle and other species, *LCORL* encodes several transcript isoforms. The exact number and composition of these isoforms varies from species to species, often dictated by the inclusion or exclusion of small cassette exons in the 5' end of transcripts. Broadly speaking, these isoforms can be classified into two major types defined by their alternate 3' exon that contributes to the majority of the protein body, including the known functional domains. In this review, these isoforms will be referred to as 'long LCORL' and 'short LCORL'. These major protein types of *LCORL* have been documented not only in the mammalian species previously discussed, but across all vertebrate taxa, including fish such as the zebrafish and blackstripe livebearer. [84].

'Long LCORL' isoforms are composed of 1750-1900 amino acids and contains an annotated domain of unknown function – *DUF4553* [84]. However, the function of 'long LCORL' may not be as mysterious as the name of its domain implies. A recent study conducted by Conway et al. [85] set about describing two previously uncharacterized accessory proteins associated with polycomb repressive complex 2 (PRC2). PRC2 is a crucial piece of cellular machinery for multicellular development. This protein complex acts as a catalyst for mono-, di-, and tri- methylation of the 27th lysine of histone H3 (H3K27), altering the chromatin architecture and resulting in the silencing of genes in the associated region [86]. H3K27me3 is able to recruit

PRC2 during DNA replication [87], making PRC2 essential for both formation and maintenance of cellular identity. This function of PRC2 is critical for early body plan development, with disruptions of PRC2 activity resulting in early embryonic mortality in mice and improper segmentation in *Drosophila*. This complex has been documented as an important controller of the expression of *Hox* and other developmental regulation genes [88]. It is composed of four core proteins and several accessory proteins which are responsible for differentiating the protein complex into two main types: PRC2.1 and PRC2.2 [86]. Conway et al. investigated an accessory protein of PRC2.1 encoded by *LCORL*'s paralog, *LCOR*, which they named "PRC2 associated LCOR isoform 1" (PALI1). The bulk of this study is devoted to characterizing this protein, however, they also discovered PALI2, the protein product of 'long LCORL', which shares high sequence homology with PALI1. Within, but not spanning the entirety of the annotated *DUF4553* found in both long isoforms of *LCORL* and *LCOR*, they identified a PALI interaction with PRC2 (PIP) domain, which enabled these proteins to interface with the repressive complex. Experiments with PALI1 demonstrated that PRC2.1 exhibited higher repressive activity compared to PRC2.2, a complex bound to a competing accessory protein, Aebp2. Alterations in PRC2 activity caused by these accessory proteins resulted in disturbances in key genes involved in embryonic development and tissue patterning, including *Wnt3, Msx3,* and *FoxA2*, among others. Mouse embryos were generated with the PALI1-specific exon knocked out, but an otherwise normal copy of *LCOR*. These embryos were viable through embryonic day 11.5 without skeletal abnormalities, which was hypothesized to be due to continued function of PALI2 offering some redundancy. However, none of these embryos were viable at birth, suggesting the balance of PRC2.1/PRC2.2 is essential for later embryonic development. While it is theorized that PALI2 functions similarly to PALI1 due to the shared PIP domain and demonstrated affinity for PRC2, this study did not further investigate the activity of PALI2 [85]. The fact that PALI2 possesses AT hooks, which PALI1 lacks, suggests PALI2 may

be able to interact with DNA and perhaps direct PRC2 to particular targets, however, as of 2024, no functional data on PALI2 has been published.

'Short LCORL' is around 550 to 650 amino acids long and is defined by its helix-turn-helix (HTH) domain encoded by its last exon. HTH domains are commonly found in transcription factors, as these structures allow the protein to interact with the cell's DNA [89]. It is thought that LCORL may be a transcription factor itself, though its potential activities have yet to be characterized [82]. In 2021, Sakaguchi et al. observed that this isoform of LCORL can form a complex with CHD4 and NOL4L which may be involved in Notch pathway suppression [90], suggesting a potential role in the regulation of early development and cell fate [91]. Notably, CHD4 is known for its role as a component of NuRD (nucleosome remodeling and deacetylase activity) complex, where it is responsible for ATP-dependent chromatin remodeling [92], though it also is thought to have its own NuRD-independent activities as well [93]. It is currently unclear whether LCORL's interaction with CHD4 and NOL4L is an independent complex or a component that may modulate NuRD in some fashion.

Documented impacts of *LCORL* mutations and modulations

Examples in other species make a compelling case for 'long LCORL' as a regulator of growth. A study published in 2019 by Plassais et al. investigated genetic variants impacting several phenotypes in dogs. *LCORL*'s association with body weight and height was again observed in a genome-wide association study utilizing 76.5 million variants identified in their working dataset of 268 dogs representing 144 breeds, as well as 54 wild canids. Their whole genome sequencing data revealed a single nucleotide insertion exclusive to medium- and large-sized breeds of dogs in the last exon of the long isoform of *LCORL*. This frameshift mutation is predicted to result in 11 amino acids substitutions and an early stop codon that effectively deletes the last third of the protein, including *DUF4553*, most likely causing a loss of function for

this protein encoded by *LCORL*, while not interfering with other isoforms [42]. Though not noted by Plassais et al., this deleted region coincides with the PIP domain described by Conway et al. It is possible these dogs are effectively PALI2 knock-outs. While Conway's findings showed PALI1 to be essential for late embryonic development, PALI2 knock-outs may be viable, at least in dogs. As established by Conway et al., decreases in PRC2.1 activity results in reduced repression, or in other words, increased expression of certain genes relevant for embryonic development and body plan formation. It is possible this change in regulation is responsible for the increased body size observed, with the continued function of PALI1 allowing these dogs to remain viable.

A similar mutation altering 'long LCORL' has been observed in goats. While investigating selective sweeps associated with body size in Pakistani goats, Saif et al. identified a single nucleotide insertion to the long isoform specific exon of *LCORL*, *Ser277Ilefs*38*. This mutation results in a frameshift and early truncation, resulting in a loss of 85% of the protein, and likely a loss of function similar to what was observed in dogs [68]. Curiously, while there is data suggestive that similar truncating variants may be causative in Swiss goat breeds, this may not be the case for Boer goats. As intense selection and distinct haplotypes around this locus are present in Boer goats, it suggests there may be other population-specific alterations to *LCORL* affecting size phenotype [66].

Notably, 'long LCORL' is absent from the broiler chicken genome assembly's annotation (b.GalGal1.mat.broiler.GRCg7b), despite its presence in other birds, including the layer chicken genome (b.GalGal1.pat.whiteleghornlayer.GRCg7w). The possibility exists that mutations in the broiler genome resulted in a failure for annotation software to recognize this splice isoform. If this is the case, it also is possible that this change is reflected biologically as well, and that commercial broilers selected for rapid lean growth may also have a loss of function of 'long LCORL'.

Setting aside loss of function of 'long LCORL', there is evidence in horses that changes in *LCORL* expression could be impacting body size. Many studies have identified the SNP *BIEC2-808543* as being highly significant for height in horses [59, 63]. Though more recent studies have suggested it is likely not the causative SNP [61], it did link the region upstream of *LCORL* with body size in horses, suggesting a change in the regulation of *LCORL* expression, rather than a mutation in the protein outright. Metzger et al. utilized qPCR to evaluate expression of several genes associated with this region (*LCORL*, *NCAPG*, and *DCAF16*) in hair root cells to assess whether there was a link between their expression and the *BIEC2-808543* genotype. Though the SNP they used may not be the true causative mutation, horses, like cattle, exhibit high LD in the region surrounding this locus [58], so while certainly not definitive, *BIEC2-808543* can be an appropriate enough marker for this haplotype in this species. This study found a statistically significant link between this genotype and expression, with C/C, the genotype linked with smallest body size, to be expressing about twice as much *LCORL* relative to T/T animals ($p<0.001$). Expression of *DCAF16* and *NCAPG* showed no apparent link to genotype for this SNP [59]. The PCR primers utilized in this study amplified the X1 and X2 isoforms of equine *LCORL*, which encode for 'long LCORL'. This suggests that a potential mutation in the regulatory region of *LCORL* in these horses could have caused a reduction in expression of 'long LCORL' which resulted in a similar increase in body size observed in animals with a loss of function of the protein.

Lindholm-Perry et al. used a similar approach in crossbred cattle, after previously observing this locus to be associated with ADG and average daily feed intake (ADFI) [33]. Expression levels of *LCORL* and *NCAPG* were quantified in muscle and adipose tissue of heifers, cows, and steers. They found increased *LCORL* expression to be significantly associated with decreased ADFI, with no significant impact on ADG. Though this link was observed across heifers, cows, and steers, it should be noted that in heifers and cows, the

association was only observed in adipose tissue, but in steers, *LCORL* expression was only significantly different in muscle tissue. Western immunoblots were also performed to measure total LCORL protein from the steers' muscle tissue. Total LCORL protein was also significantly associated with ADFI (p=0.01). Their expression analysis also found muscular expression of *NCAPG* to be significant for ADG, but only in cows. Heifers and steers had no statistically significant impact on ADG or ADFI from *NCAPG* expression. Perhaps *NCAPG* expression could contribute to fat accretion, due to its link to ADG in adult cows, but these results do not imply *NCAPG* to be causative for the lean growth phenotype observed in association with this locus. While the protein measured in the immunoblot was total LCORL, the primers they used for qPCR analysis captured three isoforms of 'long LCORL'. The lack of change in ADG suggests lack of significant change in day-by-day rate of growth related to *LCORL* expression, though reduced feed intake alongside unchanged ADG implies an improved feed conversion ratio. Curiously, unlike the evidence in dogs, horses, and goats, this study found increased expression of 'long LCORL' to be potentially associated with the lean growth phenotype, though it does not point in one direction or another on body size in cattle, as this was not a trait described by their study [94].

Copy number variants (CNVs) have long been understood to be impactful on gene expression, though they have historically been difficult to observe due technological limitations. Additional copies of genes or transcription activating sites naturally gives the cell's transcriptional machinery more binding sites to work with, and absent regulatory feedback, tends to result in increased expression for the copied region [95]. Geistlinger et al. observed copy number variation in the region containing *LCORL* in 723 Brazilian Nelore steers. They used two different methods to call CNVs, with conflicting results in this region. PennCNV, which uses signal intensity data from high-density SNP arrays to infer copy number changes, called the region as containing a rare single copy gain, with 10 animals as 3n for the region, and the

rest as 2n. However, when using CNVnator, which calls CNVs based on expected read depth, using whole genome sequencing data taken from the sires of the steers phenotyped, the region was instead called as a total deletion in some of the sires. The copy number gain called by PennCNV was not replicated in CNVnator. RNA-seq analysis did not find a change in *LCORL* expression between steers called as 2n or 3n by PennCNV. However, they found the CNV-bearing region to be orthologous to a transcription-factor binding site annotated in the human genome, including SIX5 and SAP30, which were found to be up- and down-regulated respectively in 3n steers, suggesting there could be a change in transcription factor affinity or activity caused by copy number variation in this region [96]. However, a separate study by Rafter et al. also using PennCNV did not observe any association between CNVs in the *LCORL-NCAPG* locus and body weight, conformation, or fat. This study used Charolais and Limousin cattle, which are known to be segregating for a lean growth haplotype in this region, suggesting that the CNVs they observed were not impactful on *LCORL* expression, nor were they in LD with the causative haplotype [97].

