Universität Rostock

Traditio et Innovatio

Aus der Professur für Genetik der Krankheitsresistenz
der Agrar- und Umweltwissenschaftlichen Fakultät

Identifizierung und Charakterisierung der Rolle langer nichtkodierender RNA (lncRNA) bei der Genregulation von Stoffwechselprozessen beim Rind

Kumulative Dissertation

zur Erlangung des akademischen Grades
Doktor der Agrarwissenschaften (doctor agriculturae)

an der Agrar- und Umweltwissenschaftlichen Fakultät
der Universität Rostock

vorgelegt von M.Sc. Wietje Nolte,
geboren am 27.09.1990 in der Hansestadt Stralsund

Moritzburg, den 20.12.2020

https://doi.org/10.18453/rosdok_id00003131

Gutachter:

Prof. Dr. Christa Kühn
Universität Rostock, Agrar- und Umweltwissenschaftliche Fakultät
Forschungsinstitut für Nutztierbiologie, Institut für Genombiologie

Prof. Dr. Klaus Wimmers
Universität Rostock, Agrar- und Umweltwissenschaftliche Fakultät
Forschungsinstitut für Nutztierbiologie, Institut für Genombiologie

Prof. Dr. Georg Thaller
Christian-Albrechts-Universität zu Kiel, Institut für Tierzucht und Tierhaltung

Prof. Dr. Jens Tetens
Georg-August-Universität Göttingen, Department für Nutztierwissenschaften

Jahr der Einreichung: 2020

Jahr der Verteidigung: 2021

meiner Mutter

Inhaltsverzeichnis

Abkürzungsverzeichnis

%	Prozentsatz
ACC	accretion (Akkretionstyp)
bp	Basenpaare
BTA	*Bos taurus* Autosom
cDNA	complementary DNA
ChIRP	Chromatin Isolation by RNA Purification
CFC	carcass fat content (Fettgehalt im Schlachtkörper)
DA	differenziell abundant
DE	differenziell exprimiert
DEA	Differenzielle Expressionsanalyse
DNA	Desoxyribonukleinsäure
ECM	Energiekorrigierte Milchleistung
F_1	Erste Kreuzungsstufe
F_2	Zweite Kreuzungsstufe
FBN	Leibniz-Institut für Nutztierbiologie
FC	Fold Change
FPKM	Fragments per Kilobase of Transcript per Million Reads
GO	Gene Ontology
GSEA	Gene Set Enrichment Analyse
GWAS	genomweite Assoziationsstudie
IMF	intramuskulärer Fettgehalt
IPA	Ingenuity Pathway Analysis
Kb	Kilobasen
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	lange nichtkodierende RNA
Mb	Megabasen
mRNA	Messenger-RNA
miRNA	Mikro-Ribonukleinsäure
MY	Milk Yield (Milchleistung)
NAT	natural antisense transcript
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
nt	Nukleotid
ORF	Open Reading Frame (offener Leserahmen)
p	p-Wert (Signifikanz)

P0	Parentalgeneration
PCA	Principal Component Analysis (Hauptkomponentenanalyse)
PCIT	Partial Correlation and Information Theory
PI	Pathway Impact
QTL	Quantitative Trait Locus (quantitativer Merkmalslokus)
r	Korrelationskoeffizient
RIF	Regulatory Impact Factor
RFI	Residual Feed Intake (Residuale Futteraufnahme)
RNA	Ribonukleinsäure
RNA-Seq	RNA-Sequenzierung (Transkriptomsequenzierung)
rRNA	ribosomale RNA
SEC	secretion (Sekretionstyp)
SEGFAM	segregierende Familien
tRNA	transfer RNA

Abbildungsverzeichnis

Tabellenverzeichnis

1. Kurzfassung (Abstract)

Der Großteil der Erbsubstanz von Säugetieren besteht aus genomischen Elementen, die nicht für ein Protein kodieren, darunter die Klasse der langen nichtkodierenden RNA (lncRNA), deren Rolle bei der Merkmalsausprägung von Nutztieren bisher wenig erforscht ist. In dieser Dissertation wurden in einem integrativen Ansatz das Transkriptom, Phänom und Metabolom von Bullen (N = 24) und Kühen (N = 24) einer Kreuzungspopulation (Fleischrind x Milchvieh) der zweiten Generation analysiert, um neue, unbekannte lncRNAs und ihre möglichen biologischen Funktionen und ihren Einfluss auf die Nährstoffverteilung zu ermitteln. Die Rinder wurden, basierend auf Plasmametaboliten (N = 640), ihrem Fettansatz und ihrer residualen Futteraufnahme (Bullen) oder Milchleistung (Kühe) in zwei verschiedene Stoffwechselgruppen eingeteilt. Anhand von Transkriptomdaten beider Geschlechter aus Leber, Jejunum, Skelettmuskel und Pansen wurden, basierend auf den bovinen Referenzgenomen UCD3.1 und ARS-UCD.1.2, projektspezifische Transkriptannotationen erstellt. Bioinformatisch wurden darin 7.646 bzw. 6.161 Transkripte als lncRNAs klassifiziert. LncRNAs zeigten mehrheitlich ein geringeres Expressionsniveau als Genorte, die für Proteine kodieren. Für zahlreiche lncRNAs konnte entweder eine gewebsspezifische Expression, eine differenzielle Expression zwischen Stoffwechselgruppen, und/oder eine Kolokalisation mit Genorten, die mit Futtereffizienz oder Milchleistung assoziiert sind, beobachtet werden. Anhand von Korrelations- und Netzwerkanalysen wurden lncRNAs mit einem hohen genregulatorischen Potenzial und potenziellem Einfluss auf die Zielmerkmale ermittelt. Durch die Integration von Transkriptom- und Metabolomdaten wurden für potenziell regulatorische lncRNAs über Pathway-Enrichment-Analysen biologische Funktionen abgeleitet. LncRNAs waren in diverse Stoffwechselprozesse, wie z.B. den Aminosäure- und Fettstoffwechsel, den Zitrat- und Harnstoffzyklus, die Glukoneogenese, die Fettsäure-β-Oxidation, den PPAR-Signalweg und die Proteinbiosynthese involviert. Die Koexpression mit benachbarten, proteinkodierenden Genen ließ auf eine Koregulation oder eine stabilisierende Funktion der lncRNAs schließen. Eine differenzierte Untersuchung des hepatischen Transkriptoms von Bullen und Kühen indizierte, dass verschiedene biologische Schaltkreise angesprochen waren und unterschiedliche, potenziell regulatorische lncRNAs im Vordergrund standen. Die Ergebnisse haben zur weiterten Annotation des Rindergenoms beigetragen und Kandidaten-lncRNAs für zukünftige Validierungsstudien aufgezeigt.

Abstract

The majority of the mammalian genome consists of genomic elements that do not code for a protein, including the class of long non-coding RNA (lncRNA), whose role in phenotypic modulation in livestock remains largely unknown. In this dissertation, an integrative approach was used to analyse the transcriptome, phenome and metabolome of bulls (N = 24) and cows (N = 24) of a second-generation crossbred population (beef x dairy cattle) to identify novel, unknown lncRNAs and their possible biological functions and influence on nutrient partitioning. The animals were divided into two different metabolic groups based on plasma metabolites (N = 640), their fat accretion and their residual feed intake (bulls) or milk yield (cows). Using transcriptome data of both sexes from liver, jejunum, skeletal muscle and rumen, project-specific transcript annotations were generated based on the bovine reference genomes UCD3.1 and ARS-UCD.1.2. Using bioinformatic tools 7,646 and 6,161 transcripts were classified as lncRNAs respectively. The majority of lncRNAs showed a lower expression level than protein-coding genes. For many lncRNAs, either tissue-specific expression, differential expression between metabolic groups, and/or colocalisation with loci associated with feed efficiency or milk yield were observed. Correlation and network analyses were used to identify lncRNAs with high gene regulatory potential and potential influence on target traits. By integrating transcriptomic and metabolomic data, biological functions were inferred for potentially regulatory lncRNAs via pathway enrichment analyses. LncRNAs were involved in diverse metabolic processes, such as amino acid and lipid metabolism, the citrate and urea cycle, gluconeogenesis, fatty acid-β oxidation, PPAR-signalling and protein biosynthesis. Co-expression with neighbouring protein-coding genes suggested a co-regulatory or stabilising function of the lncRNAs. Differential examination of the hepatic transcriptome of bulls and cows indicated that different biological circuits were activated and different, potentially regulatory lncRNAs were prominent. The results have contributed to the annotation of the bovine genome and provided candidate lncRNAs for future validation studies.

2. Allgemeine Einführung

Das Rind gehört nach Schafen und Ziegen zu den ersten domestizierten Tierarten und dient dem Menschen bis heute gleichzeitig als eine Hauptnahrungsquelle sowie als Zugtier (Lenstra, 2014). Es wird in Hausrinder (*Bos taurus*) und Buckelrinder (*Bos indicus*) unterschieden, die beide vom ausgestorbenen Auerochsen (*Bos primigenius*) abstammen (Ajmone-Marsan et al., 2010). In Europa werden heute über 450 Rinderrassen gehalten, die allesamt zum taurinen Hausrind zählen (Beja-Pereira et al., 2006), während Buckelrinder (Zebus) aufgrund ihrer Anpassung an die klimatischen Verhältnisse insbesondere in ariden und tropischen Regionen der Welt zu finden sind. Seit der Einführung der ersten Herdbücher im 18. Jahrhundert hat sich in Europa eine Vielzahl von Rassen herausgebildet, die gezielt für bestimmte Eigenschaften wie z.B. Milch- oder Fleischleistung selektiert wurden (Felius et al., 2014).

Mit dem Aufkommen der Genomik eröffneten sich auch für die (Milch-) Rinderzucht neue Möglichkeiten, v.a. bei der züchterischen Bearbeitung von Merkmalen, die nur unter hohem Aufwand messbar sind (Miglior et al., 2017). In jüngerer Vergangenheit konnten für das Rind Genorte mit vergleichsweise großen Effekten auf ökonomisch bedeutsame Merkmale wie Milchleistung und Milchzusammensetzung (Diacylglycerol-O-acyltransferase 1, *DGAT* (Grisart et al., 2002)), Fleischansatz (Myostatin, *MSTN* (Grobet et al., 1997)), Körpergröße und Schlachtkörpermerkmalen (Non-SMC Condensin I Complex Subunit G, *NCAPG* & Ligand Dependent Nuclear Receptor Corepressor Like, *LCORL* (Setoguchi et al., 2011), Pleiomorphic adenoma gene 1, *PLAG1* (Littlejohn et al., 2012)) identifiziert werden. Beim Milchvieh hat die Einführung der genomischen Selektion seit Beginn der 2000er Jahre zu einer substanziellen Steigerung des genetisch bedingten Leistungszuwachses bei gleichzeitiger Verkürzung des Generationsintervalls geführt (Doublet et al., 2019). Bei auf Fleischansatz gezüchteten Rassen limitieren vor allem in Deutschland, im Gegensatz zum Milchvieh, die Struktur der Zuchtprogramme, der begrenzte Einsatz von künstlicher Besamung sowie ein Mangel an konsequent durchgeführten Leistungsprüfungen für relevante Merkmale die erfolgreiche Anwendung der genomischen Selektion. Gleichwohl wäre auch hier bei erfolgreicher Implementierung ein genetisch bedingter Leistungsfortschritt von etwa 10 % jährlich zu erwarten (Ibanez-Escriche & Simianer, 2016). Heute und zukünftig liegt der Fokus in der Milchviehzucht jedoch nicht mehr ausschließlich auf einem hohen Produktionsniveau, sondern vermehrt auf Effizienz, Gesundheit und Nachhaltigkeit (Egger-Danner et al., 2014).

Futtereffizienz von Milch- und Fleischrindern hängt neben Umweltfaktoren, wie z.B. der Ernährung, von der genetischen Ausstattung eines Tieres ab (Arthur & Herd, 2008; Blake & Custodio, 1984), die beeinflusst, wie Futter in Milch oder in Fleisch umgewandelt wird. Die Aufklärung der molekulargenetischen Ursachen für die Varianz in der Nährstofftransformation

beim Rind – und damit der unterschiedlichen Stoffwechseltypen – bleibt eine Herausforderung. Basierend auf ihrem Nährstoffumsatz und -ansatz sowie der Nährstoffverwertung bzw. -verteilung lassen sich Rinder in verschiedene Stoffwechseltypen einteilen. Je nach Nutzungsrichtung, d.h. Milch- oder Fleischrind, bieten sich die Parameter Futtereffizienz, residuale Futteraufnahme, tägliche Zunahme, Milchleistung oder Fettansatz zur Charakterisierung an. Insbesondere in der westlichen Gesellschaft wird ein Stoffwechseltyp bevorzugt, der, bei geringem Fettansatz im Schlachtkörper, Futter effizient in Muskelmasse und somit Fleischansatz umwandelt oder in eine hohe Milchleistung umsetzt. Bei Kühen mit hoher Laktationsleistung führt eine stärkere Mobilisation von Fettreserven nach Beginn der Laktation vermehrt zu Stoffwechselproblemen wie z.B. Fettleber (Morrow, 1976).