A recently published study in horses identified the presence of retroCNVs of *LCORL* in equids. Batcher et al. discovered that all domestic and wild equids they investigated possessed between 17 and 35 inserted retrocopies of *LCORL* on the ninth chromosome. Curiously, despite the fact that this insertion was present across all domestic horses in their study, it was not present in the current horse genome assembly, though it was included on an unplaced contig in a previous version. These retrocopies of *LCORL* originated from the horse's 'short LCORL' transcript. Their expression analysis found these retrocopies are actually responsible for most *LCORL* expression in horses. In most tissues, the retrocopies are responsible for over 90% of *LCORL* expression, and compared to humans, horses have a much higher expression level of *LCORL*. Though highly expressed, it is currently unclear whether these retrocopy transcripts are functional, as they have accrued several mutations exclusive from the parent gene, some of

which may result in truncation or loss of function. The investigators found these retrocopies to be present across all equids investigated, however they were not found among other odd-toed ungulates like tapirs or rhinoceroses, suggesting that this mutation is exclusive to equids. This retroCNV is estimated to be around 18 million years old, and was suggested by the researchers to coincide with the rapid increase in size observed in equid evolution at this time [98]. It is important to note that these investigators were studying the expression levels of all *LCORL* transcripts using RNA-seq and specifically comparing to the retroCNV-specific transcript of *LCORL*, compared to Metzger et al., who were only observing the expression of 'long LCORL'. While much of horse *LCORL* expression indeed is being contributed by these retroCNVs of 'short LCORL', the differential expression analysis of *LCORL* between horses of differently sized breeds was only capturing the expression of 'long LCORL'. These results are not in conflict with each other, but they do pose the question of whether perturbations of 'short LCORL' expression levels due to changes involving the parent gene would be observable in horses.

Though there is a great deal of compelling evidence that variations involving *LCORL* and its expression have some link to phenotype, a mutation that can be definitively pointed to as causing the increased growth seen in cattle has yet to be identified. Due to the LD in the region, that may prove a difficult task. A selective sweep has resulted in a wide-ranging haplotype at this locus, with many variants likely to be hitchhiking along with any causative variants. Given this, it may not be possible to identify a single causative variant by sequencing alone, but it certainly is possible to further define this haplotype. By creating an exhaustive list of variants unique to this haplotype, we can narrow the list of candidate variants, which can hopefully be leveraged in future research to determine the molecular variation responsible for the increased growth associated with this region.

Chapter 2

Defining a haplotype encompassing the LCORL-NCAPG locus associated with increased lean growth in beef cattle

Abstract

Numerous studies have shown genetic variation at the *LCORL-NCAPG* locus is strongly associated with growth traits in beef cattle. However, a causative molecular variant has yet to be identified. To define all possible candidate variants, 34 Charolais-sired calves were whole genome sequenced including 17 homozygous for a long-range haplotype associated with increased growth (*QQ*), and 17 homozygous for potential ancestral haplotypes for this region (*qq*). The *Q* haplotype was refined to an 814-kb region between chr6:37,199,897-38,014,080 and contained 218 variants not found in *qq* individuals. These variants include an insertion in an intron of *NCAPG*, a previously documented mutation in *NCAPG* (rs109570900), two coding sequence mutations in *LCORL* (rs109696064 and rs384548488), and 15 variants located within ATAC peaks that were predicted to affect transcription factor binding. Notably, rs384548488 is a frameshift variant likely resulting in loss of function for long isoforms of *LCORL*. To test the association of the coding sequence variants of *LCORL* with phenotype, 405 cattle from five populations were genotyped. The two variants were in complete linkage disequilibrium. Statistical analysis of the three populations that contained *QQ* animals revealed significant ($p < 0.05$) associations with genotype and birth weight, live weight, carcass weight, hip height, and average daily gain. These findings affirm the link between this locus and growth in beef cattle and describe DNA variants that define the haplotype. However, further studies will be required to define the true causative mutation.

Introduction

Body size is a trait frequently subject to selection in domesticated animals. This trait is of particular importance in species raised for meat as body size directly correlates to the size of the carcass and thus quantity of meat produced per animal. In beef cattle, body size can generally be described as a highly complex polygenic trait, with even the most impactful DNA variants explaining only a modest fraction of the variance observed [13]. However, there are a select few loci that can be considered highly impactful for body size, one of them being the *LCORL-NCAPG* locus on chromosome 6. Over the years, many studies have indicated this locus as being associated with stature [13, 30, 80, 81], birth weight [15, 17, 23, 28], and carcass weight [12, 14, 31, 38], as well as other carcass characteristics such as increased ribeye area and reduced adiposity [33], suggesting a potential role in increasing lean growth. These findings are supported by research in other species including humans [43, 44, 53], dogs [42, 57], horses [60-62], and sheep [41, 65], that have also found variants surrounding this locus to be associated with increased body size.

However, while there is no doubt that this locus has a significant impact on body size, the biological mechanism mediating this effect remains unclear. The large effect size of this locus resulted in rapid selective sweeps in cattle [12, 13, 77, 99] and other domesticated species [68, 69, 74]; the ensuing linkage disequilibrium (LD) has long confounded attempts to discern a causative mutation within the selected haplotype. Indeed, it has been shown the same haplotype at this locus is almost fixed in some breeds of cattle such as Brown Swiss and Montbeliard [13]. That same study found that in other breeds like Charolais, this haplotype is present and very abundant, but not to the exclusion of other possible haplotypes for the region. This locus has been narrowed to a range as small as 591 kb [38], but further refinement remains complicated by the aforementioned LD.

Roughly half of this range encompasses a gene-dense region containing four genes: *LCORL, NCAPG, DCAF16,* and *FAM184B*. The other portion of this range is intergenic space upstream of *LCORL*. Many genome-wide association studies have found the most strongly associated markers with growth to be associated with the intergenic region upstream of *LCORL* [17, 25, 27, 31], suggesting a mutation affecting the regulation of *LCORL* transcription may be responsible for the change in phenotype. Earlier studies pointed to a missense mutation in *NCAPG* [22, 80], *NCAPGc.1326T>G*, but subsequent studies have found this mutation present in *q* haplotypes as well, implying this SNP is simply in high LD with the causative mutation in certain populations [24, 25].

Recent evidence in other species places more weight in favor of *LCORL* as influencing body size development. In dogs, a mutation that results in the loss of function of the long isoform of *LCORL* is exclusive to medium and large-sized breeds of dogs [42]. A similar mutation exists in large-sized Pakistani goat breeds [68], however in both cases there has yet to be functional validation of these mutations. This long isoform of *LCORL* encodes for a protein dubbed PALI2 by Conway et al. [85], which, while not directly characterized, is postulated to have a similar role to PALI1, a very similar protein encoded by the paralog of *LCORL, LCOR*.

PALI1 acts as an accessory protein to polycomb repressive complex 2 (PRC2), an essential protein complex that is responsible for the creation and maintenance of cellular identity by mono-, di-, and trimethylation of H3K27. This methylation represses transcription and is maintained from early development onward [100]. Disturbances of PRC2 function have been shown to cause abnormalities in body plan formation in animals, demonstrating that this complex plays a key role in this process [88]. The repressive activity of PRC2 is increased when the complex is accessorized with PALI1 [85]. It is thought that due to a shared domain between PALI1 and PALI2, PALI2 can likely interact with PRC2 in the same way; though as of this writing, no functional data has been published for PALI2.

It is presently unclear what molecular variation is responsible for the influence this region has on animal growth. Thus, the objective of this study was to comprehensively define the extended haplotype associated with increased growth by determining all potential variants that can be considered in phase or not in phase with the haplotype using 34 Charolais-sired calves, 17 of which are *QQ* (homozygous for bearing the mutant haplotype associated with increased growth) and 17 are *qq* (homozygous for the 'wild-type' of this haplotype; or in other words not bearing the *Q* haplotype). These variants were then used to explore the associated effects on phenotype by genotyping a larger, multi-breed population for which several growth and carcass traits had been collected.

Materials and Methods

Sample collection

Whole blood samples were used from 439 cattle across six previous studies conducted at University of Illinois in accordance with animal care and use committee procedures associated with their respective projects. The number of individuals and sire breed composition of each group are presented in Table 2.1. As the current experiments were performed post hoc, there was no need for an animal care and use protocol for this project. DNA was extracted from these samples using a classic salting out procedure [101] for populations 0, 1, 2 and 3. For populations 4 and 5, DNA was extracted using a *Quick*-DNA kit (Zymo Research, Irvine, CA) following the manufacturer's protocol.

For the initial objective, the 34 Charolais-sired calves from population 0 were selected for whole-genome sequencing. These animals were previously shown to be homozygous for the *QQ* (*n*=17) and *qq* (*n*=17) haplotype within the region of interest [27]. DNA libraries were prepared using an Illumina DNA Prep library kit (Illumina, San Diego, CA), following the manufacturer's instructions. Libraries were paired-end sequenced using both lanes of a

Table 2.1 The populations of cattle used for this project.

Population	Sire Breed	*N*
0	Charolais	34
1	Simmental-Angus & Angus	30
2	Shorthorn	87
3	Angus	83
4	Simmental & Simmental-Angus	78
5	Simmental-Angus	127

Novaseq 6000 S2 flowcell at 2 x 150 cycles.

Read quality was assessed using FastQC version 0.11.9 [102], and trimming was carried out using Trimmomatic version 0.39 [103] in paired-end mode using the following parameters: HEADCROP:1 ILLUMINACLIP:2:30:10 LEADING:28 SLIDINGWINDOW:15:28 MINLEN:75. Reads were aligned to the ARS-UCD 2.0 assembly of the bovine genome using BWA-MEM [104] using the default settings. Alignments with a MAPQ below 30 were filtered out, to remove poor quality mappings and multi-mapped reads, before being passed to GATK 4.2.2.0 for final processing and variant calling [105]. Duplicates were flagged using MarkDuplicatesSpark and variants were called using HaplotypeCaller, both using default parameters. The variant call files were merged with GenomicsDBImport and joint genotyped with GenotypeGVCFs to create a combined VCF with all animal genotypes. Lastly, variants were hard filtered with VariantFiltration with the following filters: "QD < 5.0", "SOR > 2.5", -filter "FS > 20.0", and "MQ < 50.0". Mapping Quality Rank Sum (MQRS) and Read Position Rank Sum (ReadPosRankSum) were not used, due to their inability to be calculated when there are no samples heterozygous for a variant present. The choice of using *QQ* and *qq* animals for sequencing required that the variants of interest not be heterozygous, so filtering by those parameters would have removed them. For similar reasons, as well as for quality control, multiallelic variants were also removed.