Während bisher hauptsächlich proteinkodierende Gene im Zentrum der genomischen Forschung standen, häufen sich Indizien dafür, dass auch das nichtkodierende Genom einen wesentlichen Beitrag zur Merkmalsausprägung leistet. Wiederholt wurden auch beim Nutztier beispielsweise Quantitative Merkmalsloci (QTL), die hochsignifikant mit ökonomisch bedeutsamen Merkmalen assoziiert sind, außerhalb von proteinkodierenden Genen im intergenischen bzw. nichtkodierenden Bereich gefunden wurden (Bouwman et al., 2018; Fang & Pausch, 2019; Pausch et al., 2016). Die zwischen eukaryotischen Individuen und Spezies beobachtete phänotypische Variation liegt womöglich in Unterschieden der nichtkodierenden Genomregionen begründet, die durch die Kontrolle der Architektur des Genoms die Expression von Genen regulieren (Mattick, 2001). Zum nichtkodierenden Genom zählt u.a. die Klasse der langen nichtkodierenden RNAs (lncRNAs), die allem voran durch eine minimale Länge von 200 Basenpaaren (bp) und die Abwesenheit von proteinkodierendem Potenzial gekennzeichnet sind (Mercer et al., 2009). Trotzdem ist von einigen lncRNAs bekannt, dass sie für kleine, funktionale Peptide kodieren (z.B. Anderson et al., 2015; Ruiz-Orera et al., 2014).

Während schon erste Studien zur Katalogisierung und Annotation von lncRNAs bei Nutztieren unternommen wurden, sind sie im Vergleich zum Menschen und dem Modelltier Maus bisher unzureichend untersucht (Kosinska-Selbi et al., 2020). Die Identifizierung und funktionelle Charakterisierung von lncRNAs wird zur weiteren Aufklärung der Ursachen phänotypischer Variation einschließlich der ihr zugrundeliegenden physiologischen Prozesse in domestizierten Tieren beitragen (Weikard et al., 2017) und ist unumgänglich, um die Lücke zwischen Phänotyp und Genotyp zu schließen (Kosinska-Selbi et al., 2020).

Ein wichtiges Hilfsmittel zur funktionalen Charakterisierung von lncRNAs und zur weiteren molekularen Beschreibung von Tieren ist ihr Metabolom. Sowohl für die Futtereffizienz von Fleischrindern (Novais et al., 2019) und Milchvieh (Wang & Kadarmideen, 2019) und für den Laktationsbeginn bei Milchkühen (Luo et al., 2019) bzw. für verschiedene Laktationsstadien (Ilves et al., 2012) wurden bereits charakteristische Metabolitenprofile im Blut beschrieben.

Auch Stoffwechselprobleme wie die Ketose sind von Veränderungen der Metabolitenprofile im Blut begleitet (Li et al., 2014), und beispielsweise wurden Serumspiegel von FGF21 als diagnostische Biomarker für Ketose vorgeschlagen (Xu et al., 2016).

Bisherige Studien, die an einer Kreuzungspopulation von Rassen des Fleisch-und Milchtyps durchgeführt wurden, unterstützen die Hypothese, dass divergente Stoffwechseltypen mittels systematischer Untersuchungen auf molekularer Ebene charakterisiert und genetische Schalter und physiologische Regelkreise für divergente Nährstoffverwertung und Leistungsausprägung identifiziert werden können (Weikard et al., 2010; Widmann et al., 2015; Widmann et al., 2013).

3. Zielsetzung

Die Zielrichtung des Forschungsvorhabens bestand in der molekularen Charakterisierung von phänotypisch divergenten Stoffwechseltypen des Rindes. Dafür wurde die Identifizierung und Charakterisierung der Rolle langer nichtkodierender RNA (lncRNA) bei der Genregulation von Stoffwechselprozessen beim Rind anvisiert. Folgende Fragestellungen sollten im Rahmen des Projektes beantwortet werden:

- Ist der lncRNA-Transkriptomkatalog beim Rind mit dem anderer Spezies zu vergleichen?
- Welche Rolle spielt das lncRNA-Transkriptom des Rindes bei der Differenzierung der Nährstoffverteilung?
- Kann einzelnen lncRNAs eine funktionale Rolle bei der Merkmalsdifferenzierung über Einbeziehung von Metabolomdaten und Pathway-Enrichment-Analysen zugewiesen werden?
- Was sind wesentliche biologische Regelkreise, die zur Differenzierung der Närstoffverteilung beitragen?

Zur Beantwortung dieser Fragestellungen wurden von Kühen und Bullen, die phänotypisch durch eine unterschiedliche Nährstoffverteilung charakterisiert waren, vier Gewebe mit hoher Relevanz für den (Nährstoff-) Stoffwechsel in eine Transkriptomsequenzierung einbezogen (siehe Tabelle 1) und umfassende Metabolitenprofile aus dem Plasma für ergänzende Analysen integriert.

Tabelle 1: Auswahl der beprobten und RNA-sequenzierten Gewebe

	Gewebe	Funktion/ Begründung
1	Leber	stoffwechselaktivstes Organ
2	Pansen	Organ der Fermentation der aufgenommenen Nahrung
3	Jejunum	Organ zur Verdauung und Resorption der fermentierten Nahrung
4	Skelettmuskel	größtes metabolisch aktives Organ, u.a. zur Energiespeicherung

4. Literaturübersicht und aktueller Wissensstand

4.1. Transkriptomsequenzierung

Der Begriff Transkriptom umfasst die Gesamtheit aller von der DNA transkribierten RNA-Moleküle (Transkripte) zu einem gegebenen Zeitpunkt, unabhängig davon, ob diese anschließend in ein Protein translatiert werden. Zum Transkriptom zählen damit messenger RNAs (mRNA) von proteinkodierenden Genen, und nichtkodierende RNAs (LaRossa, 2013; National Human Genome Research Institute (NHGRI), 2020) wie Transfer-RNAs, ribosomale RNAs (rRNAs), small nuclear RNAs, small nucleolar RNAs, Mikro-RNAs (miRNAs) oder lange nichtkodierende RNAs (lncRNAs). Durch die unterschiedliche Abundanz von Transkripten ergeben sich spezifische Transkriptionsprofile, die in Abhängigkeit von Zeitpunkt, Entwicklungsstadium, Krankheits- oder Gesundheitszustand, Gewebe, Tierart und Geschlecht variieren können.

Die Transkriptomsequenzierung, die auch als RNA-Sequenzierung (kurz RNA-Seq) bezeichnet wird, wurde in den frühen 2000er Jahren entwickelt (Nagalakshmi et al., 2008). Mit Next-Generation-Sequencing (NGS) [1] -Technologien können bei entsprechender Sequenziertiefe auch Transkripte erfasst werden, die in geringer Menge bzw. Kopienzahl in der Probe vorliegen. Dadurch ist eine holistische Abbildung des Transkriptoms möglich (Martin & Wang, 2011). Mit Etablierung der strangspezifischen RNA-Seq (Parkhomchuk et al., 2009) wurde ein Meilenstein erreicht, da hier die Information, von welchem DNA-Strang ein Transkript abgeschrieben wurde, konserviert wird. Insbesondere für die Detektion von antisense-Transkripten, d.h. von Transkripten, deren zugrundeliegende Loci sich auf den beiden DNA-Strängen gegenüberliegen bzw. überlappen, ist dieses Verfahren eine Voraussetzung.

Die RNA-Seq ermöglicht, im Gegensatz zu cDNA-Mikroarray-Designs, bei denen nur bereits bekannte Gene und Genomregionen erfasst werden, einen agnostischen Ansatz, bei dem alle in der Zelle bzw. im Gewebe transkribierten Genomabschnitte erfasst werden. Dieser Ansatz schafft die Voraussetzungen für die Detektion neuer Loci, da er frei von existierendem Vorwissen über Gene und Loci greift (Martin & Wang, 2011).

Die mehrheitlich gegenwärtig verwendeten NGS-Ansätze ermöglichen noch keine vollständige Sequenzierung einzelner Transkripte in voller Länge, sondern basiert auf der Sequenzierung relativ kurzer Abschnitte („reads“, 35 - 500 bp Länge), die anschließend über bioinformatische

[1] Next Generation Sequencing (NGS) bezeichnet alle Verfahren, in denen parallel eine große Mengen von Molekülen bzw. Molekülfragmenten einer Probe sequenziert werden.

Schritte wieder zum originalen Transkript in voller Länge zusammengesetzt werden müssen („assembly“). Bei der Assemblierung nutzen bioinformatische Algorithmen Stranginformationen der Transkripte, um auch sich überlappende und antisense liegende Transkripte korrekt zu rekonstruieren (Martin & Wang, 2011). Als Sequenziertiefe („sequencing depth“) wird die Gesamtanzahl von generierten Reads in Relation zur Länge des betrachteten Genoms bezeichnet, wobei der Schwellwert von 40 Millionen Reads bei Säugetieren und einem durchschnittlich komplexen Transkriptom als Grenze für eine ausreichende Sequenziertiefe angenommen wird, um alle biologisch relevanten RNAs abzubilden, auch wenn sie in nur niedriger Menge in der Zelle vorliegen (Mortazavi et al., 2008).

Im Anschluss an die Assemblierung der Reads erfolgt die Kartierung der Transkripte, d.h. die Zuordnung der physischen Stelle (Position) des Transkriptes auf der DNA-Sequenz (de Sá et al., 2018). Sofern bereits ein Referenzgenom[2] und eine Kartierung von Genen verfügbar ist („annotation“), können Transkripte den jeweiligen Loci zugeschrieben werden („alignment“). RNA-Seq bietet alternativ die Möglichtkeit, sofern kein Referenzgenom mit Annotation[3] vorliegt, Transkripte *de-novo* zu assemblieren. Mit modernen Assemblierungsprogrammen wie StringTie (Pertea et al., 2015) lassen sich zudem aus sequenzierten Transkriptomen in Kombination mit einem Referenzgenom und gegebenenfalls auch unter Nutzung einer bestehenden Annotation neue Annotationen erstellen, die bisher unbekannte Loci integrieren (Pertea et al., 2016). Nachdem alle Transkripte kartiert wurden, lässt sich das Expressionsnivau eines Gens oder Transkriptes ermitteln.

Bei der RNA-Seq ist es üblich, bei der Aufbereitung der Gewebeprobe rRNA herauszufiltern. Ribosomale RNA macht mit 80 % bis 95 % einen Großteil des Transkriptoms einer Zelle aus (Darnell et al., 2000; Morlan et al., 2012) und das Interesse liegt i.d.R. auf anderen RNA-Molekülen. Diese würden von den Expressionswerten der rRNAs überschattet, wenn selbige in der Sequenzierung berücksichtigt werden würden.

Transkripte der Klasse der mRNA verfügen mehrheitlich über einen schwanzähnlichen Anhang von Adenin-Molekülen (Polyadenylierung), einen sogenannten poly-A-Anhang (Yosim & Fry, 2015). Im poly-A-basierten Sequenzierungsprotokoll wird für die RNA-Seq ein Filter auf polyadenylierte mRNAs ermöglicht (LaRossa, 2013). Auch lncRNAs weisen häufig mRNA-ähnliche Charakteristiken inklusive der Polyadenylierung auf (Sultan et al., 2014). Die Verwendung von poly-A-Protokollen führt jedoch dazu, dass eine substanzielle Anzahl nicht-polyadenylierter lncRNAs unentdeckt bleibt (Kashi et al., 2016). Um auch nicht-polyadenylierte Transkripte zu erfassen, empfiehlt es sich, ein alternatives Protokoll durchzuführen, das gezielt

[2] Als Referenzgenom wird die bekannte Abfolge aller Basenpaare der DNA einer Spezies bezeichnet.
[3] Es wird in örtliche und funktionale Annotation differenziert. Die örtliche Annotation gibt die Position eines Lokus im Genom wieder, während bei der funktionalen Annotation einem Lokus oder Gen eine konkrete biologische Funktion zugeordnet wird (de Sá et al., 2018).

rRNA eliminiert (ribosomale Depletion) (Zhao et al., 2014). Je nachdem, ob der Studienfokus mehr auf mRNA oder lncRNA liegt, ist bei der RNA-Seq folglich auf das passende Aufbereitungsprotokoll für die Bibliotheken zu achten. Speziell wenn lncRNAs im Fokus der Studie stehen und eine ribosomale Depletion angewendet werden soll, hat sich das Illumina Sequenzierungskit RiboZero als geeignet erwiesen (Schuierer et al., 2017).