Identification of haplotype-defining mutations

To identify the variants that could be causing the *QQ* growth phenotype, a subtractive approach was taken. Using 17 *QQ* and 17 *qq* animals known from previous haplotype and phenotype information, the haplotype in this population was refined to a span between chr6:37,199,897-38,014,080, where almost all variants are entirely fixed in *QQ* individuals. To assess the validity of variants that did not present as fixed, they were visualized in IGV. If called variants could be attributed to a single read or PCR duplicate or were found in an area where

alignment might be impeded like a structural variant or repetitive element, they were removed from the final list of variants for consideration. This approach removed all variants that were not in phase with the *Q* haplotype. To determine variants in the haplotype that defined *Q*, the list was further narrowed to variants fixed in *QQ* animals, but where all *qq* individuals were homozygous for the opposite allele. Under the assumption that Dominette was *qq,* any variants where the *Q*-exclusive variant matched the reference were also removed. By filtering based on these criteria, the potential variant set was refined to variants exclusive to and completely in phase with the *Q* haplotype in this population.

Using Ensembl's Variant Effect Predictor (VEP) [106], this set of variants was annotated based on location in the annotated genome (intronic, coding sequence, untranslated regions, intergenic, etc.), as well as impacts on coding sequence and reading frame. Effects on splicing were predicted with Pangolin [107] using the default settings. To investigate potential changes in transcription factor binding, variants were loaded into the UCSC genome browser as a custom track [108], alongside the bovine ATAC-seq peak catalog published by Yuan et al [109]. Variants located within ATAC-seq peaks and thus potentially affecting transcription factor (TF) binding were identified by visualizing the data in the browser. For further analysis, only variants within peaks reaching a signal score of at least 0.5 were considered. The effects of those variants were further interrogated using the Transcription Factor Binding Site Prediction tool provided by AnimalTFDB4.0 [110] to identify specific TFs that could bind to the affected regions. FIMO [111] was then used to calculate scores for the original and mutated version of the sequence, using all motifs associated with that transcription factor from CIS-BP [112]. For the final comparison, the motif with the highest FIMO score possible was selected for each TF.

Individual variant genotyping

To validate the presence of these variants in a broader population and to assess their potential as surrogates for the haplotype and predictors of phenotype, the remaining 405

animals were genotyped for two coding sequence variants found in LCORL exclusive to the Q haplotype, rs109696064 (chr6:g.37403795T) and rs384548488 (chr6:g.37401771_37401772del). For populations 1, 2, and 3, genotypes were called using a PCR-RFLP assay. Primers and enzymes used are described in Table 2.2. PCR was conducted in 20µL reactions containing 100ng genomic DNA, 0.5U Hot-StarTaq DNA polymerase (QIAGEN, Valencia, CA), 1x PCR buffer, 200µM each dNTP, and 0.5µM each of forward and reverse primers. Amplification was performed with an initial incubation at 95˚C for 3 minutes, followed by 34 cycles of 94˚C for 20 s, 45 s, and 72˚C for 45 s, with a final incubation at 72˚C for 5 m. Following PCR, 10 µL of master mix (1.5µL of 10X enzyme buffer, 5 Units of restriction enzyme, and 8µL nuclease-free water) was added to each PCR reaction and incubated for one hour at 37˚C. The digested fragments were subjected to electrophoresis in 1.5% agarose, 0.5x tris-borate-EDTA gels with 0.1 µg/mL ethidium bromide. Genotypes were visualized by UV illumination. For populations 4 and 5, a fluorescent 5'-3' exonuclease assay was per-formed using primers and probes as described in Table 2.3. Quantitative PCR was performed in 10 µL reactions containing 1x PrimeTime™ Gene Expression Master Mix (Integrated DNA Technologies, Coralville, IA), 0.5µM each of forward and reverse primers, 0.2 µM of each allele-specific probe. Amplification was performed with an initial incubation at 95˚C for 3 minutes followed by 40 cycles of 95˚C for 15 s, 60˚C for 45 s. Genotypes were called using CFX Maestro software (Bio-Rad Laboratories, Hercules, CA).

Statistical analyses

Statistical analyses were performed in R version 4.3.2. lmer() [113] was used to construct linear mixed-effect models using genotype and sex as a fixed effect and farm as a random variable. Due to a lack of individuals bearing the alleles used as surrogates for the *QQ* genotype, populations 1 and 3 were not used for analysis. The final models for the first set were constructed using populations 2, 4, and 5 combined, using genotype and sex as fixed effects

Table 2.2 Primers, enzymes, and expected fragment sizes for the PCR-RFLP genotyping assay.

Variant	Primer Type	Primer Sequence	Enzyme	Fragment Sizes
rs109696064	Forward	CAGGTGCTGTTCTGGACTGT	HpyAV	C: 129, 152, 394
	Reverse	GGTGAGTCAGGAGAGCTGTG		T: 281, 394
rs384548488	Forward	CTGCAGTGTTTACAGATGATATGA	HinfI	ACT: 75, 162, 344
	Reverse	AGGCTTGTGATTTGGTCGAA		A: 75, 504

Table 2.3 Primers and probes for the 5’-3’ exonuclease genotyping assay.

Variant	Primer Type	Primer Sequence
rs109696064	Forward	AGATCTCGAAGTCCCTCACC
	Probe *C*	5’-6-FAM/AGGCTGAAGGA-3’iABkFQ
	Probe *T*	5’-SUN/CAGGCTAAAGGAT-3’iABkF
	Reverse	GCTGAACACTCCGAGATGAC
rs384548488	Forward	GACCAAACAGACATTGCCAT
	Probe *ACT*	5’-6-FAM/CCCAGAGTCTCT-3’iABkF
	Probe *A*	5’-SUN/GCCCAGTCTCTCA-3’iABkF
	Reverse	TCTTGACCTTCCATAAATAACAGT

and population group as a random effect to test for association between genotype and 13 phenotypes: birth weight (BW), adjusted weaning weight (WW), three weight points through life (W1, W2, and W3), average daily gain (ADG), hip height (HH), dry matter intake (DMI), hot carcass weight (HCW), backfat thickness (BF), ribeye area (REA), kidney pelvic heart fat (KPH), and marbling (MB). As of the writing of this manuscript, DMI and carcass phenotypes were not available for populations 4 and 5, so only population 2 was considered for DMI, HCW, BF, REA, KPH, and MB. Additionally, only steers had HH and W3 measured in populations 4 and 5, though there were measurements from steers and heifers for these traits in population 2. Due to being calved considerably later than their contemporaries, five animals from population 4, and 34 from population 5 were not used for analyses. To adjust for multiple testing, the Benjamini-Hochberg correction was used [114]. All phenotypes within classes broken down by population and sex passed the Shapiro-Wilks normalcy test (p>0.05), with the exception of W2, ADG, and MB for the steers of population 2, and birth weight for the steers of populations 4 and 5.

Results

Defining mutations exclusive to the *Q* haplotype

Thirty-four Charolais calves were whole genome sequenced (17 *QQ* and 17 *qq*), with the goal of building a complete list of the potential causative mutations to better understand the functional impact of this haplotype. An average of 119.1 million read pairs were generated per sample, with 103.3 million read pairs remaining after trimming. After removal of reads with a mapping quality below 30, each sample had 170.7 million reads mapped to the genome on average, resulting in a final coverage of roughly 9.2x for sample, with the range of coverage for individual samples being between 7.6x and 11.5x.

As the ultimate objective was to investigate the *Q* haplotype at *LCORL-NCAPG* locus, the first step was to identify the span of LD within this population. Based on the evidence from

previous studies [27, 38], the initial area of exploration encompassed 37,000,000 and 38,200,000 on BTA6, which contained *LAP3, MED28, FAM184B, DCAF16, NCAPG, LCORL,* and the 600kb intergenic region upstream of *LCORL*. Within this area, an 814-kb region between 37,199,897-38,014,080 was shown to be in complete LD among the *QQ* animals, removing *LAP3, MED28,* and a portion of the 3' end of *FAM184B* from consideration. In total, 7,278 variants were called in this 814-kb region, 147 of which did not present as fixed in the *Q* haplotype. To validate these mutations, they were visualized in IGV. Of these variant calls, 81 could be attributed to a single read or PCR duplicate. Another 35 were in repetitive regions and could be construed to issues with alignment. There were three regions containing 21 total variants that were the result of failed alignment of larger repetitive regions. Five variants were clustered around apparent structural variants that are in phase with the haplotype, and the remaining five (rs109576691, rs379524098, chr6:g.37726299T>A, rs109270787, and rs383633472) did not fit any of these criteria, and could be true germline variants existing within the *Q* haplotype or somatic mutations within individual animals. None of these five variants presented as being due to recombination events; the *Q* haplotype continued uninterrupted on either side of the variant in the individuals where they did not fit the defined *Q* genotype. Because of this, these variants were not useful in refining the *Q* haplotype further, nor are likely to be causative.

After quality control pruning a total of 7,131 SNPs and indels were identified as being fixed among *QQ* animals in this region. To be considered a variant 'defining' the haplotype, a variant had to be completely absent in *qq* individuals. In other words, while *QQ* animals were homozygous for one allele, the *qq* cattle had to be homozygous for the opposite allele. There were 217 variants that met these criteria. These variants were subject to further investigation to identify potential causative mechanisms using three prediction methods: VEP, for coding sequence and effects on reading frame; Pangolin, to detect changes in splicing sites; and

combining published ATAC-seq data for cattle with FIMO prediction of binding affinity for transcription factors obtained from AnimalTFDB.

Of the 217 variants, only 22 were within coding sequence or an ATAC peak (Table 2.4). No variants had a notable impact on splicing; the highest increase in splice probability score calculated by Pangolin was 0.02, and no decreases in splice score were observed. Of these 22 variants, three were detected in coding sequence: rs109570900 (the *NCAPGc.1326T>G* seen in previous studies), and rs3845484488 and rs109696064.

SIFT scores calculated for the SNPs indicate that rs109696064 is tolerated (0.38), but rs109570900 is deleterious (0.01). The last coding sequence variant, rs3845484488 is a frameshift that results in a truncation of the long isoform of the LCORL protein, so it is likely to be impactful. The other 19 variants found in ATAC peaks were distributed across 15 regions labelled A through O, with A, C, and G containing variants from two different peaks (labelled as A-1 or A-2, etc.) Among these, G encompassed the promoter region directly upstream of *LCORL.* However, neither of the variants in this region were in the major promoter ATAC peak, but in smaller adjacent peaks. Region B included the promoter for *FAM184B,* but the other regions were distributed among intronic and intergenic regions. A summary of the ATAC peaks these variants are in can be found in Table 2.5, and the differential transcription factor binding between the *Q* and *q* alleles for these variants is shown in Table 2.6.

While four of the variants in ATAC peaks had no predicted change to transcription factor affinity or any transcription factor affinity in their site, the remaining 15 had some degree of change. Most of the ATAC peaks harboring variants were relatively small, with only six having a signal score greater than 0.5. Notably for region G, though G-2 did not have any motif hits, the variant at G-1 results in a loss of affinity for EGR1 and EGR2 among others, but simultaneously a gain in affinity for FOXA1. Region L, an embryonic ATAC peak with the highest signal score in its upstream region, was also remarkable for its mutation causing a loss of affinity for all TFs

Table 2.4 The 22 variants detected in coding sequence or ATAC peaks.