4.2. Eigenschaften und Funktionen langer nichtkodierender RNA

Zum Arsenal der nichtkodierenden Elemente, die im menschlichen Genom über 97 % der DNA umfassen (Derrien et al., 2012), gehört auch die Klasse der lncRNA. Seit Entdeckung der ersten lncRNAs *H19* und *Xist* Ende der 1980er und Anfang der 1990er Jahre (Borsani et al., 1991; Brannan et al., 1990; Brockdorff et al., 1991) und nach anfänglicher Skepsis über die Bedeutung und biologische Relevanz dieser Genomelemente, hat die Anzahl der Studien zu lncRNAs in den 2000er Jahren ernorm zugenommen (Miao et al., 2017) (siehe Abbildung 1).

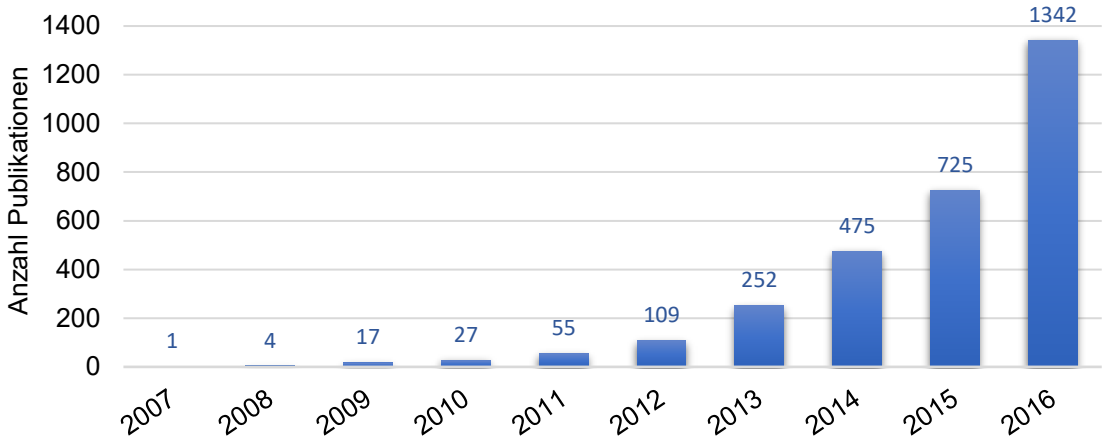

Abbildung 1 | Anzahl der von 2007 bis 2016 in der Web of Science Core Collection Datenbank gelisteten, englischsprachigen Publikationen mit Fokus auf langer nichtkodierender RNA (lncRNA). [Darstellung nach Miao et al., 2017]

LncRNAs sind allem voran durch die überwiegende Abwesenheit von proteinkodierendem Potenzial gekennzeichnet und haben konventionell eine definierte Länge von mindestens 200 Nukleoriden (nt) (Mercer et al., 2009). Diese willkürlich anmutende Grenze ist höchstwahrscheinlich auf den Umstand zurückzuführen, dass erste Bestimmungen des Kodierungspotenzials anhand von DNA-Sequenzen erst bei Längen über 200 nt eine Fehlerquote von unter 5 % aufwiesen (Fickett, 1982). Im Jahr 2001 postulierte John Mattick, dass intronische und andere nichtkodierende RNAs in Eukayroten grundlegend zur Genexpression und ihrer Regulation beitrügen und ein Verständnis höherer Organismen nur

erreicht werden könne, wenn neben dem Proteom auch das nichtkodierende Genom identifiziert und charakterisiert werden würde (Mattick, 2001).

Allgemein verfügen lncRNAs über ein geringeres Expressionsniveau als mRNAs (Derrien et al., 2012). Beim Hund wurden für lncRNAs in 26 Geweben Expressionsniveaus festgestellt, die im Durchschnitt ein Zwanzigstel der mRNA-Expression betrugen (mit Ausnahme der Expressionsmuster im Hoden) (Le Beguec et al., 2018). Untersuchungen an menschlichen Zelllinien haben gezeigt, dass die für mRNAs charakteristische Polyadenylierung nur bei einem Teil der bisher entdeckten lncRNAs vorzukommen scheint (Sultan et al., 2014). Weiteres zum Merkmal der Polyadenylierung und sich daraus ergebende Implikationen für das zu wählende Aufbereitungsprotokoll der Proben für die RNA-Seq sind Kapitel 4.1 zu entnehmen.

In einem Übersichtsartikel wiesen Pennachio & Rubin (2001) darauf hin, dass ein hoher Sequenzkonservierungsgrad zwischen Spezies ein Indiz für funktionale Wichtigkeit der jeweiligen Sequenz ist, wobei mittels derartiger komparativer Ansätze trotzdem nicht alle genregulatorischen Elemente aufgedeckt werden können (Pennacchio & Rubin, 2001). LncRNA-Konservierung kann laut Diederichs in vier Kategorien aufgeteilt werden (Diederichs, 2014):

1. Sequenzkonservierung
2. Strukturkonservierung
3. Funktionskonservierung
4. Expressionskonservierung syntenischer Loci

Hinsichtlich der ersten Kategorie (Sequenz) konnten Derrien et al. (2012) beobachten, dass lncRNAs insgesamt über ein niedrigeres Konservierungsniveau verfügen als mRNAs (Derrien et al., 2012). Die seltener untersuchte Sekundärstruktur (2. Kategorie) von lncRNAs lässt sich durch eine Reihe bioinformatischer Methoden vorhersagen (zusammengefasst durch (Yan et al., 2016)). Beispielhaft ausgewählt zeigte der Vergleich der Sekundärstruktur der lncRNA *lncTCF7* zwischen Mensch und anderen Säugern, neben einer Sequenzkonservierung, die strukturelle Konservierung einzelner Regionen auf (Owens et al., 2019). Ein Beispiel für die positionelle Konservierung wurde von Ulitsky (2011) präsentiert: Während die DNA-Sequenz von lncRNAs in Zebrafischen mehrheitlich über ein geringes Konservierungsniveau in anderen Spezies verfügte, konnten in anderen Organismen trotzdem positionell konservierte genomische Loci gefunden werden. Funktionelle Analysen dieser konservierten Regionen im Zebrafisch ergaben, dass eine Blockade der Loci über Antisense-Oligos zu embryonalen Entwicklungsstörungen führt (Ulitsky et al., 2011), was auf eine funktionale Konservierung der lncRNAs (3. Kategorie) schließen ließe. Bezüglich der vierten Kategorie (Expression) deuten Ergebnisse erster komparativer Studien auf eine vermehrt spezies-spezifische Expression

(Ulitsky et al., 2011; Washietl et al., 2014) und insbesondere eine gewebsspezifische Expression von lncRNAs hin (Cabili et al., 2011; Ulitsky et al., 2011). Jedoch fanden Washietl et al. (2014), dass für etwa 80 % der beobachteten menschlichen intergenischen lncRNAs Orthologe in Schimpansen zu finden waren und für 35 % bis 39 % der lncRNAs Orthologe bei Ratte, Maus und Rind vorlagen. Die Autoren fanden weiterhin, dass Gewebsspezifität stärker ausgeprägt war, wenn die betreffende lncRNA weit (über 10 Kilobasen (kb)) von einem proteinkodierenden Gen entfernt war. Derrien und Kollegen (2012) stellten nach Sequenzvergleichen von humanen lncRNAs mit allen anderen verfügbaren Säugetiergenomen (N=18) fest, dass 30 % der lncRNAs (sequenz-)spezifisch für Primaten waren. In einer komparativen Studie konnten Hezroni und Kollegen zeigen, dass Hunderte von lncRNAs aus dem menschlichen Genom über homologe lncRNAs mit vergleichbarem Expressionsniveau in anderen Säugetieren verfügen (Hezroni et al., 2015). In Folgeuntersuchungen wurde gezeigt, dass etwa 5 % der konservierten lncRNAs in Säugetieren ursprünglich aus proteinkodierenden Genen hervorgegangen sind, die ihre ursprüngliche Funktion verloren haben (Hezroni et al., 2017).

Neben der gewebsspezifischen Expression zeigen lncRNAs in verschiedenen ontogenetischen Phasen differenzielle Expression. Insbesondere in sich entwickelnden Organen ist ihre Expression angereichert, und die betreffenden lncRNAs demonstrieren eine vergleichsweise hohe evolutionäre Konservierung, vornehmlich in Promotorregionen (Darbellay & Necsulea, 2020). Studien deuten darauf hin, dass lncRNAs neben der Involvierung in die Ontogenese, auch in die Pathophysiologie diverser Krankheiten wie humane Autoimmunerkrankungen, Krebs und genetisch bedingte Erbkrankheiten involviert sind (Übersichtsartikel z.B.: Bhan et al., 2017; Hu et al., 2018; Sparber et al., 2019; Zou & Xu, 2020).

Der Umstand, dass lncRNAs über eine räumlich-zeitliche Spezifität zu verfügen scheinen, d.h. spezifische Expressionsmuster in diversen Geweben und während der Ontogenese und Pathogenese zeigen, prädestiniert sie als indikatorische Biomarker. Die Eignung von lncRNAs als Biomarker, speziell in der Krebsdiagnostik im Humanbereich, wurde von Bolha et al. (2017) zusammengefasst. Aufgrund ihrer differenziellen und relativ stabilen Expression in Körperflüssigkeiten wie Blut, die nichtinvasiv entnommen werden können, bieten sich lncRNAs besonders für die Diagnostik an (Bolha et al., 2017). Im Umkehrschluss kann die Inhibition onkogener lncRNAs einen neuen Therapieansatz für Krebspatienten darstellen und erste Studien, in denen gezielt eine onkogene lncRNA duch Antisenseoligonukleotide in ihrer Expression reduziert wurde, liefern vielversprechende Ergebnisse bezüglich eines reduzierten Tumorwachstums (Xia et al., 2017).

LncRNAs entfalten ihre genregulatorische Wirkung durch eine Vielfalt an Mechanismen (siehe Abbildung 2). Zu den bisher bekannten Wirkungsweisen gehören die Chromatin-

Remodellierung und die Aufrechterhaltung des Chromatinzustands. Weiterhin ermöglichen sie eine transkriptionelle Verstärkung („enhancer“) oder Unterdrückung („repressor“), z.B. durch die Bindung an Transkriptionsfaktoren, was sowohl *in cis*[4] als auch *in trans* geschehen kann (Long et al., 2017; Marchese et al., 2017; Rinn & Chang, 2012). Außerdem können lncRNAs eine sogenannte Köderfunktion („decoy“) erfüllen, bei der sie an DNA-regulierende Proteine wie Transkriptionsfaktoren binden und dadurch die Transkription beeinflussen. Weiterhin erfüllen sie strukturelle Aufgaben und dienen als Adapter für andere Proteine („scaffolding“); sie binden miRNAs und unterbinden damit deren Wirkungsentfaltung in der Zelle („sponging“), sie lenken und führen Proteinkomplexe zeitlich und räumlich differenziert an Zielstellen auf der DNA („guide“) (Marchese et al., 2017). Darüber hinaus können lncRNAs die Chromatinstruktur modifizieren und so die Transkription sowohl ermöglichen als auch hemmen, z.B. durch Interaktion mit chromatinmodifizierenden Komplexen (Marchese et al., 2017; Rinn & Chang, 2012). Insbesondere wird angenommen, dass natürliche Antisense-Transkripte (NATs), die sich auf dem gegenüberliegenden Strang eines anderen, in der Regel proteinkodierenden Gens befinden, den Chromatinzustand beeinflussen und somit ein epigenetisches regulatorisches Potenzial besitzen (Magistri et al., 2012). Eine große Anzahl der bisher entdeckten lncRNAs sind NATs zu anderen Genorten (Katayama et al., 2005; Pelechano & Steinmetz, 2013; Zhang et al., 2006). Im Weiteren wurde festgestellt, dass lncRNAs häufiger starke positive Korrelationen ($r > 0,9$) mit proteinkodierenden Genen aufweisen als Korrelationen von mRNAs mit anderen mRNAs (Derrien et al., 2012).