Variant	Location[1]	*q* allele	*Q* allele	Nearby Gene	Type, Consequence
rs109438687	37,214,389	T	C	*FAM184B*	ATAC peak, intron
rs109467519	37,214,736	C	T	*FAM184B*	ATAC peak, intron
rs109145748	37,301,160	G	C	*FAM184B*	ATAC peak, 5' UTR
rs109570900	37,343,379	T	G	*NCAPG*	Coding sequence, missense
rs210386983	37,379,506	A	G	*LCORL*	ATAC peak, 3' UTR
rs207496787	37,379,507	A	T	*LCORL*	ATAC peak, 3' UTR
rs379449143	37,381,106	A	G	*LCORL*	ATAC peak, intron
rs384548488	37,401,770	ACT	A	*LCORL*	Coding sequence, frameshift
rs109696064	37,403,795	C	T	*LCORL*	Coding sequence, missense
rs517494305	37,452,882	C	CA	*LCORL*	ATAC peak, intron
rs110293947	37,479,269	G	C	*LCORL*	ATAC peak, intron
rs379787611	37,487,010	T	C	*LCORL*	ATAC peak, intron
rs109114124	37,555,677	C	A	*LCORL*	ATAC peak, intron
rs109092727	37,559,117	A	G	*LCORL*	ATAC peak, upstream
rs110470694	37,608,504	C	T		ATAC peak, intergenic
rs109060347	37,627,776	G	C		ATAC peak, intergenic
rs207689046	37,669,453	A	G		ATAC peak, intergenic
rs109331793	37,681,968	C	T		ATAC peak, intergenic
rs110458346	37,934,068	C	T		ATAC peak, intergenic
rs110888204	37,946,012	C	T		ATAC peak, intergenic
rs110930653	37,962,887	G	T		ATAC peak, intergenic
rs110658468	37,997,160	C	T		ATAC peak, intergenic

1 - All variants are located on chromosome 6, ARS-UCD2.0 (NC_037333.1).
2 - Consequence classification is determined by VEP.

Table 2.5 Consensus ATAC peaks with variants, their associated tissues, and signal scores.

Site	Consensus Peak	Variant	Tissue	Signal Score[1]
A-1	chr6_37214241_37214456_NMF12_0.33	rs109438687	Liver & Testicle	0.353
A-2	chr6_37214731_37214965_NMF10_0.92	rs109467519	Muscle	0.209
B	chr6_37300595_37301446_NMF10_0.14	rs109145748	Ubiquitous	0.909
C-1	chr6_37379341_37379516_NMF13_1.00	rs210386983 & rs207496787	8-cell Embryo	0.260
C-2	chr6_37381082_37381259_NMF13_1.00	rs379449143	8-cell Embryo	0.302
D	chr6_37452831_37452975_NMF16_0.67	rs517494305	Adipose	0.222
E	chr6_37479127_37479316_NMF7_0.58	rs110293947	Cerebellum	0.277
F	chr6_37487007_37487238_NMF13_1.00	rs379787611	8-cell Embryo	0.337
G-1	chr6_37555562_37555843_NMF9_0.15	rs109114124	Ubiquitous	0.522
G-2	chr6_37558579_37559199_NMF14_0.35	rs109092727	Embryo	0.866
H	chr6_37608344_37608531_NMF5_1.00	rs110470694	Colon & Embryo	0.422
I	chr6_37627598_37627866_NMF5_0.33	rs109060347	Colon, Rumen Epithelial & Embryo	0.947
J	chr6_37669337_37669559_NMF5_1.00	rs207689046	Colon	0.715
K	chr6_37681884_37682015_NMF13_1.00	rs109331793	8-cell Embryo	0.294
L	chr6_37933851_37934401_NMF13_1.00	rs110458346	8-cell Embryo	0.758
M	chr6_37945955_37946153_NMF10.88	rs110888204	Cerebrum	0.324
N	chr6_37962696_37962988_NMF16_0.49	rs110930653	Epididymis	0.198
O	chr6_37996798_37997347_NMF5_0.69	rs110658468	Colon	0.401

1 - Signal score was determined by the highest signal in the entire consensus peak where the variant was found.

Table 2.6 Transcription factors predicted to bind to the ATAC regions containing variants.

Site	Shared TFs	*qq* TFs	*QQ* TFs
A-1	ZNF621	-	NR2C2, PAX6
A-2	-	-	-
B	GCM1, MAZ, SP2, ZNF180, ZNF212, ZNF341, ZNF467, ZNF527, ZNF548, ZNF596, ZNF792	PAX6, ZBTB14, ZFP64, ZNF264	KLF11, SP1, ZBTB17, ZNF329
C-1	-	-	-
C-2	-	BATF3	POU6F1
D	NFATC1, RELA, ZNF484	NFATC3, REST	NFYA, NFYB, NFYC, ZNF280A, ZNF619
E	NFE2L2, NHLH1, NHLH2, OLIG2, TCF12, ZBTB18, ZNF273, ZNF331	ASCL2, MYOG, ZNF549, ZNF69, ZSCAN31	-
F	IRF1, STAT2, ZIM3, ZNF225, ZNF487, ZNF502	MEF2A, ZNF394	IRF2, IRF3, IRF4, IRF5, IRF8, IRF9, ZNF573
G-1	RREB1, ZNF263, ZNF283, ZNF785, ZNF805	EGR1, EGR2, MAZ, ZNF460, ZNF580	FOXA1
G-2	-	-	-
H	KLF15, ZNF383, ZNF432, ZNF880	-	ZNF449
I	ZNF335	NKX2-5	SOX18, ZNF200, ZNF808
J	ESR1, NR1H3, NR2C2, YY1	CREB3L1, CREB3L2, RORA	RXRG
K	CUX1, CUX2	ZNF667	ZNF605
L	-	MEF2B, POU6F1, SRF, THRA, THRB, ZNF774, ZNF823	-
M	ZNF768	PPARG, ZBTB12, ZNF543, ZNF621, ZNF768	ZNF440
N	-	NR1I2	-
O	-	-	-

Shared TFs had motif scores passing the q-value threshold (<0.05) for both the *Q* and *q* allele sequence, while *qq* and *QQ* TFs only had passing scores for their respective allele.

predicted for that region in *qq*, including thyroid hormone receptors THRA, THRB, and SRF.

Discovery of structural variants

As part of quality control, regions with ambiguous calls were visualized in IGV. This led to the discovery of three apparent structural variants: 1) a 157-bp deletion within the first intron of *NCAPG* (chr6:37,326,536-37,326,693) (Figure 2.1), 2) an insertion within the fifth intron of *NCAPG* (chr6:37,336,715-37,336,716) (Figure 2.2), and 3) a small insertion in the intergenic space upstream of *LCORL* (chr6:37,619,145-37,619,149). As all three of these were within the fixed haplotype region for *Q*, they were homozygous in all *QQ* samples. With respect to *qq* samples, the smaller insertion upstream of *LCORL* was very common; 10 of the *qq* individuals were homozygous for having this insertion and 3 were heterozygous for it. While rarer, there were *qq* animals heterozygous for the deletion in the *NCAPG* intron as well. However, the insertion within *NCAPG* seems to be completely absent among *qq* animals, and thus would qualify as being a defining mutation for this haplotype. Mate pairs of reads entering the insertion point at this region appear on multiple different chromosomes, suggesting the insertion of a repetitive element. As the insertion is large enough to not have any reads appear on both sides of it, it is difficult to say at this time what exactly it is, or how it could impact the transcription of *NCAPG* or perhaps neighboring genes like *LCORL*. Figure 2.3 displays this structural variant, alongside the other 22 variants found in coding and ATAC sequence, and their relative position within the haplotype.

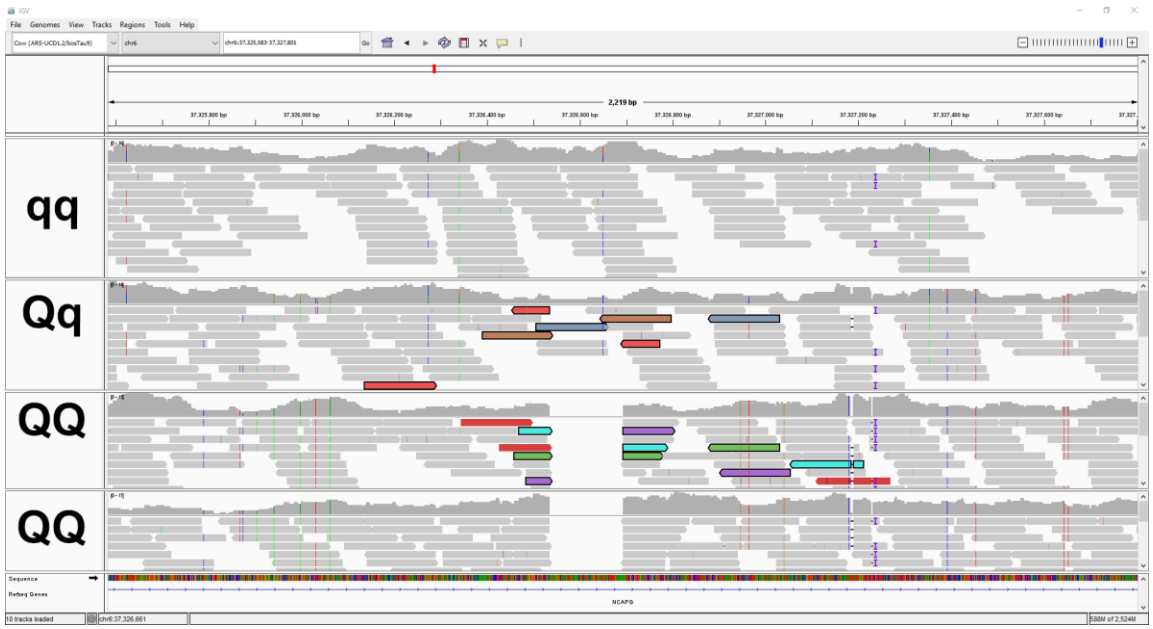


Figure 2.1 A 157-bp deletion within the first intron of *NCAPG*. The second and third track have read pairs highlighted to demonstrate that this structural variant is a simple deletion; reads are soft-clipped and able to pick back up on the other side. The *Qq* individual shows reads that can pass through the gap uninterrupted, as well as the soft-clipped reads showing it carries the deletion.



Figure 2.2 An insertion within the fifth intron of *NCAPG,* between chr6:37,336,715-37,336,716.