[4] *In cis* beschreibt die Wirkung auf nahe gelegende Loci, die sich beispielsweise auf demselben Chromosom oder in direkter physischer Nachbarschaft befinden. *In trans* beschreibt Interaktionen mit weit entfernt gelegenen Loci.

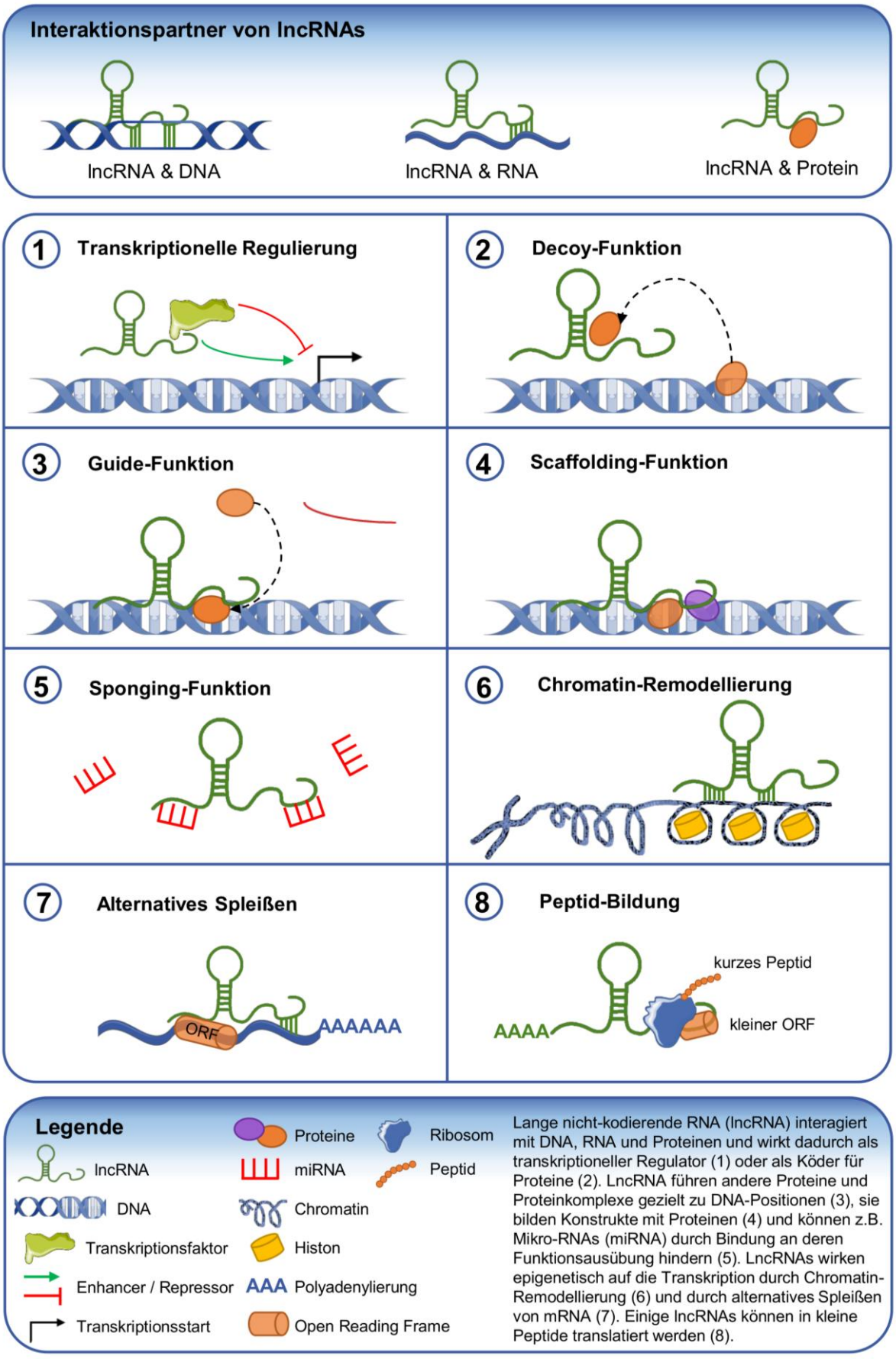

Abbildung 2 | Interaktionspartner und Funktionsweisen von langer nichtkodierender RNA (lncRNA) [Darstellung nach Cheng et al., 2019; Morlando et al., 2015; Salehi et al., 2017]

4.2.1. Identifikation neuer langer nichtkodierender RNA

Zur Identifikation neuer, unbekannter lncRNAs wurden in den vergangenen Jahren zahlreiche Algorithmen entwickelt, um sequenzbasiert das Kodierungspotenzial von Transkripten zu ermitteln und diese so in proteinkodierende mRNAs und (lange) nichtkodierende RNAs zu differenzieren. Zu den populären bioinformatischen Anwendungen gehören CPC (Kong et al., 2007), CPAT (Wang et al., 2013) oder FEELnc (Wucher et al., 2017). Antonov stellte in einem Methodenvergleich fest, dass FEELnc und CPAT aktuell die genauesten lncRNA-Vorhersagen bei Säugetieren liefern (Antonov et al., 2019). In der vorliegenden Dissertation wurde das Programm FEELnc zur Identifikation und positionellen Charakterisierung von lncRNAs genutzt (siehe Kapitel 5).

Nach dem Ausschluss von Loci, die als proteinkodierend annotiert sind, sind die drei prinizipiellen Kernelemente und RNA-intrinsischen Eigenschaften zur Vorhersage einer lncRNA im Programm FEELnc: (1) die Länge von mindestens 200 nt sowie die Exonanzahl, (2) das Vorhandensein und die (relative) Länge eines Open Reading Frames (ORF), sowie (3) multiple *k-mer*-Frequenzen (Wucher et al., 2017). Der Parameter der Länge ist selbsterklärend und in der Anwendung ein einfacher Filterschritt, der bei Bedarf angepasst werden kann. Hinsichtlich der ORFs ist die Länge des längsten ORF einer Sequenz von Bedeutung (Wucher et al., 2017), da lange ORFs in der Regel eher in ein Protein übersetzt werden (Brown, 2002). Ein ORF beinhaltet alle Codons (Basen-Triplets) vom Start- bis zum Stop-Codon (Lewin, 2000). Neben der Länge und der Existenz von ORFs unterscheiden sich lncRNAs von mRNAs hinsichtlich der relativen Häufigkeit von Oligonukleotiden (*k-mers*) (Li et al., 2014), wobei *k* die Länge des Oligonukleotids bezeichnet. Anhand der ORFs und der *k-mer*-Frequenz-Information wird in FEELnc mittels der Maschine-Learning-Methode *Random Forest* (Breiman, 2001) eine Kodierungspotenzialpunktzahl („Coding Potential Score“) für jedes Transkript kalkuliert. Mittels Schwellwerten, die durch den Vergleich mit einem Trainingsset ermittelt werden, erfolgt die Auftrennung der RNAs in lncRNAs und mRNAs sowie RNAs mit unsicherem Kodierungspotenzial (Wucher et al., 2017).

Nach der initialen Identifikation von lncRNAs können diese in Kategorien eingeteilt werden, die sich an deren Position im Verhältnis zum nächst gelegenen (proteinkodierenden) Gen orientieren (Wucher et al., 2017). LncRNAs können demnach intergenisch, d.h. zwischen anderen Genen liegen, oder sich auf demselben („sense“) oder gegenüberliegenden („antisense“) DNA-Strang teils oder vollständig mit einem Gen überlappen („genic“). Wenn eine Überlappung vorliegt („overlapping“) oder die lncRNA sogar vollständig innerhalb eines Gens liegt („nested“, „containing“), kann die Positionierung in bzw. Überlappung mit einem Exon („exonic“) oder Intron („intronic“) vorliegen (Wucher et al., 2017).

4.2.2. Funktionale Charakterisierung neuer langer nichtkodierender RNA

Während zur Identifikation von lncRNAs schon eine Reihe zuverlässiger Programme zur Verfügung steht, ist der Aspekt der funktionalen Annotation noch nicht ausreichend beachtet worden und sollte zukünftig mehr Aufmerksamkeit erfahren (Ramakrishnaiah et al., 2020). Da lncRNAs kein Protein als Produkt haben, definiert sich ihre Funktion nicht eigenständig, sondern über ihr komplexes Zusammenspiel mit und ihren Einfluss auf Gene und Proteine (Ramakrishnaiah et al., 2020). Wegen der geringen Sequenzkonservierung ist eine Vorhersage der Funktion basierend auf der Sequenz allein schwierig (Kashi et al., 2016).

Gudenas und Kollegen (2019) haben in einem Übersichtsartikel folgende Eigenschaften zur funktionalen Charakterisierung von humanen lncRNAs benannt und näher beleuchtet: die Assoziation von lncRNAs mit Krankheiten, die Analyse der Expressionsmuster, die zelluläre Lokalisation (Zellkern, Zytoplasma), RNA-Protein-Interaktionen, RNA-Modifikationen sowie die funktionale Motivbindung (Gudenas et al., 2019). Weil lncRNAs spezifisch in Geweben und Entwicklungsstadien exprimiert werden, ist das Expressionsmuster ein Indiz für die biologische Funktion, wobei Gruppenvergleiche (differenzielle Expression) besonders hilfreich sind. In der vorliegenden Dissertation wurden vor allem die positionelle Assoziation mit merkmalsbeeinflussenden Genorten (z.B. für Milchleistung[5] anstatt Krankheiten) und die Expressionsmusteranalyse genutzt, inklusive Koexpressionnetzwerkanalysen, die für die funktionale Annotation von unbekannten Loci ein hilfreiches Werkzeug sind (van Dam et al., 2017). Bei der sogenannten *guilt-by-association*-Heuristik folgt man bei der funktionalen Annotation der Prämisse, dass unbekannte Loci dieselben oder ähnliche biologische Funktionen haben bzw. im selben Schaltkreis agieren wie die Gene, mit denen sie in ihrer Expression korreliert sind (Gudenas et al., 2019). Dafür werden entweder Module von Genen inklusive lncRNAs erstellt, die in ihrer Expression miteinander korrelieren oder es werden hochvernetzte (Hub-) lncRNAs aus Netzwerken isoliert, gemeinsam mit durch vorliegende Korrelationen verbundene Netzwerkknoten (Gene). Diese Genmodule oder Gengruppen werden anschließend einer Gene Set Enrichment Analyse (GSEA) (Gudenas et al., 2019) bzw. einer Pathway Enrichment Analyse zugeführt. Die Funktionsvorhersage mittels GSEA kann eine hohe Spezifität und Präzision erreichen, wobei beides stark vom festgesetzten Signifikanzniveau und der Anzahl der eingespeisten Genpartner abhängt (Liao et al., 2011). Neben der Korrelation einer lncRNA mit vielen anderen Genen, kann auch, sofern vorliegend, eine besonders starke Korrelation zu einem benachbarten oder überlappend liegenden Gen Aufschluss über eine mögliche Funktion geben, sofern dieses Gen bereits funktional annotiert ist. Auch hier wird der *guilt-by-association*-Ansatz verfolgt, wonach die lncRNA zum selben Funktionskreis gehört, wie das mit ihr korrelierte Gen.

[5] Assoziation in Form von bekannten QTL aus der Animal QTL Database (https://www.animalgenome.org/)

Die Integration von *omics*-Daten (wie Expressiondaten (Transkriptom), DNA-Methylierung (Methylom), Genomvariation (Genom) und Interaktion mit miRNAs (Interaktom)) stellt eine wirksame Strategie zur effizienten Charakterisierung von lncRNAs dar (Li et al., 2020). Auch die Integration von Transkriptom und Metabolom über Pathway-Analysen schafft durch die Lieferung zuverlässigerer Ergebnisse im Vergleich zu einer einseitigen Auswertung von Genexpressions- oder Metabolitendaten einen Mehrwert bei der Untersuchung biologischer Fragestellungen (Murakami et al., 2015). Sowohl von Netzwerkanalysen (Hub-lncRNAs mit korrelierten Genen) als auch Pathway Enrichment Analysen wurde unter Berücksichtigung mehrerer *omics*-Datenebenen (Transkriptom, Phänom, Metabolom) im Rahmen dieser Arbeit Gebrauch gemacht (siehe Kapitel 5).

Es empfiehlt sich, mehrere Methoden zu kombinieren, um die Involvierung von lncRNAs in biologische Schaltkreise zu prognostizieren. Die Vorhersagen dienen unter anderem zur Priorisierung von funktional und regulatorisch vielversprechenden lncRNAs bevor eine eingehende experimentelle Untersuchung und Validierung der Funktion erfolgt.

Zur experimentellen Funktionsvalidierung, die deutlich aufwändiger und zeitintensiver ist als die *in-silico* Prognose von Funktionen, gehören beispielsweise strukturierte Knock-Down-Studien mit ausgewählten lncRNAs. Hier wird untersucht, wie sich die induzierte Herunterregulierung einer einzelnen lncRNA auf die Expression übriger Gene, d.h. auf den molekularen Phänotypen auswirkt (Ramilowski et al., 2020). Zusätzlich lassen sich durch Methoden wie Chromatin Isolation by RNA Purification (ChIRP) Bindungsstellen der lncRNAs im Genom ausmachen (Chu et al., 2012), über die Funktionen und Interaktionen abgeleitet werden können. Eine Methodenübersicht sowie Empfehlungen zu sukzessiven Arbeitsschritten zur Identifikation und funktionalen Charakterisierung von lncRNAs sind in der Literatur gegeben (Kashi et al., 2016; Kunz et al., 2020; Ramakrishnaiah et al., 2020).

4.3. Nutzung von experimentellen Tierpopulationen

Die Nutzung von experimentellen Tierpopulationen bietet Vorteile bei der Untersuchung genetischer Kausalitäten für die Merkmalsausprägung. So lässt die dauerhafte Haltung und Aufzucht unter standardisierten Bedingungen eine Reduzierung der Umwelteffekte zu, wodurch beobachtete Gruppeneffekte und unterschiedliche Merkmalsausprägungen mit größerer Sicherheit auf genotypische Unterschiede zurückzuführen sind. Dasselbe Prinzip wird in der Tierzucht bei der Leistungsprüfung auf Station angewendet, bei der systematische Umwelteinflüsse auf die Tierleistung soweit wie möglich reduziert werden sollen (William & Simianer, 2011). Wichtig bei den standardisierten Haltungsbedingungen ist, dass diese mit den Bedingungen in der Praxis vergleichbar sind (William & Simianer, 2011). Weiterhin ist in

experimentellen Tierpopulationen die ausführliche Abstammungsdokumentation und eine standardisierte, strukturierte Merkmalserfassung leichter umsetzbar.

Heutige Nutztierarten, z.B. Rinder, werden in Rassen auf eine spezielle, in der Regel uniforme Nutzungsrichtung gezüchtet und weisen rassetypische Merkmalsausprägungen auf (Encyclopædia Britannica, 2018). Zur Analyse der grundlegenden genetischen Hintergründe komplexer Merkmale ist jedoch eine hohe phänotypische und genotypische Varianz vorteilhaft, wie sie durch die Kreuzung spezialisierter Rassen oder Linien erreicht werden kann. So ermöglicht die Kreuzungszucht die Ausnutzung einer größeren genetischen Bandbreite als die Nutzung einer einzelnen Rasse (Swan & Kinghorn, 1992). Die Kreuzung von Rassen, die auf extreme Merkmalsausprägung gezüchtet wurden, ermöglicht die Einführung einer maximalen Merkmalsvarianz in der zweiten Kreuzungsstufe (F_2) (Kühn et al., 2002), die somit eine Testpopulation von hohem Informationsgehalt darstellt (Kogelman et al., 2013). Zudem bietet ein Kreuzungsdesign die Möglichkeit, die Korrealtionen von Merkmalen innerhalb einer Rasse aufzubrechen, zwischen denen kein ursächlich genetischer Zusammenhang besteht (Kühn et al., 2002).

Im Nutztierbereich finden F_2-Designs wiederholt Anwendung bei der Analyse komplexer Merkmale mittels genomweiter Assoziationsstudien (GWAS) (siehe z.B. (Lutz et al., 2017) beim Huhn, (Falker-Gieske et al., 2019) beim Schwein und (Alexander et al., 2007) beim Rind). F_2-Populationen sind für die Untersuchung quantitativer Merkmale besonders wertvoll, da alle genetischen Effekte (additiv, dominant, epistatisch) geschätzt werden können (Zhang, 2012).

4.3.1. Kreuzungspopulation SEGFAM

Die vorliegende Arbeit basiert auf Tiermaterial aus einer genotypisch und phänotypisch tiefgehend charakterisierten Kreuzungspopulation von Rindern mit segregierenden Familien (SEGFAM) [6], die in einem deutschlandweit einzigartigen Langzeitversuch (Leibniz-Gemeinschaft, 2008) dauerhaft am Leibniz-Institut für Nutztierbiologie (FBN) in Dummerstorf unter standardisierten Bedingungen gehalten und aufgezogen wurden. Speziell wurden Tiere der zweiten Kreuzungsstufe (F_2) genutzt, die in der Parentalgeneration (P0) auf fünf Charolais-Bullen (Fleischtyp) und 40 Holstein-Kühe (Milchtyp) zurückgehen (siehe Abbildung 3). Diese hochinformative Test- und Ressourcenpopulation wurde eingehend in Kühn et al. (2002) beschrieben und dient fortwährend als Tiermodell zur Untersuchung der physiologischen und genetischen Grundlagen der phänotypischen Variation der Nährstoffverteilung bzw. -umsetzung.

[6] Die Ressourcenpopulation SEGFAM wurde von 1997 bis 2013 am FBN in Dummerstorf gehalten und die letzten Einzeltiere wurden 2018 geschlachtet.

Als Gründerrassen wurden Charolais und Deutsche Holstein-Friesian ausgewählt, da beide Rassen für komplexe Merkmale hoher ökonomischer Relevanz (Futtereffizienz und Fleischansatz, Milchleistung) extreme Ausprägungen aufweisen, über eine vergleichbare Ontogenese und Körpergröße im adulten Stadium verfügen (Kühn et al., 2002). Trotz relativ enger verwandtschaftlicher Beziehungen wiesen Tiere der F_2-Population eine hohe phänotypische Diversität auf. Die für statistische Auswertungen notwendige Populations- und Famliengröße wurde durch induzierte Superovulation und anschließenden Embryotransfer in der P0- und F_1-Generation realisiert (Kühn et al., 2002).

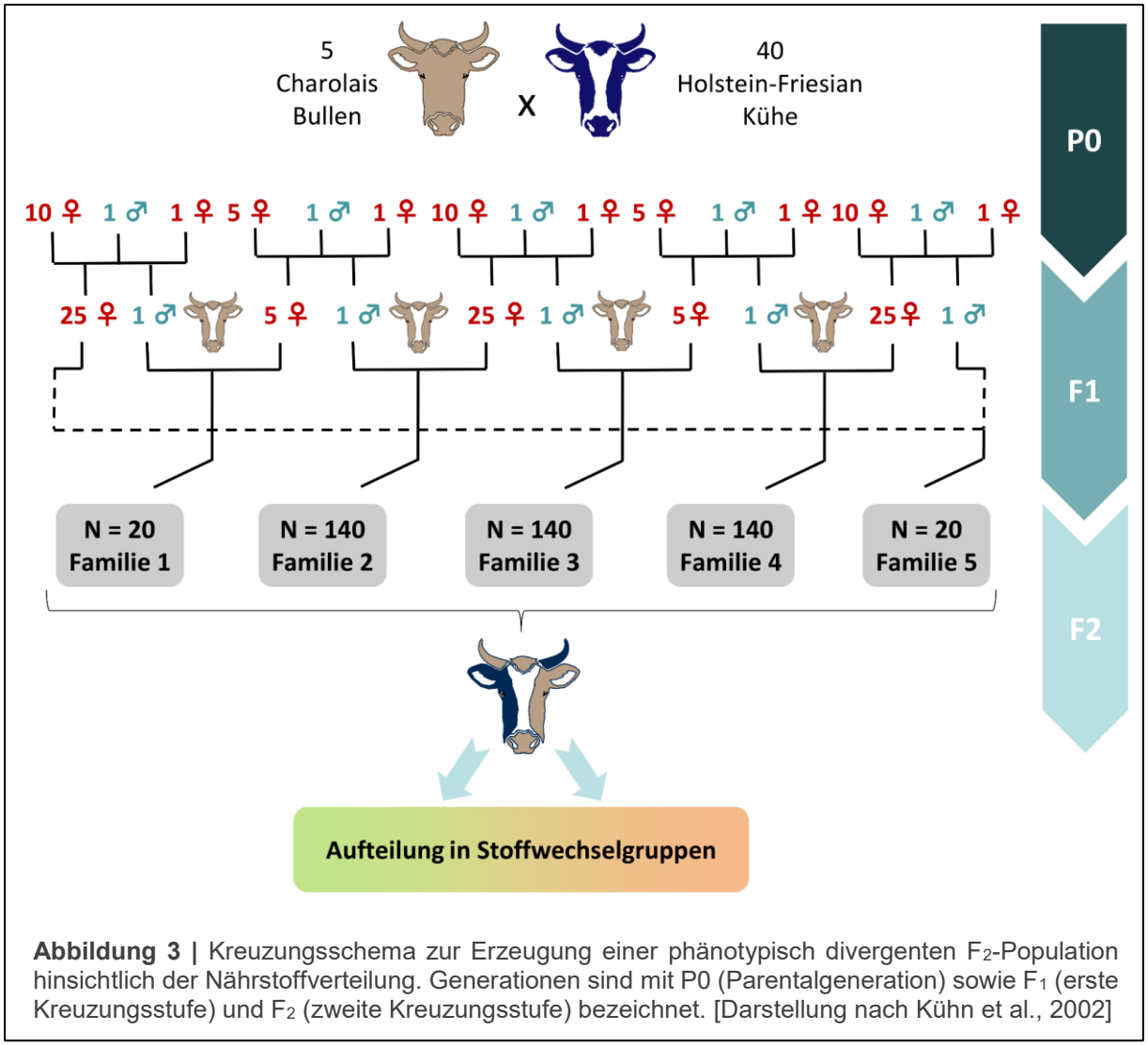

Abbildung 3 | Kreuzungsschema zur Erzeugung einer phänotypisch divergenten F_2-Population hinsichtlich der Nährstoffverteilung. Generationen sind mit P0 (Parentalgeneration) sowie F_1 (erste Kreuzungsstufe) und F_2 (zweite Kreuzungsstufe) bezeichnet. [Darstellung nach Kühn et al., 2002]

Schwarzbunte Holstein-Kühe sind die in Deutschland am häufigsten vertretene Milchviehrasse und haben einen Anteil von 77 % am Gesamtbestand des deutschen Milchviehs. Charolais gehören mit einem Anteil von 8 % am Gesamtbestand der Fleischrinder neben Limousin, Fleischfleckvieh und Deutsch Angus zu den bedeutensten reinrassig gezogenen Fleischrindern in Deutschland (BMEL (Bundesministerium für Ernährung und Landwirtschaft), 2020). Bisher konnte die F_2-Population zur Untersuchung der molekulargenetischen Hintergründe verschiedener komplexer und monogener Merkmale dienen (z.B. Milchleistung,

Futtereffizienz, Fettsäureprofil im Muskel, Lipidstoffwechsel und -ansatz, Wachstum, Verhalten, Bovine neonatale Panzytopenie, Rat-Tail-Syndrom) und stellte die Grundlage für eine Vielzahl von publizierten Studien dar (unter anderem: (Demasius et al., 2013; Eberlein et al., 2010; Friedrich et al., 2016; Hammon et al., 2007; Knaust et al., 2016; Krappmann et al., 2011, 2012; Liu et al., 2018; Melzer et al., 2017; Widmann et al., 2013, 2015, 2011)). Die Aufdeckung der genetischen Ursachen für eine differente Umsetzung von Nährstoffen kann letztlich in der gezielten Zucht auf ressourceneffizientere Tiere münden.