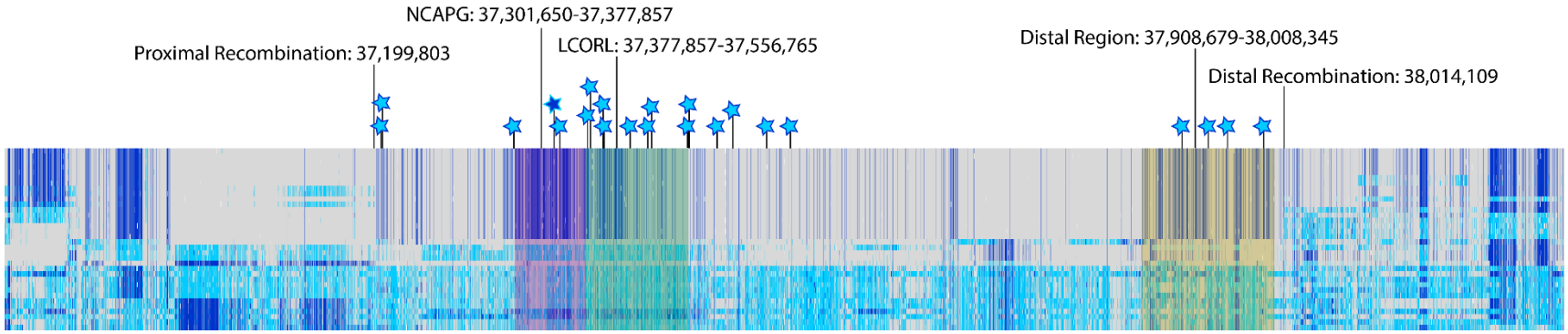


Figure 2.3 A display of all variants observed between chr6:37,000,000-38,200,000. Each horizontal bar represents one of the genotyped cattle, and each vertical line represents their respective genotype for each variant. Grey represents homozygous for matching the reference (0/0), light blue is heterozygous (0/1), and dark blue is homozygous for the mutant allele (1/1). White indicates that a genotype was not able to be confidently called for that individual at that variant. The *qq* animals are on the left, and the *QQ* haplotype is visible on the right. Each star marks one of the 22 variants in phase with the haplotype that were found in coding sequence or an ATAC peak, and the inverse-colored star marks where the structural variant in phase with the haplotype roughly is, though the insertion itself is not one of the variants on this map. The mutant variants in this haplotype tend to cluster around *NCAPG* (highlighted purple), *LCORL* (highlighted green), and the distal region around chr6:37,900,000-38,000,000 (highlighted yellow).

Genotype-phenotype relationship

To confirm the existence of some of these mutations in a broader population, and to assess the association between selected SNPs and phenotype, an additional 405 cattle from five populations were genotyped for the rs384548488 and rs109696064 variants. These variants were selected due to their location in coding sequence, the potential significant impact of rs384548488, and to assess if either of these two very closely neighboring variants were found independent of one another. The variants were found to be in complete LD, with all animals being either homozygous ACT-C/ACT-C (as in wild type, or *qq*), A-T/A-T (*QQ*), or heterozygous for both (ACT-C/A-T, or *Qq*). Genotype distribution by population is listed in Table 2.7. Due to the proximity of these variants, and that these were the only variants genotyped, it cannot be confirmed these cattle truly have the extended Q haplotype. However, given the evidence presented in the whole genome sequenced animals, it may be possible to use these variants as a surrogate for the haplotype.

To assess the association between genotype and phenotype, linear mixed-effects models were constructed. As there were no *QQ* individuals in populations 1 and 3, only populations 2, 4, and 5 were used. Due to unavailability of DMI and carcass phenotypes for populations 4 and 5, only population 2 was used for the DMI, HCW, REA, BF, KPH, and MB phenotypes. The model constructed for each phenotype is presented in Table 2.8. BW, W2, W3, HH, ADG, and HCW all passed the significance threshold, with *Qq* and *QQ* cattle trending toward higher body weight, increased stature, and greater average daily gain than qq animals. This is consistent with the known effect of this locus on phenotype and demonstrates these variants can serve as effective markers for the haplotype.

Table 2.7 Genotype frequencies by population.

Population	*qq*	*Qq*	*QQ*
1	18	12	0
2	49	31	7
3	63	20	0
4	14	47	17
5	62	59	6

Possible alleles were *q* (C for rs109696064 and ACT for rs384548488) or *Q* (T for rs109696064 and A for rs384548488), associated with the wild-type and mutant haplotype, respectively.

Table 2.8 Phenotype and the statistical significance of genotype in the association model.

Phenotype	*n*	*p*-value
Birth Weight (BW)	253	3.67e-4 (*)
Adjusted Weaning Weight (WW)	241	0.327
Weight 1 (W1)	247	0.0603
Weight 2 (W2)	242	4.10e-5 (*)
Weight 3 (W3)	157	1.40e-4 (*)
Average Daily Gain (ADG)	244	2.27e-5 (*)
Hip Height (HH)	157	0.0185 (*)
Dry Matter Intake (DMI)	87	0.221
Hot Carcass Weight (HCW)	83	0.0152 (*)
Backfat (BF)	83	0.130
Ribeye Area (REA)	83	0.0423
Kidney Pelvic Heart Fat (KPH)	83	0.102
Marbling (MB)	83	0.267

Statistical significance (*) was determined using the Benjamini-Hochberg Method to correct for multiple testing.

Table 2.9 Parameter estimates for each phenotype.

Phenotype	(Intercept)	β_*Qq*	β_*QQ*	β_Steer	Pop2	Pop4	Pop5
BW, kg (*)	31.7 ± 2.3	0.4 ± 0.7	4.2 ± 1.1	3.1 ± 0.6	+3.5	-3.9	+0.5
WW, kg	206.8 ± 17.4	5.3 ± 4.6	9.5 ± 7.2	13.9 ± 4.1	+5.6	-31.6	+26.0
W1, kg	161.3 ± 54.5	7.0 ± 2.9	3.7 ± 4.7	12.5 ± 2.7	+108.9	-51.1	-57.8
W2, kg (*)	476.4 ± 12.9	22.6 ± 5.9	35.6 ± 9.7	56.0 ± 5.5	+20.7	-19.3	-1.4
W3, kg (*)	517.8 ± 21.4	31.0 ± 7.7	32.2 ± 13.0	55.3 ± 9.6	+37.3	-18.1	-19.2
ADG, kg (*)	1.46 ± 0.12	0.10 ± 0.03	0.18 ± 0.04	0.37 ± 0.02	-0.21	+0.01	+0.20
HH, cm (*)	120.9 ± 1.3	1.8 ± 0.7	2.6 ± 1.2	3.4 ± 0.9	+1.6	-1.1	-0.6
DMI, kg	9.02 ± 0.15	0.31 ± 0.21	0.44 ± 0.36	0.47 ± 0.20	-	-	-
HCW, kg (*)	350.2 ± 5.4	21.4 ± 7.4	16.7 ± 13.6	32.5 ± 7.0	-	-	-
BF, cm	1.59 ± 0.06	0.04 ± 0.09	-0.30 ± 0.16	-0.20 ± 0.08	-	-	-
REA, cm^2	82.4 ± 1.5	5.2 ± 2.0	2.8 ± 3.8	7.0 ± 1.9	-	-	-
KPH %	2.11 ± 0.03	-0.06 ± 0.04	-0.15 ± 0.08	-0.20 ± 0.04	-	-	-
MB	538.5 ± 15.5	32.1 ± 21.2	-12.5 ± 39.0	-73.3 ± 20.2	-	-	-

Intercept for these models is *qq* heifer, with β from *Qq*, *QQ,* and/or sex effect for steers being added as applicable. Farm is included as a random effect added to the intercept, with (*) denoting models where the genotype was statistically significant for phenotype. For DMI, HCW, BF, REA, KPH, and MB, there is no farm effect, due to only population 2 being phenotyped for these traits.

Discussion

The results of this investigation support the findings of numerous prior studies demonstrating the link between the *NCAPG-LCORL* locus and increased lean growth, and several variants have been identified that could underly the changes in phenotype. In agreement with the results presented by Bouwman et al. [13] and others [33, 38, 40, 99], there is a large region of LD in *QQ* animals sequenced, which unfortunately was not narrowed further compared to previous studies and continues to confound attempts to find a causative variant. Nevertheless, the list has been refined further by identifying variants exclusive to the haplotype and several potential molecular explanations for how this locus exerts its influence.

The most straightforward and tempting answers lie within the coding sequence mutations. The *NCAPGc.1326T>G* substitution, rs109570900, has been previously identified as a putative quantitative trait nucleotide (QTN) because the resulting missense substitution is predicted to be functionally damaging. Indeed, the *NCAPGc.1326G* allele was present in the 34 cattle sequenced here and was exclusive to the *Q* haplotype. Many studies investigating this region have found this SNP to be significantly associated with various phenotypes [22, 38, 80], and without doubt it is in LD with this haplotype. However, other studies have questioned whether this mutation is causative. At least two studies have documented animals that would be heterozygous for the growth haplotype at this locus (*Qq*) being homozygous for the G allele, or animals that should not have the haplotype (*qq*) being carrying a copy of this allele [24, 25]. While none of *qq* animals used in this study carried the G allele, it seems likely that this variant may exist outside of this haplotype in the broader population and may simply have been present on the chromosome on which the causative mutation first arose.

While evidence from previous studies suggests that rs109570900 is most likely not the causative variant, *NCAPG* itself cannot be fully ruled out, as indeed, expression of *NCAPG*

seems to be important for muscle development. A recent study has found in myogenic differentiation to be impaired in fetal myoblasts where *NCAPG* has been knocked down [115]. None of the 217 variants exclusive to *Q* were located within ATAC peaks in or directly upstream of *NCAPG*. However, the existence of an insertion within an intron of *NCAPG*, 84 bp downstream from an exon-intron junction, could potentially impact the transcription, splicing, or function of this gene, though more direct evidence will be necessary to confirm this impact.

Alternatively, the frameshift mutation of the long isoform of *LCORL* is quite compelling. The fact that other studies have documented a loss of function for this long isoform to be linked to increased stature in dogs and goats implies a similar mechanism could be at work here [42, 68]. The frameshift variant in cattle, rs384548488, has been noted within the last year by both Sanchez et al., Gualdrón Duarte et al., and Cai et al. to be associated with several beef production traits in mostly Charolais, Brown Swiss, and original Braunvieh cattle, height and length in Belgian Blues, and reduced young stock survival in Nordic Red cattle, respectively [116-118].

While the association with growth and carcass traits is unsurprising, given the body of literature surrounding this locus at this point, the influence on young stock survival observed by Cai et al. [117] is perhaps less expected. It may be easy to rationalize this effect as due to dystocia caused by the increase in birth weight associated with this genotype. Indeed, when considering stillbirths, almost the exact same region found in the current study's results (chr6:37,236,226-38,027,078) was the most highly associated with the trait. The immediately adjacent proximal region (chr6:36,679,547-37,179,665) was the most significant region for calf survival within their first year, but the previously mentioned haplotype region (chr6: 37,236,226-38,027,078) still exceeded the significance threshold as well. It is curious that a mutation mostly known for its effect on growth is linked to early death, even after the first month of life. This locus was also linked to a decrease in longevity in the same study that identified the frameshift

mutation in dogs [42]. That connection could simply be accounted for by correlation over causation with large breeds within a species tending towards a shorter lifespan, but it may also be that epigenetic changes spurred by the loss of function of PALI2, while also resulting in increased body frame, can impact longevity and survival. In their research on the epigenetic clock in dogs and humans, Horvath et al. [119] found regions that gained methylation with age were enriched for PRC2 targets and genes involved in development. Under the hypothesis that the long isoform of *LCORL* modulates PRC2 activity, there is an inviting connection to be made. However, much more evidence, particularly regarding epigenetic changes associated with this frameshift mutation, would be needed to draw this conclusion.

In a similar vein to the young stock survival locus being adjacent to, but not directly within the *LCORL-NCAPG* haplotype, Sanchez et al. acknowledge that the frameshift mutation is not in very high LD with the lead SNP in Belgian Blues. Given the extensive LD in this region and the nature of GWAS for quantitative traits, caution should be exercised when looking at a lead SNP, as it may exist in a small amount of higher performing *Qq/qq* animals in addition to being attached to the *Q* haplotype. The fact that this region is in such LD is strong evidence for a selective sweep, and that the causative mutation is likely exclusive to the haplotype. Confirmation of further meiotic events to narrow the haplotype further would be ideal. It is not lost on the author that over a decade ago, Setoguchi et al. suggested a considerably smaller, 591-kb haplotype region in Japanese Black cattle [38]. Translated locations of their markers to ARS-UCD2.0, would place their range at chr6:37,278,524-37,869,348, pruning roughly 144-kb from the region furthest upstream of *LCORL*. This recombination can be traced to sire C in their study. While the methods and reference genome then were different from today, it would be interesting to see how the haplotype presents in Japanese Black cattle, or if other recombination events can be found.

Setting aside the coding sequence mutations, the variants located in ATAC peaks, while perhaps not as straightforward an answer, are worthy of consideration as well. It has been shown previously that changes in *LCORL* expression have been linked both to this haplotype [27], as well as to feed intake [94]. Thus, changes in regulation of *LCORL* may be contributing to this phenotype as well. Overall, changes to transcription factor binding trended toward loss of TF affinity in the *Q* haplotype; there are 37 TFs exclusive to *qq,* 22 exclusive to *QQ*, and 46 shared between the two haplotypes.

Though the distal region around chr6:37,900,000 would be excluded by the findings from Setoguchi et al., this region is fixed in the *Q* haplotype in the population genotyped in the present study. As this Charolais population is the same as those used in a study by Martins Rodrigues [27] to demonstrate increased LCORL expression in *QQ* individuals, it may still be worth considering the effects on transcription from this region. Site L is of particular interest, as it showed some of the strongest changes in TF affinity, and all resulted in a predicted loss of TF binding (Table S2) in *QQ.* Additionally, this region is among the ATAC sites with a stronger signal and also is an embryonic-exclusive peak, suggesting a potential role in early developmental regulation. The most impacted transcription factors among these were THRA and THRB, thyroid hormone receptors. Thyroid hormone is important for normal embryonic development [120], and the presence of thyroid hormone receptor binding sites implies regulation of the genes at this locus may be thyroid hormone sensitive. However, the directionality of expression in response to T3 cannot be inferred from sequence alone, as these receptors can promote or repress transcription in the presence of T3, depending on the other proteins involved in the com-plex at the locus [121]. The TF with the highest affinity for this site, SRF, is known for its important role in both development and skeletal muscle accretion [122, 123], but similarly has fairly complicated and nuanced activity. SRF is most commonly known for promoting expression of its target genes in response to growth factor stimulation [124], but it

also can have a repressive action in competing with other transcription factors for binding sites [125].

In the case of region G-1, the most impactful changes to TF binding are the loss of affinity for the early growth response (EGR) transcription factors, and the gain of FOXA1 binding. EGR1 and EGR2 are crucial regulators of cellular proliferation and apoptosis [126]. EGR1 has been demonstrated to activate pro-apoptotic and pro-survival pathways, again, depending on the context of the cell's status as a whole [127]. FOXA1 is actually able to act as a 'pioneer factor', binding to condensed regions of chromatin and promoting the opening and transcription of previously inaccessible regions of the chromosome, though this is also dependent on the epigenetic marks of the histone as well as the sequence [128]. Thus, this gain of FOXA1 affinity interestingly suggests that individuals with the *QQ* haplotype may be able to promote *LCORL* transcription when it may otherwise be inaccessible for expression. Transcription factor activity ultimately relies on coordinated activity of likely many cofactors interacting with each transcription factor associated with these binding sites, making it difficult to predict direct impacts caused by any individual variants by sequence alone. However, it does demonstrate these changes likely affect the TF binding environment in some way and could warrant further investigation into these changes and their impact on transcription in the region and phenotype as a whole.

Conclusions

The *LCORL-NCAPG* locus a critical region for growth in beef cattle, as evidenced by the findings presented here and the overwhelming body of evidence in the literature. This study has identified 218 mutations exclusive to the haplotype in the region associated with increased growth, including a structural variant in *NCAPG*, a frameshift variant causing a loss of function of the long isoform of *LCORL*, and several mutations affecting transcription factor binding and

thus potential regulation of genes in this locus. Genotyping for some of these variants showed statistically significant associations for birth weight, carcass weight, and average daily gain, making them useful markers for selection and prediction of performance. Though the true causative variant has yet to be determined due to the extensive LD in this region, these findings further clarify details underpinning this region and hopefully can contribute to future studies investigating how this locus mediates its effects.

List of References

1. OECD, FAO: **OECD-FAO Agricultural Outlook 2022-2031**; 2022.
2. Asem-Hiablie S, Battagliese T, Stackhouse-Lawson KR, Alan Rotz C: **A life cycle assessment of the environmental impacts of a beef system in the USA**. *The International Journal of Life Cycle Assessment* 2019, **24**(3):441-455.
3. Mekonnen MM, Hoekstra AY: **A Global Assessment of the Water Footprint of Farm Animal Products**. *Ecosystems* 2012, **15**(3):401-415.
4. de Vries M, de Boer IJM: **Comparing environmental impacts for livestock products: A review of life cycle assessments**. *Livestock Science* 2010, **128**(1):1-11.
5. Myers N, Kent J: **New consumers: The influence of affluence on the environment**. *Proceedings of the National Academy of Sciences* 2003, **100**(8):4963-4968.
6. Peñagaricano F: **Chapter 6 - Genetics and genomics of dairy cattle**. In: *Animal Agriculture.* Edited by Bazer FW, Lamb GC, Wu G: Academic Press; 2020: 101-119.
7. Zuidhof MJ, Schneider BL, Carney VL, Korver DR, Robinson FE: **Growth, efficiency, and yield of commercial broilers from 1957, 1978, and 2005**. *Poultry Science* 2014, **93**(12):2970-2982.
8. White RR, Capper JL: **An environmental, economic, and social assessment of improving cattle finishing weight or average daily gain within U.S. beef production**. *Journal of Animal Science* 2013, **91**(12):5801-5812.
9. Zaborski D, Grzesiak W, Szatkowska I, Dybus A, Muszynska M, Jedrzejczak M: **Factors Affecting Dystocia in Cattle**. *Reproduction in Domestic Animals* 2009, **44**(3):540-551.
10. Bourdon RM, Brinks JS: **Genetic, Environmental and Phenotypic Relationships among Gestation Length, Birth Weight, Growth Traits and Age at First Calving in Beef Cattle**. *Journal of Animal Science* 1982, **55**(3):543-553.
11. Pryce JE, Hayes BJ, Bolormaa S, Goddard ME: **Polymorphic Regions Affecting Human Height Also Control Stature in Cattle**. *Genetics* 2011, **187**(3):981-984.
12. Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, Sugimoto Y, Takasuga A: **Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle**. *BMC Genetics* 2012, **13**(1):40.
13. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M *et al*: **Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals**. *Nature Genetics* 2018, **50**(3):362-367.
14. Purfield DC, Evans RD, Berry DP: **Reaffirmation of known major genes and the identification of novel candidate genes associated with carcass-related metrics based on whole genome sequence within a large multi-breed cattle population**. *BMC Genomics* 2019, **20**(1):720.
15. Smith JL, Wilson ML, Nilson SM, Rowan TN, Oldeschulte DL, Schnabel RD, Decker JE, Seabury CM: **Genome-wide association and genotype by environment interactions for growth traits in U.S. Gelbvieh cattle**. *BMC Genomics* 2019, **20**(1):926.
16. Zhang F, Wang Y, Mukiibi R, Chen L, Vinsky M, Plastow G, Basarab J, Stothard P, Li C: **Genetic architecture of quantitative traits in beef cattle revealed by genome wide association studies of imputed whole genome sequence variants: I: feed efficiency and component traits**. *BMC Genomics* 2020, **21**(1):36.
17. Smith JL, Wilson ML, Nilson SM, Rowan TN, Schnabel RD, Decker JE, Seabury CM: **Genome-wide association and genotype by environment interactions for growth traits in U.S. Red Angus cattle**. *BMC Genomics* 2022, **23**(1):517.

18. Schrooten C, Bovenhuis H, Coppieters W, Van Arendonk JAM: **Whole Genome Scan to Detect Quantitative Trait Loci for Conformation and Functional Traits in Dairy Cattle**. *Journal of Dairy Science* 2000, **83**(4):795-806.
19. Casas E, Shackelford SD, Keele JW, Stone RT, Kappes SM, Koohmaraie M: **Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin2**. *Journal of Animal Science* 2000, **78**(3):560-569.
20. Kneeland J, Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Hansen C, Moore SS: **Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of Bos taurus1**. *Journal of Animal Science* 2004, **82**(12):3405-3414.
21. Maltecca C, Weigel KA, Khatib H, Cowan M, Bagnato A: **Whole-genome scan for quantitative trait loci associated with birth weight, gestation length and passive immune transfer in a Holstein × Jersey crossbred population**. *Animal Genetics* 2009, **40**(1):27-34.
22. Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, Klopp N, Fürbass R, Weikard R, Kühn C: **Dissection of Genetic Factors Modulating Fetal Growth in Cattle Indicates a Substantial Role of the Non-SMC Condensin I Complex, Subunit G (NCAPG) Gene**. *Genetics* 2009, **183**(3):951-964.
23. Gutiérrez-Gil B, Williams JL, Homer D, Burton D, Haley CS, Wiener P: **Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle1**. *Journal of Animal Science* 2009, **87**(1):24-36.
24. Gutiérrez-Gil B, Wiener P, Williams JL, Haley CS: **Investigation of the genetic architecture of a bone carcass weight QTL on BTA6**. *Animal Genetics* 2012, **43**(6):654-661.
25. Markey A: **Mapping of monogenic and quantitative trait loci using a whole genome scan approach and single nucleotide polymorphism platforms**. *Dissertation.* University of Illinois at Urbana-Champaign; 2013.
26. Sahana G, Höglund JK, Guldbrandtsen B, Lund MS: **Loci associated with adult stature also affect calf birth survival in cattle**. *BMC Genetics* 2015, **16**(1):47.
27. Martins Rodrigues F: **Transcriptional variation in muscle from cattle with alternative NCAPG/LCORL QTL genotypes**. *Thesis.* University of Illinois; 2017.
28. Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, Sonstegard TS, Thallman RM, Bennett GL: **Genome-wide association study of growth in crossbred beef cattle**. *Journal of Animal Science* 2010, **88**(3):837-848.
29. Weng Z, Su H, Saatchi M, Lee J, Thomas MG, Dunkelberger JR, Garrick DJ: **Genome-wide association study of growth and body composition traits in Brangus beef cattle**. *Livestock Science* 2016, **183**:4-11.
30. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes BJ, Goddard ME: **A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle**. *PLOS Genetics* 2014, **10**(3):e1004198.
31. Keogh K, Carthy TR, McClure MC, Waters SM, Kenny DA: **Genome-wide association study of economically important traits in Charolais and Limousin beef cows**. *Animal* 2021, **15**(1):100011.
32. Takasuga A, Watanabe T, Mizoguchi Y, Hirano T, Ihara N, Takano A, Yokouchi K, Fujikawa A, Chiba K, Kobayashi N *et al*: **Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping**. *Mammalian Genome* 2007, **18**(2):125-136.