In die Untersuchungen dieser Arbeit wurden F_2-Tiere beider Geschlechter einbezogen, die sich phänotypisch hinsichtlich ihres Stoffwechseltyps unterschieden. Die weiblichen Tiere wurden am 30. Tag der zweiten Laktation und die männlichen Tiere am 180. Lebenstag im institutseigenen Schlachthaus getötet. Gewebeproben wurden unmittelbar nach Schlachtung entnommen und in flüssigem Stickstoff schockgefroren und bei -80°C eingelagert. Weiterhin standen Plasmaproben zur Verfügung, die vor der Schlachtung genommen wurden und in bei -80 °C gelagert wurden. Das experimentelle Design umfasste je zwei Gruppen je Geschlecht, die in phänotypisch divergierende Stoffwechseltypen unterteilt waren. Die Auswahl der Tiere erfolgte in zwei Schritten. Im ersten wurde eine Vorauswahl von 30 Tieren je Geschlecht anhand von vorliegenden phänotypischen Daten getroffen. Die männlichen Tiere wurden nach der Futtereffizienz (bestimmt als Residual feed intake (RFI) im letzten Lebensmonat) als Schlüsselmerkmal und dem intramuskulären Fettgehalt im Skelettmuskel (IMF im *M. longissimus dorsi*) als akzessorisches Merkmal ausgewählt. Die Vorauswahl der weiblichen Tiere erfolgte nach der durchschnittlichen energiekorrigierten Milchleistung (ECM) in der Woche vor dem Schlachtzeitpunkt als Schlüsselmerkmal (30. Tag der Laktation) und dem Gehalt des IMF als akzessorisches Merkmal. In einem zweiten Schritt wurden die zur Verfügung stehenden Plasmaproben der vorausgewählten Tiere einer globalen Metabolomanalyse unterzogen (ausgeführt durch Metabolon Inc., Morrisville, USA). Die erhaltenen metabolomischen Parameter ermöglichten eine tiefere phänotypische Charakterisierung der Tiere und dienten somit für die finale Gruppeneinteilung nach divergierenden Phänotypen. Die Gewebe (Leber, Skelettmuskel, Pansen und Jejunum) der final ausgewählten Tiere (je Geschlecht zwei Gruppen mit N = 12 Tieren) wurden einer holistischen RNA-Seq mit rRNA-Depletion unterzogen.

Tabelle 2: Phänotypische Charakterisierung des Tiermaterials der drei Studien (Mittelwert in der ersten und Standardabweichung in der zweiten Zeile)

	Studie 1		Studie 2		Studie 3	
Merkmals-bezeichnung	Metabolische Effizienz metabolic efficiency		Futtereffizienz feed efficiency		Nährstoffverteilung nutrient partitioning	
IMF[1]	3,46 ±1,30	5,51 ± 2,34	2,77 ± 0,95	4,59 ± 1,17	4,16 ± 1,12	6,18 ± 2,30
CFC[2]	15,93 ± 3,16	22,93 ± 4,88	14,39 ± 2,86	20,28 ± 3,27	17,09 ± 2,59	25,32 ± 3,27
RFI[3]	-21,30 ± 4,44	20,83 ± 4,41	-20.91 ± 4.47	20.48 ± 4.40	*NA*	*NA*
ECM[4]	190,87 ± 22,02	30,97 ± 9,18	*NA[5]*	*NA*	190,87 ± 22,02	29,28 ± 7,92
Geschlecht	männlich, weiblich		männlich		weiblich	
Tieranzahl	48		26		25	
Gewebe	Leber, Skelettmuskel, Jejunum, Pansen		Leber		Leber	
Gruppen-bezeichnung in der Publikation	high metabolic efficiency	low metabolic efficiency	high feed efficiency	low feed efficiency	secretion type (SEC)	accretion type (ACC)

[1]IMF = Intramuskulärer Fettgehalt im *M. Longissimus dorsi* in Prozent, [2]CFC = Content Fat in Carcass (Fettgehalt im Schlachtkörper) in Prozent, [3]RFI = Residual Feed Intake (residuale Futteraufnahme) im letzten Lebensmonat vor der Schlachtung, [4]ECM = Energiekorrigierte Milch in kg in den letzten 7 Tagen vor der Schlachtung, [5]NA = nicht analysiert

In den drei vorliegenden Studien (siehe Kapitel 6) wurden verschiedene Bezeichnungen für den jeweiligen Phänotypen bzw. die Gruppeneinteilung nach divergenter phänotypischer Merkmalsausprägung verwendet. Die zugrunde liegende Einteilung und die Auswahlkriterien waren jedoch identisch. Die verwendeten Bezeichnungen und Merkmalskategorien sind in Tabelle 2 zusammengefasst. In den folgenden Kapiteln wird die einheitliche Bezeichnung der „Nährstoffverteilung“ („nutrient partitioning“) zur Bezeichnung des Phänotyps verwendet werden.

5. Identifikation und funktionale Annotation von lncRNAs beim Rind

Während lncRNAs beim Menschen und der Maus bereits umfassend untersucht werden, ist die Anzahl von Studien am Nutztier in diesem Feld vergleichsweise klein (Kosinska-Selbi et al., 2020). Im November 2020 waren in der NONCODE-Datenbank (Version 6) für das Rind 23.515 lncRNA-Transkripte bzw. 22.227 lncRNA-Loci hinterlegt, während etwa jeweils das Vierfache für Maus und Mensch abgelegt war (NONCODE, 2020). Durch die relativ geringe Sequenzkonservierung (siehe Kapitel 4.2) zwischen Spezies, ist es notwendig, die Identifikation, Katalogisierung und Charakterisierung von lncRNAs für einzelne Arten separat durchzuführen.

Detaillierte Beschreibungen zur Probenentnahme, Aufbereitung, Metabolomuntersuchung, Tranksriptomsequenzierung und Einstellungen in den sich anschließenden Analysen können den Studien 1 und 2 entnommen werden (Nolte et al., 2019, 2020). In dieser Arbeit wurde grundlegend die Prämisse verfolgt, dass lncRNAs zu demselben biologischen Regelkreis gehören wie die Elemente, mit denen sie signifikant und stark korreliert sind (*guilt-by-associatio*n-Heuristik) bzw. bedingt durch Koregulation mit benachbarten und hochkorrelierten Genen Einfluss auf deren Funktionalität nehmen. Im Folgenden werden wichtige Elemente bei der Aufbereitung und Auswertung der Daten benannt und eine schematische Übersicht der Arbeitsschritte und verwendeten Methoden und Programme ist in Abbildung 4 gegeben.

Auswahl eines lncRNA-orientierten Protokolls für die Transkriptomsequenzierung

Die Gewebeproben, die für diese Arbeit zur Verfügung standen, wurden nach der Schlachtung entnommen, schockgefroren und bei -80°C bis zur Aufbereitung gelagert. Für die Erstellung der Bibliotheken für die Transkriptomsequenzierung wurde das Aufbereitungskit TruSeq Stranded RNA-Ribo-Zero H/M/R Gold Kit (Illumina, San Diego, CA, USA) mit dem zugehörigen Protokoll angewendet, das eine rRNA-Depletion einschließt, wodurch auch nicht polyadenylierte Transkripte erfasst werden. Die Auswahl des Protokolls beeinflusst maßgeblich, welche Art von Transkripten anschließend sequenziert werden kann (siehe Kapitel 4.1).

Erstellung einer projektspezifischen Annotation

Aufgrund der Tatsache, dass die Mehrheit der lncRNAs bei vielen Spezies nach wie vor unbekannt und nicht annotiert ist, steht die ausschließliche Nutzung von Referenzgenomen und -annotationen der Entdeckung neuer Loci entgegen. Die holistische RNA-Seq ermöglicht die Aufdeckung und Assemblierung neuer, unbekannter Transkripte, die unter Zuhilfenahme des Referenzgenoms und gegebenenfalls einer verfügbaren Annotation positionell annotiert

werden können. Die Funktion *merge* im Programm StingTie (Pertea et al., 2015) implementiert dieses Prinzip und ermöglicht eine unkomplizierte Integration multipler Datensätze zur Erstellung einer projektspezifischen Annotation. In dieser Arbeit wurden insgesamt 204 Proben genutzt, die sich aus vier Geweben und insgesamt 52 Tieren beider Geschlechter rekrutieren. In Studie 1 wurde das zum damaligen Zeitpunkt verfügbare Referenzgenom UMD.3.1 genutzt und nach Veröffentlichung des neuen bovinen Referenzgenoms ARS-UCD.1.2 im Dezember 2018 wurde diese Version für die Studien 2 und 3 genutzt. Mit der Novellierung des Referenzgenoms wurden dementsprechend sowohl eine neue Kartierung der Transkripte als auch eine neue Erstellung der projektspezifischen Annotation notwendig.

In-silico Identifikation von lncRNAs

Im Anschluss an die Transkriptomsequenzierung, Qualitätsfilter, das Assemblieren, Kartieren und Quantifizieren der Expression der Transkripte, folgte die Identifikation möglicher lncRNAs im Datensatz. In der vorliegenden Arbeit wurde die Software FEELnc (Wucher et al., 2017) zur Identifikation und positionellen Charakterisierung sowie zur Aufdeckung möglicher benachbarter Partnergene verwendet.

Identifikation von lncRNAs mit genregulatorischem Potenzial

Der RIF ist ein etablierter an, Transkriptionsfaktoren erprobter Algorithmus zur Prognostizierung von genregulatorischem Potenzial einzelner genomischer Elemente (Reverter et al., 2010), der auch in den ersten beiden Studien dieser Dissertation angewendet wurde. RIF nutzt zwei komplementäre Metriken (RIF1 und RIF2), die Loci mit hohem genregulatorischen Potenzial eine Punktzahl zuweisen. Die Punktzahl hängt von der Veränderung der Korrelation zwischen dem regulierenden Locus (z.B. eine lncRNA) und seinem Zielgen in zwei verschiedenen Gruppen ab sowie davon, ob das Zielgen differenziell exprimiert (DE) ist und wie hoch sein allgemeines Expressionsniveau ist. Eine hohe RIF1-Punktzahl wird vergeben, wenn das regulatorische Element durchgängig in beiden Gruppen mit einem Zielgen koexprimiert ist. Ein Regulator erhält dagegen eine hohe RIF2-Punktzahl, wenn sich die Korrelationsrichtung (positiv, negativ) zwischen Regulator und Zielgen in den beiden Merkmalsgruppen unterscheidet oder eine Korrelation nur in einer von beiden Gruppen besteht. In dieser Arbeit erfolgte die Berechnung der RIF-Metriken für ein priorisiertes Set, das Loci von besonderem Interesse für den jeweiligen Phänotypen beherbergte, d.h. DE Loci, lncRNAs, gewebsspezifsch exprimierte Loci, sowie Loci, die mit einem QTL für Milchleistung oder residuale Futteraufnahme kolokalisiert waren.

Erstellung von Korrelationsnetzwerken

Zur Erstellung von Korrelationsnetzwerken wurde in Studie 1 und 2 die PCIT Analyse verwendet (Reverter & Chan, 2008). Die PCIT berechnet in einem Locus-Set alle paarweisen Korrelationen und testet, ob signifikante Korrelationen zwischen zwei Elementen unanhängig

von einem dritten Lokus bestehen bleiben. Dafür werden alle möglichen Dreierkombinationen im Datensatz berücksichtigt. Unabhängige, signifikante Korrelationen werden losgelöst von der Stärke der Korrelation erkannt. Im Anschluss kann auf eine minimale Korrelationsstärke von beispielsweise $|r| \geq 0{,}8$ gefiltert werden.

Ergänzend zur Koexpression von Loci wurden Pearson-Korrelationen zwischen Genexpression und Abundanz von Metaboliten im Blutplasma berechnet. Insbesondere in Studie 3 wurde von der klassischen Pearson-Korrelation anstelle der PCIT Gebrauch gemacht, um Beziehungen zwischen lncRNA-Expression und Metabolitenniveau im Plasma abzubilden.

Anhand der berechneten Korrelationen wurden Koexpressions- bzw. Korrelationsnetzwerke erstellt, deren Visualisierung mit Cytoscape 3.6.1 (Shannon et al., 2003) realisiert wurde.

Vorhersage potenzieller biologischer Funktionen von lncRNAs durch Pathway Enrichment Analysen und Integration von Metabolom und Transkriptom

In Studie 1 und 2 wurden hochvernetzte (Hub-) lncRNAs im Koexpressionsnetzwerk identifiziert und die mit ihnen korrelierten Gene und Metabolite wurden einer Pathway Enrichment Analyse in Ingenuity Pathway Analysis (IPA: QIAGEN Inc., https://www.qiagenbioinformatics.com/ products/ingenuitypathway-analysis (Kramer et al., 2014)) unterzogen, um angereicherte biologische Schaltkreise aufzudecken (Ansatz 1 in Abbildung 4).

In Studie 3 wurden alle zwischen den zwei Stoffwechselgruppen different exprimierten (DE) Gene und different abundanten (DA) Plasmetaboliten in eine Pathway Enrichment Analyse und eine gemeinsame Netzwerkanalyse in MetaboAnalyst 4.0 (Chong et al., 2019) eingespeist, die im Wesentlichen KEGG-Pathways betrachtet. Anschließend wurden die Gene und ausgewählte Metabolite der angereicherten biologischen Schaltkreise auf Korrelationen mit lncRNAs untersucht (Ansatz 2 in Abbildung 4).