33. Lindholm-Perry AK, Sexten AK, Kuehn LA, Smith TPL, King DA, Shackelford SD, Wheeler TL, Ferrell CL, Jenkins TG, Snelling WM *et al*: **Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle**. *BMC Genetics* 2011, **12**(1):103.
34. Zhang W, Li J, Guo Y, Zhang L, Xu L, Gao X, Zhu B, Gao H, Ni H, Chen Y: **Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle**. *Scientific Reports* 2016, **6**(1):38073.
35. Nkrumah JD, Sherman EL, Li C, Marques E, Crews DH, Jr., Bartusiak R, Murdoch B, Wang Z, Basarab JA, Moore SS: **Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle**. *Journal of Animal Science* 2007, **85**(12):3170-3181.
36. Han YJ, Chen Y, Liu Y, Liu XL: **Sequence variants of the LCORL gene and its association with growth and carcass traits in Qinchuan cattle in China**. *Journal of Genetics* 2017, **96**(1):9-17.
37. Chang T, Xia J, Xu L, Wang X, Zhu B, Zhang L, Gao X, Chen Y, Li J, Gao H: **A genome-wide association study suggests several novel candidate genes for carcass traits in Chinese Simmental beef cattle**. *Animal Genetics* 2018, **49**(4):312-316.
38. Setoguchi K, Furuta M, Hirano T, Nagao T, Watanabe T, Sugimoto Y, Takasuga A: **Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate**. *BMC Genet* 2009, **10**:43.
39. Xia J, Fan H, Chang T, Xu L, Zhang W, Song Y, Zhu B, Zhang L, Gao X, Chen Y *et al*: **Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle**. *Scientific Reports* 2017, **7**(1):42048.
40. Niu Q, Zhang T, Xu L, Wang T, Wang Z, Zhu B, Gao X, Chen Y, Zhang L, Gao H *et al*: **Identification of Candidate Variants Associated With Bone Weight Using Whole Genome Sequence in Beef Cattle**. *Frontiers in Genetics* 2021, **12**.
41. Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C: **Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight**. *Genetics Selection Evolution* 2015, **47**(1):66.
42. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, Decker B, Parker HG, Ostrander EA: **Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology**. *Nature Communications* 2019, **10**(1):1489.
43. Helgeland Ø, Vaudel M, Juliusson PB, Lingaas Holmen O, Juodakis J, Bacelis J, Jacobsson B, Lindekleiv H, Hveem K, Lie RT *et al*: **Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth**. *Nature Communications* 2019, **10**(1):4448.
44. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, Hennig BJ, Bradfield JP, St Pourcain B, Evans DM, Charoen P *et al*: **New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism**. *Nat Genet* 2013, **45**(1):76-82.
45. Lin Y-J, Liao W-L, Wang C-H, Tsai L-P, Tang C-H, Chen C-H, Wu J-Y, Liang W-M, Hsieh A-R, Cheng C-F *et al*: **Association of human height-related genetic variants**

with familial short stature in Han Chinese in Taiwan. *Scientific Reports* 2017, **7**(1):6372.
46. Liu JZ, Medland SE, Wright MJ, Henders AK, Heath AC, Madden PA, Duncan A, Montgomery GW, Martin NG, McRae AF: **Genome-wide association study of height and body mass index in Australian twin families**. *Twin Res Hum Genet* 2010, **13**(2):179-193.
47. Paternoster L, Howe LD, Tilling K, Weedon MN, Freathy RM, Frayling TM, Kemp JP, Smith GD, Timpson NJ, Ring SM *et al*: **Adult height variants affect birth length and growth rate in children**. *Hum Mol Genet* 2011, **20**(20):4069-4075.
48. Shriner D, Adeyemo A, Gerry NP, Herbert A, Chen G, Doumatey A, Huang H, Zhou J, Christman MF, Rotimi CN: **Transferability and Fine-Mapping of Genome-Wide Associated Loci for Adult Height across Human Populations**. *PLOS ONE* 2009, **4**(12):e8398.
49. Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, Hammond N, Stolk L, Nica A, Inouye M, Hofman A *et al*: **Meta-Analysis of Genome-Wide Scans for Human Adult Stature Identifies Novel Loci and Associations with Measures of Skeletal Frame Size**. *PLOS Genetics* 2009, **5**(4):e1000445.
50. Sovio U, Bennett AJ, Millwood IY, Molitor J, O'Reilly PF, Timpson NJ, Kaakinen M, Laitinen J, Haukka J, Pillas D *et al*: **Genetic Determinants of Height Growth Assessed Longitudinally from Infancy to Adulthood in the Northern Finland Birth Cohort 1966**. *PLOS Genetics* 2009, **5**(3):e1000409.
51. van der Valk RJP, Kreiner-Møller E, Kooijman MN, Guxens M, Stergiakouli E, Sääf A, Bradfield JP, Geller F, Hayes MG, Cousminer DL *et al*: **A novel common variant in DCST2 is associated with length in early life and height in adulthood**. *Human Molecular Genetics* 2014, **24**(4):1155-1168.
52. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS *et al*: **Genome-wide association analysis identifies 20 loci that influence adult height**. *Nat Genet* 2008, **40**(5):575-583.
53. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z *et al*: **Defining the role of common variation in the genomic and biological architecture of adult human height**. *Nature Genetics* 2014, **46**(11):1173-1186.
54. Zhao J, Li M, Bradfield JP, Zhang H, Mentch FD, Wang K, Sleiman PM, Kim CE, Glessner JT, Hou C *et al*: **The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature**. *BMC Medical Genetics* 2010, **11**(1):96.
55. Hayward JJ, White ME, Boyle M, Shannon LM, Casal ML, Castelhano MG, Center SA, Meyers-Wallen VN, Simpson KW, Sutter NB *et al*: **Imputation of canine genotype array data using 365 whole-genome sequences improves power of genome-wide association studies**. *PLOS Genetics* 2019, **15**(9):e1008003.
56. Momozawa Y, Merveille A-C, Battaille G, Wiberg M, Koch J, Willesen JL, Proschowsky HF, Gouni V, Chetboul V, Tiret L *et al*: **Genome wide association study of 40 clinical measurements in eight dog breeds**. *Scientific Reports* 2020, **10**(1):6520.
57. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, Seppälä EH, Hansen MST, Lawley CT *et al*: **Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping**. *PLOS Genetics* 2011, **7**(10):e1002316.
58. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, Loredo AI, Bellone RR, Mezey JG, Brooks SA *et al*: **Four Loci Explain 83% of Size Variation in the Horse**. *PLOS ONE* 2012, **7**(7):e39929.

59. Metzger J, Schrimpf R, Philipp U, Distl O: **Expression Levels of LCORL Are Associated with Body Size in Horses**. *PLOS ONE* 2013, **8**(2):e56497.
60. Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, Rieder S: **A Genome-Wide Association Study Reveals Loci Influencing Height and Other Conformation Traits in Horses**. *PLOS ONE* 2012, **7**(5):e37282.
61. Staiger EA, Abri MAA, Pflug KM, Kalla SE, Ainsworth DM, Miller D, Raudsepp T, Sutter NB, Brooks SA: **Skeletal variation in Tennessee Walking Horses maps to the LCORL/NCAPG gene region**. *Physiological Genomics* 2016, **48**(5):325-335.
62. Tetens J, Widmann P, Kühn C, Thaller G: **A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses**. *Animal Genetics* 2013, **44**(4):467-471.
63. Tozaki T, Sato F, Ishimaru M, Kikuchi M, Kakoi H, Hirota K-i, Nagata S-i: **Sequence variants of BIEC2-808543 near LCORL are associated with body composition in Thoroughbreds under training**. *Journal of Equine Science* 2016, **27**(3):107-114.
64. La Y, Zhang X, Li F, Zhang D, Li C, Mo F, Wang W: **Molecular Characterization and Expression of SPP1, LAP3 and LCORL and Their Association with Growth Traits in Sheep**. *Genes* 2019, **10**(8):616.
65. Posbergh CJ, Huson HJ: **All sheeps and sizes: a genetic investigation of mature body size across sheep breeds reveals a polygenic nature**. *Animal Genetics* 2021, **52**(1):99-107.
66. Graber JK, Signer-Hasler H, Burren A, Drögemüller C: **Evaluation of truncating variants in the LCORL gene in relation to body size of goats from Switzerland**. *Anim Genet* 2022, **53**(2):237-239.
67. Khalid A, Ashraf H, Asim H, Ayman M, Saif R: **Association Analysis of LCORL Genetic Variant (rs657074013A>AT) with Body-Height and Skeletal Frame in Pakistani Goats**. *bioRxiv* 2022:2022.2010.2009.511453.
68. Saif R, Henkel J, Jagannathan V, Drögemüller C, Flury C, Leeb T: **The LCORL Locus Is under Selection in Large-Sized Pakistani Goat Breeds**. *Genes* 2020, **11**(2):168.
69. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg Ö, Jern P, Jørgensen CB *et al*: **Strong signatures of selection in the domestic pig genome**. *Proceedings of the National Academy of Sciences* 2012, **109**(48):19529-19536.
70. Guo J, Qu L, Dou T-C, Shen M-M, Hu Y-P, Ma M, Wang K-H: **Genome-wide association study provides insights into the genetic architecture of bone size and mass in chickens**. *Genome* 2020, **63**(3):133-143.
71. Liu J, Zhou J, Li J, Bao H: **Identification of candidate genes associated with slaughter traits in F2 chicken population using genome-wide association study**. *Animal Genetics* 2021, **52**(4):532-535.
72. Liu R, Sun Y, Zhao G, Wang F, Wu D, Zheng M, Chen J, Zhang L, Hu Y, Wen J: **Genome-Wide Association Study Identifies Loci and Candidate Genes for Body Composition and Meat Quality Traits in Beijing-You Chickens**. *PLOS ONE* 2013, **8**(4):e61172.
73. Lyu S, Arends D, Nassar MK, Brockmann GA: **Fine mapping of a distal chromosome 4 QTL affecting growth and muscle mass in a chicken advanced intercross line**. *Animal Genetics* 2017, **48**(3):295-302.
74. Carneiro M, Hu D, Archer J, Feng C, Afonso S, Chen C, Blanco-Aguiar JA, Garreau H, Boucher S, Ferreira PG *et al*: **Dwarfism and Altered Craniofacial Development in Rabbits Is Caused by a 12.1 kb Deletion at the HMGA2 Locus**. *Genetics* 2017, **205**(2):955-965.