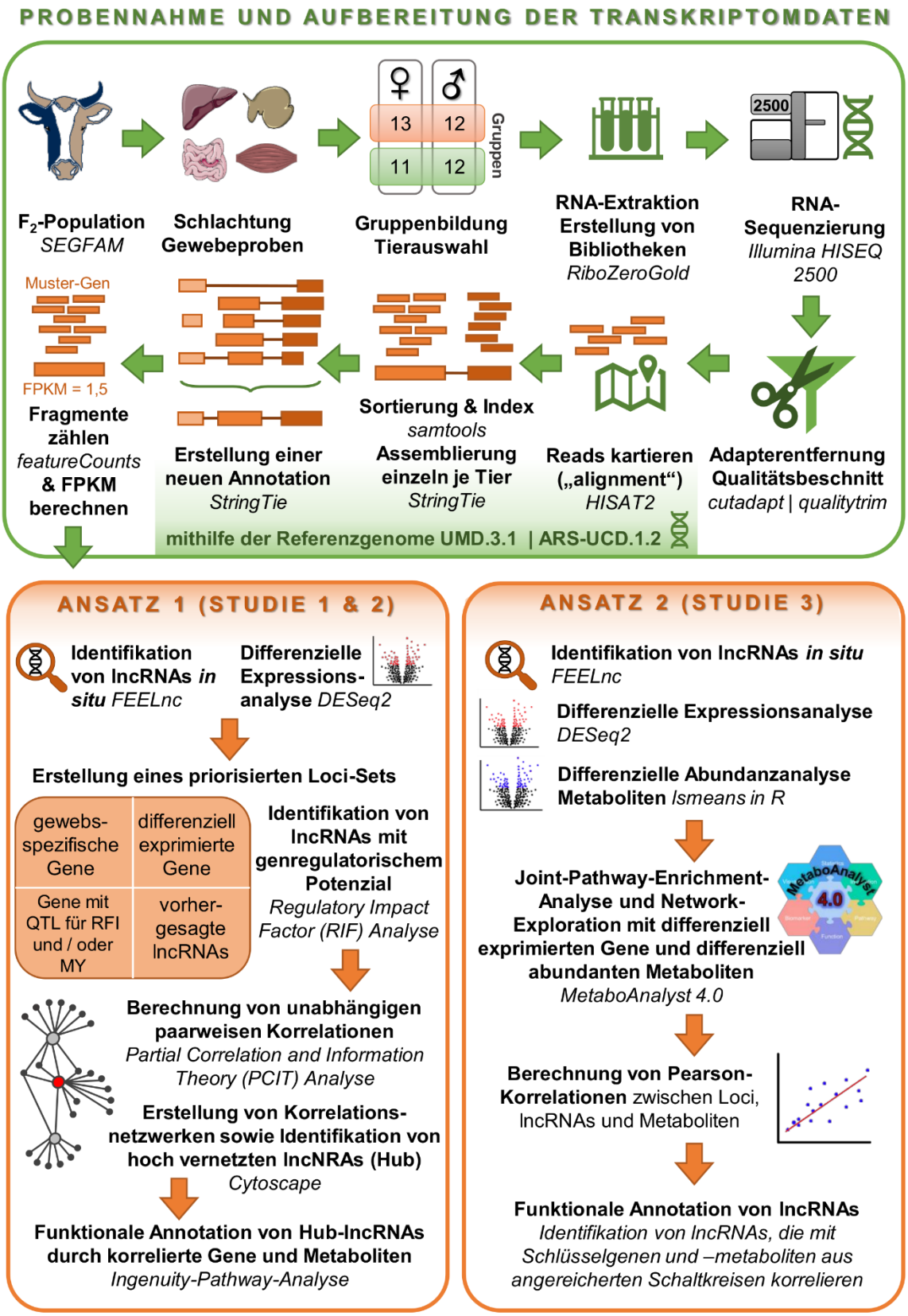

Abbildung 4 | Arbeitsschritte (fettgedruckt) zur Identifikation und funktionalen Annotation von lncRNAs mit potenziell regulatorischem Einfluss auf Stoffwechselprozesse und Nährstoffverteilung beim Rind unter Angabe der Methode oder Programme (kursivgedruckt). RFI = Residuale Futteraufnahme, MY = Milchleistung, FPKM = Fragments per Kilobase of Transcript per Million Reads, QTL = Quantitative Trait Locus [eigene Darstellung]

6. Auflistung der Veröffentlichungen

Hiermit erkläre ich, dass sich mein Beitrag zu den eingebrachten Publikationen wie folgt gestaltete:

Publizierte Fachartikel (Publikationen 1 und 2)

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2019). Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated with Metabolic Efficiency in Cattle. *Frontiers in Genetics* 10:1130. doi: 10.3389/fgene.2019.01130. **(IF 3,789** [2019]**)**

- Aufbereitung Transkriptomdaten: Adapter- und Qualitätsfilter, Alignment, Assembly, Erstellung der projektspezifischen Annotation, Read-Zählung
- Datenanalyse: RIF, PCIT, Korrelationsberechnungen, Netzwerk- und Pathway-Analysen
- Dateninterpretation und Diskussion
- Erstellung von Grafiken und Tabellen
- Erstellung des ersten Manuskriptentwurfes inkl. Literaturrecherche
- Überarbeitung des Manuskriptes
- **Gesamtbeitrag: 70 %**

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2020). Identification and Annotation of Potential Function of Regulatory Antisense Long Non-Coding RNAs Related to Feed Efficiency in *Bos taurus* Bulls. *International Journal of Molecular Sciences* 21(9). doi: 10.3390/ijms21093292. **(IF 4,556** [2019]**)**

- Aufbereitung Transkriptomdaten: Adapter- und Qualitätsfilter, Alignment, Assembly, Erstellung der projektspezifischen Annotation, Read-Zählung
- Datenanalyse: RIF, PCIT, Korrelationsberechnungen, Netzwerk- und Pathway-Analysen
- Dateninterpretation und Diskussion
- Erstellung von Grafiken und Tabellen
- Erstellung des ersten Manuskriptentwurfes inkl. Literaturrecherche
- Überarbeitung des Manuskriptes
- **Gesamtbeitrag: 70 %**

Fachartikel unter Begutachtung (Publikation 3)

Nolte W, Weikard R, Albrecht E, Hammon HM, Kühn C. Metabogenomic analysis to functionally annotate the regulatory role of long non-coding RNAs in the liver of cows with different nutrient partitioning phenotype. **(eingereicht am 7. November 2020 und unter Begutachtung seit dem 28. November 2020 bei Genomics, IF 6,205** [2019]**)**

- Aufbereitung Transkriptomdaten: Adapter- und Qualitätsfilter, Alignment, Assembly, Erstellung der projektspezifischen Annotation, Read-Zählung
- Datenanalyse: Korrelationsberechnungen, Netzwerk- und Pathway-Analysen
- Erstellung von Grafiken und Tabellen
- Beitrag zur Erstellung des ersten Manuskriptentwurfes inkl. Literaturrecherche
- **Gesamtbeitrag: 50 %**

6.1. Zusammenfassung der Veröffentlichungen

1. Studie – Fragestellungen:

- Welche lncRNAs werden bei Bullen und Kühen in den Geweben Leber, Pansen, Jejunum und Skelettmuskel exprimiert und liegen gewebsspezifische Expressionsmuster vor?
- Welche lncRNAs sind mit den Zielmerkmalen signifikant assoziiert, d.h. sind mit bekannten QTLs für Futtereffizenz oder Milchleistung kolokalisiert?
- Welche lncRNAs sind in ihrer Expression in den vier Geweben zwischen den beiden Stoffwechselgruppen signifikant verschieden?
- Welche lncRNAs weisen ein hohes genregulatorisches Potenzial innerhalb dieser vier Gewebe auf?
- Welche lncRNAs haben eine Schlüsselrolle bei der Regulation der Nährstoffverteilung?
- In welche biologischen Schaltkreise sind diese lncRNAs involviert?

In der **ersten Studie** (Nolte et al., 2019) wurden Bullen (N = 24) und Kühe (N = 24) mit divergenter Nährstoffverteilung untersucht. Durch die Integration von phänotypischen, transkriptomischen und metabolomischen Daten sollten lncRNAs mit einer potenziellen regulatorischen Schlüsselrolle für die Nährstoffverteilung beim Rind identifiziert und funktionell charakterisiert werden. Basierend auf Transkriptomdaten aus vier Geweben (Pansen, Jejunum, Leber, Skelettmuskel) wurde eine projektspezifische Transkriptannotation mit 30.072 Loci erstellt, in der *in silico* 7.646 Transkripte, die zu 3.287 Loci gehörten, als lncRNAs klassifiziert wurden. Für alle Tiere lagen zudem Metabolitenprofile aus dem Plasma für je 640 Metaboliten vor. Zunächst wurden für jedes Gewebe ein priorisiertes Locus-Set erstellt, das Loci enthielt, die eine besondere Bedeutung für die Nährstoffverteilung beim Rind aufwiesen, d.h. zwischen den Stoffwechselgruppen differenziell exprimiert waren, mit einem QTL für Futtereffizienz oder Milchleistung kolokalisiert waren, gewebsspezifisch exprimiert waren oder als lncRNA identifiziert worden waren.

Insgesamt waren 2.154 Loci zwischen den Stoffwechselgruppen DE (davon 238 lncRNAs). Eine gewebsspezifische Expression wurde für 279 Loci im Jejunum, für 283 Loci in der Leber, für 204 Loci im Skelettmuskel und für 164 Loci im Pansen festgestellt. Unter diesen 930 gewebsspezifisch exprimierten Loci waren 204 lncRNAs. In dieser Studie konnten keine lncRNAs gefunden werden, die mit einem bekannten QTL für RFI oder Milchleistung überlappten. Zur Auffindung von lncRNAs mit hohem (gen-) regulatorischen Potenzial in einem der vier Gewebe wurde der RIF angewendet. Für 92 (Jejunum), 55 (Leber), 35 (Skelettmuskel)

und 73 (Pansen) lncRNAs wurden hohe RIF-Punktzahlen berechnet, was auf ein regulatorisches Potenzial im jeweiligen Gewebe schließen ließ. Mithilfe der PCIT-Analyse wurden jeweils innerhalb der zwei Stoffwechselgruppen Korrelationen zwischen lncRNAs und proteinkodierenden Genen aus den priorisierten Locus-Sets berechnet. Anhand der Korrelationen wurden zwei Netzwerke generiert (eines spezifisch für jede Stoffwechsegruppe), die 1.522 bzw. 1.732 Loci enthielten. Insgesamt wurden acht lncRNAs, die eine hohe Vernetzung (über 100 Loci) in einem der zwei Netzwerke vorweisen konnten, als Hub-lncRNAs klassifiziert: *MSTRG.4740*, *MSTRG.4926*, *MSTRG.9051*, *MSTRG.10337*, *MSTRG.17681*, *MSTRG.18433*, *MSTRG.19098* und *MSTRG.19312.* Für diese lncRNAs wurde eine mögliche Schlüsselrolle bei der Regulation der bovinen Nährstoffverteilung postuliert. Die Expressionsprofile dieser Hub-lncRNAs wurden zusätzlich auf Korrelationen mit der Abundanz von Plasmametaboliten untersucht. Basierend auf den je mit einer Hub-lncRNA (N = 8) signifikant korrelierten Genen und Metaboliten wurden Pathway-Enrichment-Analysen in IPA durchgeführt. Die Ergebnisse dieser Analysen indizierten, dass die Hub-lncRNAs u.a. funktionell am hepatischen Aminosäurestoffwechsel, Proteinsynthese sowie am Kalzium-Signalweg und am neuronalen Stickoxid-Synthase-Signalweg in Skelettmuskelzellen beteiligt waren.

2. Studie – Fragestellungen:

- Welche lncRNAs zeigen in der Leber von Bullen mit unterschiedlicher Futtereffizienz bzw. Nährstoffverteilung ein hohes genregulatorisches Potenzial?
- Welche lncRNAs sind in ihrer Expression in der Leber zwischen den beiden Bullengruppen signifikant verschieden?
- Welche Metaboliten sind in ihrer Abundanz im Plasma zwischen den beiden Bullengruppen signifikant verschieden?
- Mit welchen und mit wie vielen Genen und oder Plasmametaboliten sind diese lncRNAs in ihrer Expression korreliert?
- Welche lncRNAs haben eine Schlüsselrolle bei der Regulation der Nährstoffverteilung von Bullen?
- In welche biologischen Schaltkreise sind diese lncRNAs involviert?
- Welche dieser lncRNAs sind natürliche antisense Transkripte zu proteinkodierenden Genen und wie sind die Korrelationen der Expressionswerte charakterisiert?