75. Krutikova A, Dementieva N, Nikitkina E: **PSXIV-23 Polymorphism of the LCORL gene in the reindeer**. *Journal of Animal Science* 2021, **99**(Supplement_3):256-257.
76. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG: **Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings**. *PLOS Genetics* 2006, **2**(3):e41.
77. Bongiorni S, Mancini G, Chillemi G, Pariset L, Valentini A: **Identification of a Short Region on Chromosome 6 Affecting Direct Calving Ease in Piedmontese Cattle Breed**. *PLOS ONE* 2012, **7**(12):e50137.
78. Yurchenko AA, Deniskova TE, Yudin NS, Dotsev AV, Khamiruev TN, Selionova MI, Egorov SV, Reyer H, Wimmers K, Brem G *et al*: **High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia**. *BMC Genomics* 2019, **20**(3):294.
79. Gutierrez-Gil B, Williams JL, Homer D, Burton D, Haley CS, Wiener P: **Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle**. *J Anim Sci* 2009, **87**(1):24-36.
80. Setoguchi K, Watanabe T, Weikard R, Albrecht E, Kühn C, Kinoshita A, Sugimoto Y, Takasuga A: **The SNP c.1326T>G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle**. *Animal Genetics* 2011, **42**(6):650-655.
81. Doyle JL, Berry DP, Veerkamp RF, Carthy TR, Walsh SW, Evans RD, Purfield DC: **Genomic Regions Associated With Skeletal Type Traits in Beef and Dairy Cattle Are Common to Regions Associated With Carcass Traits, Feed Intake and Calving Difficulty**. *Frontiers in Genetics* 2020, **11**.
82. Consortium TU: **UniProt: the Universal Protein Knowledgebase in 2023**. *Nucleic Acids Research* 2022, **51**(D1):D523-D531.
83. Murphy LA, Sarge KD: **Phosphorylation of CAP-G is required for its chromosomal DNA localization during mitosis**. *Biochemical and Biophysical Research Communications* 2008, **377**(3):1007-1011.
84. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S *et al*: **Database resources of the national center for biotechnology information**. *Nucleic Acids Res* 2022, **50**(D1):D20-d26.
85. Conway E, Jerman E, Healy E, Ito S, Holoch D, Oliviero G, Deevy O, Glancy E, Fitzpatrick DJ, Mucha M *et al*: **A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities**. *Mol Cell* 2018, **70**(3):408-421.e408.
86. Laugesen A, Højfeldt JW, Helin K: **Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation**. *Mol Cell* 2019, **74**(1):8-18.
87. Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappsilber J, Lerdrup M, Helin K: **A model for transmission of the H3K27me3 epigenetic mark**. *Nature Cell Biology* 2008, **10**(11):1291-1300.
88. Margueron R, Reinberg D: **The Polycomb complex PRC2 and its mark in life**. *Nature* 2011, **469**(7330):343-349.
89. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM: **The many faces of the helix-turn-helix domain: Transcription regulation and beyond⋆**. *FEMS Microbiology Reviews* 2005, **29**(2):231-262.
90. Sakaguchi C, Ichihara K, Nita A, Katayama Y, Nakayama KI: **Identification and characterization of novel proteins associated with CHD4**. *Genes to Cells* 2022, **27**(1):61-71.
91. Bray SJ: **Notch signalling: a simple pathway becomes complex**. *Nature Reviews Molecular Cell Biology* 2006, **7**(9):678-689.

92. Denslow SA, Wade PA: **The human Mi-2/NuRD complex and gene regulation**. *Oncogene* 2007, **26**(37):5433-5438.
93. O'Shaughnessy A, Hendrich B: **CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now**. *Biochemical Society Transactions* 2013, **41**(3):777-782.
94. Lindholm-Perry AK, Kuehn LA, Oliver WT, Sexten AK, Miles JR, Rempel LA, Cushman RA, Freetly HC: **Adipose and Muscle Tissue Gene Expression of Two Genes (NCAPG and LCORL) Located in a Chromosomal Region Associated with Cattle Feed Intake and Gain**. *PLOS ONE* 2013, **8**(11):e80882.
95. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME *et al*: **Copy number variation: new insights in genome diversity**. *Genome Res* 2006, **16**(8):949-961.
96. Geistlinger L, da Silva VH, Cesar ASM, Tizioto PC, Waldron L, Zimmer R, Regitano LCdA, Coutinho LL: **Widespread modulation of gene expression by copy number variation in skeletal muscle**. *Scientific Reports* 2018, **8**(1):1399.
97. Rafter P, Gormley IC, Purfield D, Parnell AC, Naderi S, Berry DP: **Genome-wide association analyses of carcass traits using copy number variants and raw intensity values of single nucleotide polymorphisms in cattle**. *BMC Genomics* 2021, **22**(1):757.
98. Batcher K, Varney S, Raudsepp T, Jevit M, Dickinson P, Jagannathan V, Leeb T, Bannasch D: **Ancient segmentally duplicated LCORL retrocopies in equids**. *PLOS ONE* 2023, **18**(6):e0286861.
99. Zhao G, Liu Y, Niu Q, Zheng X, Zhang T, Wang Z, Xu L, Zhu B, Gao X, Zhang L *et al*: **Runs of homozygosity analysis reveals consensus homozygous regions affecting production traits in Chinese Simmental beef cattle**. *BMC Genomics* 2021, **22**(1):678.
100. Wiles ET, Selker EU: **H3K27 methylation: a promiscuous repressive chromatin mark**. *Current Opinion in Genetics & Development* 2017, **43**:31-37.
101. Miller SA, Dykes DD, Polesky HF: **A simple salting out procedure for extracting DNA from human nucleated cells**. *Nucleic Acids Research* 1988, **16**(3):1215-1215.
102. Andrews S: **FastQC: A Quality Control Tool for High Thoroughput Sequence Data**. In. [Online]; 2010.
103. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
104. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. *arXiv preprint arXiv:13033997* 2013.
105. Van der Auwera GA OCB: **Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)**: O'Reilly Media; 2020.
106. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F: **The Ensembl Variant Effect Predictor**. *Genome Biology* 2016, **17**(1):122.
107. Zeng T, Li YI: **Predicting RNA splicing from DNA sequence using Pangolin**. *Genome Biology* 2022, **23**(1):103.
108. Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D *et al*: **Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser**. *Bioinformatics* 2013, **30**(7):1003-1005.
109. Yuan C, Tang L, Lopdell T, Petrov VA, Oget-Ebrad C, Moreira GCM, Gualdrón Duarte JL, Sartelet A, Cheng Z, Salavati M *et al*: **An organism-wide ATAC-seq peak catalog for the bovine and its use to identify regulatory variants**. *Genome Res* 2023, **33**(10):1848-1864.

110. Shen WK, Chen SY, Gan ZQ, Zhang YZ, Yue T, Chen MM, Xue Y, Hu H, Guo AY: **AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations**. *Nucleic Acids Res* 2023, **51**(D1):D39-d45.
111. Grant CE, Bailey TL, Noble WS: **FIMO: scanning for occurrences of a given motif**. *Bioinformatics* 2011, **27**(7):1017-1018.
112. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K *et al*: **Determination and inference of eukaryotic transcription factor sequence specificity**. *Cell* 2014, **158**(6):1431-1443.
113. Bates D, Mächler M, Bolker B, Walker S: **Fitting Linear Mixed-Effects Models Using lme4**. *Journal of Statistical Software* 2015, **67**(1):1 - 48.
114. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing**. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995, **57**(1):289-300.
115. Hu X, Xing Y, Fu X, Yang Q, Ren L, Wang Y, Li Q, Li J, Zhang L: **NCAPG Dynamically Coordinates the Myogenesis of Fetal Bovine Tissue by Adjusting Chromatin Accessibility**. *International Journal of Molecular Sciences* 2020, **21**(4):1248.
116. Sanchez M-P, Tribout T, Kadri NK, Chitneedi PK, Maak S, Hozé C, Boussaha M, Croiseau P, Philippe R, Spengeler M *et al*: **Sequence-based GWAS meta-analyses for beef production traits**. *Genetics Selection Evolution* 2023, **55**(1):70.
117. Cai Z, Wu X, Thomsen B, Lund MS, Sahana G: **Genome-wide association study identifies functional genomic variants associated with young stock survival in Nordic Red Dairy Cattle**. *Journal of Dairy Science* 2023, **106**(11):7832-7845.
118. Gualdrón Duarte JL, Yuan C, Gori A-S, Moreira GCM, Takeda H, Coppieters W, Charlier C, Georges M, Druet T: **Sequenced-based GWAS for linear classification traits in Belgian Blue beef cattle reveals new coding variants in genes regulating body size in mammals**. *Genetics Selection Evolution* 2023, **55**(1):83.
119. Horvath S, Lu AT, Haghani A, Zoller JA, Li CZ, Lim AR, Brooke RT, Raj K, Serres-Armero A, Dreger DL *et al*: **DNA methylation clocks for dogs and humans**. *Proceedings of the National Academy of Sciences* 2022, **119**(21):e2120887119.
120. Ashkar FA, Semple E, Schmidt CH, St. John E, Bartlewski PM, King WA: **Thyroid hormone supplementation improves bovine embryo development in vitro**. *Human Reproduction* 2009, **25**(2):334-344.
121. Wondisford FE: **Chapter 77 - Thyroid Hormone Action**. In: *Endocrinology: Adult and Pediatric (Seventh Edition).* Edited by Jameson JL, De Groot LJ, de Kretser DM, Giudice LC, Grossman AB, Melmed S, Potts JT, Weir GC. Philadelphia: W.B. Saunders; 2016: 1336-1349.e1333.
122. Vlahopoulos S, Zimmer WE, Jenster G, Belaguli NS, Balk SP, Brinkmann AO, Lanz RB, Zoumpourlis VC, Schwartz RJ: **Recruitment of the Androgen Receptor via Serum Response Factor Facilitates Expression of a Myogenic Gene ***. *Journal of Biological Chemistry* 2005, **280**(9):7786-7792.
123. Schratt G, Philippar U, Hockemeyer D, Schwarz H, Alberti S, Nordheim A: **SRF regulates Bcl-2 expression and promotes cell survival during murine embryonic development**. *The EMBO Journal* 2004, **23**(8):1834-1844-1844.
124. Miano JM: **Role of serum response factor in the pathogenesis of disease**. *Laboratory Investigation* 2010, **90**(9):1274-1284.
125. Lee HJ, Yun CH, Lim SH, Kim BC, Baik KG, Kim JM, Kim WH, Kim SJ: **SRF is a nuclear repressor of Smad3-mediated TGF-β signaling**. *Oncogene* 2007, **26**(2):173-185.
126. Kumbrink J, Kirsch KH, Johnson JP: **EGR1, EGR2, and EGR3 activate the expression of their coregulator NAB2 establishing a negative feedback loop in cells of**

neuroectodermal and epithelial origin. *Journal of Cellular Biochemistry* 2010, **111**(1):207-217.
127. Baron V, Adamson ED, Calogero A, Ragona G, Mercola D: **The transcription factor Egr1 is a direct regulator of multiple tumor suppressors including TGFβ1, PTEN, p53, and fibronectin**. *Cancer Gene Therapy* 2006, **13**(2):115-124.
128. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M: **FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription**. *Cell* 2008, **132**(6):958-970.

Vita

Leif Majeres was raised in Orland Park, Illinois. He attended University of Illinois in Urbana-Champaign, where he obtained a Bachelor of Science degree in Animal Science with a Science, Pre-Veterinary and Medical Concentration. During his time at University of Illinois, Leif completed and presented his undergraduate research project evaluating algae grown as part of a carbon-capture program as a feedstuff for beef cattle. While his first research project was in nutrition, his primary research interest is in genetic improvement of livestock and molecular mechanisms underlying production phenotypes such as growth. This led him to choose to attend University of Tennessee, Knoxville to pursue a Master of Science degree with Dr. Jon Beever. After completing his Masters, he will be remaining in Dr. Beever's lab as a PhD student to continue his research on *LCORL*.