In der **zweiten Studie** wurde der Arbeitsablauf der ersten Studie verfeinert und an das neue Referenzgenom ARS-UCD.1.2 adaptiert: Hier wurde der Schwerpunkt auf die funktionale Charakterisierung unbekannter lncRNAs in der Leber von Bullen (N = 24) gelegt, welche sich hinsichtlich ihrer Futtereffizienz unterschieden (Nolte et al., 2020). Zudem wurde das Metabolom eingehender untersucht und berechnet, welche Plasmametaboliten zwischen den

beiden Stoffwechselgruppen in signifikant verschiedenen Niveaus vorlagen. Von 640 Metaboliten zeigten 45 eine signifkant differenzielle Abundanz, darunter die Aminosäuren Asparagin, Methionin, Glutamin und Cystein. Eine PCA, basierend auf den Metabolitendaten, zeigte eine klare Auftrennung der Tiere in die zwei Stoffwechselgruppen anhand der ersten beiden Hauptkomponenten. Die aktualisierte projektspezifische Transkriptannotation enthielt nun 30.806 Loci, wovon 6.161 Transkripte, die zu 3.495 Loci gehörten, als lncRNA eingestuft wurden. Anhand eines priorisierten Sets mit 4.666 Loci und einer darauf aufbauenden RIF-Analyse wurden 238 lncRNAs identifiziert, für die ein hohes genregulatoriscches Potenzial prognostiziert wurde. Insgesamt waren 745 Loci zwischen den beiden Stoffwechselgruppen DE, darunter 219 lncRNAs. Im priorisierten Locus-Set befanden sich 35 lncRNAs, die mit einem bekannten QTL für RFI überlappten.

Insgesamt wurden vier lncRNAs in dieser Studie für eine tiefergehende Charakterisierung ausgewählt: *MSTRG.4390*, *MSTRG.4802*, *MSTRG.5042* sowie *MSTRG.7472*. Die lncRNAs *MSTRG.5042* sowie *MSTRG.7472* waren NATs zu proteinkodierenden Genen (*Apolipoprotein A1* bzw. *Haptoglobin*) und durch eine starke signifikante Korrelation mit ihrem jeweiligen Wirtsgen gekennzeichnet ($r \geq 0.97$). Die lncRNAs *MSTRG.4390* und *MSTRG.5042* waren im Koexpressions-Netzwerk sowohl miteinander als auch mit vielen Plasmametaboliten korreliert (44 bzw. 45 Metaboliten). Hervorzuheben ist die lncRNA *MSTRG.4802,* die durch Korrelationen mit über 50 annotierten Genen hochvernetzt war und außerdem als DE und QTL-überlappend kategorisiert war. Weiterhin war diese lncRNA ein NAT zum Gen *UQCRB*, mit dem sie durch eine signifikante Korrelation mittlerer Stärke ($r = 0.69$) verbunden war. Pathway-Enrichment-Analysen dieser lncRNAs, basierend auf den mit ihnen korrelierten Genen und Metaboliten, legten nahe, dass sie u.a. funktional an der mitochondrialen Funktion, dem Signalweg der Akuten-Phase-Reaktion, dem Zitratzyklus, der Fettsäure-β-Oxidation und vermutlich der Gluconeogenese beteiligt sind. Die Hälfte der lncRNAs mit einer sehr hohen RIF-Punktzahl ($N = 119$) befand sich in antisense-Ausrichtung und Überlappung mit einem proteinkodierenden Gen. Eine Untersuchung der Koexpression dieser NAT-lncRNAs und ihren sogenannten Wirts-Genen ergab, dass für 44 lncRNA-Gen-Paare eine signifikante Korrelation vorlag, die in 95 % der Fälle positiv war. Bei acht lncRNA-Gen-Paaren bestand sogar eine signifikante Korrelation mit $r \geq 0.9$. Diese Gen-lncRNA-Beziehungen deuten darauf hin, dass die NAT-lncRNAs eine stabilisierende Funktion für ihre *cis*-korrelierten Gene und eine putative regulatorische Rolle bei der Genexpression haben könnten.

3. Studie – Fragestellungen:

- Welche lncRNAs werden in der Leber von laktierenden Kühen exprimiert?
- Welche lncRNAs sind in ihrer Expression in der Leber von laktierenden Kühen mit unterschiedlicher Nährstoffverteilung signifikant unterschiedlich?
- Welche Metaboliten sind in ihrer Abundanz im Plasma zwischen den beiden Kuhgruppen signifikant unterschiedlich?
- Welche lncRNAs haben eine Schlüsselrolle bei der Regulation der Nährstoffverteilung von laktierenden Kühen?
- Gibt es lncRNAs, die natürliche antisense Transkripte zu proteinkodierenden Genen sind, die eine zentrale Rolle in angereicherten biologischen Schaltkreise innehaben?

In der **dritten Studie** wurden in Korrelationsanalysen Metabolomdaten und Transkriptomdaten integriert, um über Netzwerk- und Pathwayanalysen die funktionale Rolle von unbekannnten lncRNAs in der Leber von laktierenden Kühen (N = 24) mit divergenter Nährstoffverteilung zu untersuchen (Nolte et al., unveröffentlicht). Zu diesem Zweck wurde neben dem hepatischen Transkriptom das Metabolom im Plasma (640 Metaboliten) eingehender untersucht. Insgesamt waren 183 Metaboliten zwischen den Gruppen different abundant, wobei signifkant höhere Plasmawerte für eine Vielzahl von Metaboliten des Lipidstoffwechsels bei Kühen festgestellt wurden, die durch eine hohe Milchleistung und einen geringeren Fettansatz gekennzeichnet waren, d.h. im Sekretionstyp standen. Eine differenzielle Expressionsanalyse (DEA) ergab, dass 2.114 Loci zwischen den Stoffwechselgruppen DE waren, darunter 247 lncRNAs. Auf Seiten der proteinkodierenden Gene wurden die stärksten Unterschiede im Expressionsniveau von *FGF21*, *MFSD2A*, *ANGPTL4*, *APOA4, SLC25A47*, *LIPG* und *ADIPOR2* detektiert. Anhand von DE Genen und DA Metaboliten wurden mithilfe der KEGG-Pathway-basierten Online-Plattform MetaboAnalyst 4.0 Pathway-Enrichment-Analysen und Netzwerkanalysen durchgeführt. Die Ergebnisse indizierten, dass v.a. die biologischen Schaltkreise der Argininbiosynthese, des Glycerolipid-Stoffwechsels, der Proteinprozessierung im Endoplasmatischen Retikulum, des PPAR-Signalweges sowie des Aminosäure- und Fettstoffwechsels angereichert waren. Eine eingehende Betrachtung der Ergebnisse ließ vermuten, dass den Unterschieden zwischen den Stoffwechselgruppen Unterschiede in der Aktivität des Energie- und Lipidstoffwechsel, des Harnstoff- und Zitratzyklus‘ und der Gluconeogenese zugrunde lagen. Insbesondere das *FGF21*-Gen, dessen Expression mit einer Fülle von differenziell exprimierten Genen, differenziell abundanten Metaboliten sowie lncRNAs korrelierte, deutete sich als wichtiger Stoffwechselregulator an. Die lncRNA *XLOC_010660* war ein NAT zu *FGF21* und beide Loci waren ko-exprimiert und durch eine starke signifikante Korrelation (r = 98) miteinander verbunden. Weitere lncRNAs mit signifikanter und starker Korrelation ($|r| \geq 0{,}8$) zu *FGF21* waren u.a. *XLOC_015138*, *XLOC_023300*, *XLOC_006438*, *XLOC_010593* and

XLOC_028970. Eine Schlüsselrolle bei der Regulation der Nährstoffverteilung von laktierenden Kühen wurde jenen lncRNAs zugeschrieben, die mit DE-Genen bzw. Schlüsselgenen oder -metaboliten aus diesen Signal- und Stoffwechselwegen signifikant korreliert waren. Zu solchen funktional bedeutsamen lncRNAs gehörten z.B. *XLOC_028970*, *XLOC_030378*, *XLOC_004875*, *XLOC_018415*, *XLOC_022793*, *XLOC_018898* und *XLOC_006438*. Sowohl intergenische lncRNAs in enger physischer Nachbarschaft zu kodierenden Genen sowie lncRNAs, die NATs zu proteinkodierenden Schlüsselgenen aus den angereicherten Schaltkreisen waren, üben wahrscheinlich eine *cis*-regulatorische Funktion auf Gene aus, die in den Fett-, Aminosäure- und Energiestoffwechsel involviert sind oder eine wichtige Rolle in der Entgiftung über den Harnstoffzyklus spielen. Diese lncRNAs scheinen eine Feinabstimmungsfunktion in der Expression dieser Gene zu besitzen.

7. Studienübergreifende Ergebnisse und Diskussion

Der Konsum tierischer Produkte hat sich weltweit innerhalb der letzten 60 Jahre drastisch verändert. Seit Beginn der 1960er Jahre ist der durchschnittliche Milchverbrauch pro Kopf und Jahr um 29 % angestiegen und hinsichtlich des Fleischkonsums konnte sogar eine Verdopplung von ehemals 23,08 kg (1961) auf 43,22 kg (2013) festgestellt werden (Ritchie, 2017). In Deutschland lag der Pro-Kopf-Verbrauch von Frischmilcherzeugnissen 2019 bei 86,43 kg, wobei sich eine leicht rückläufige Tendenz abzeichnet (Statista GmbH, 2015). Der Fleischkonsum lag zum selben Zeitpunkt bei 59,5 kg pro Kopf, davon 10 kg Rindfleisch (Statista GmbH, 2020). In industrialisierten Regionen wie Europa lässt sich bezüglich des Rindfleisches eine Präferenz für fettarmes Fleisch ausmachen, das von Konsumenten als gesünder eingeschätzt wird (Verbeke et al., 2010). Wie von Ferguson (2010) zusammengefasst, wurde der erhöhte Konsum von rotem Fleisch wiederholt mit der Entwicklung von (kolorektalem) Krebs in Verbindung gebracht, wobei es Hinweise darauf gibt, dass das erhöhte Krebsrisiko u.U. auf den Fettgehalt und die Fettsäurezusammensetzung des Fleisches zurückzuführen ist (Ferguson, 2010). Damit ist ein Umsatz von Nährstoffen in die Milchproduktion oder den Ansatz von Muskelmasse einer Einlagerung in Körperfett vorzuziehen. Neben der Komponente der Fütterung, beeinflussen auch die Genetik bzw. die Rasse den Fettgehalt und Fettsäurespiegel von Fleisch (Hocquette et al., 2010; Sevane et al., 2014) und Milch (Van Eijndhoven et al., 2011). Eine gezielte Züchtung von Rindern für den effizienten Fleischansatz ist auch deswegen vorteilhaft, weil die Umwandlung von Nährstoffen aus dem Futter in Proteine kalorisch etwas weniger aufwändig ist als die Fetteinlagerung (Jeroch et al., 2008).

Die Definition eines Merkmals für die Präferenz der Nährstoffverteilung („nutrient partitioning“) beim Rind gestaltet sich schwierig, da keine einheitliche Messgröße für diesen Phänotyp vorliegt. Ein konventionell verwendeter Parameter zur Messung der Futtereffizienz ist der RFI (Koch et al., 1963). Bei Kühen ist zusätzlich die Erhebung der Energie-korrigierten Milchleistung (ECM) (Kirchgeßner, 1997), ggf. in Kombination mit der Futteraufnahme, eine anerkannte Messgröße. Die Nährstoffverteilung in Form von Fettakkretion lässt sich beispielsweise als Fettgehalt des Schlachtkörpers (CFC) oder als IMF messen. Mit der SEGFAM-Population wurde gezielt eine F_2-Population erzeugt, die hinsichtlich der Nutzung von Nährstoffen für den Fett-und Muskelansatz bzw. die Sekretion von Proteinen durch Milch große Variabliltät aufwies und sich zur Untersuchung der genetischen Hintergründe des Stoffwechseltyps und somit der Präferenz für die Nährstoffverteilung eignet (Kühn et al., 2002).

7.1. Charakterisierung von Stoffwechseltypen beim Rind über Metabolom- und Gesamttranskriptomanalysen

In der vorliegenden Arbeit zeigte die Hauptkomponentenanalyse (PCA) anhand von Metabolomdaten sowohl in Bullen (Studie 2) als auch in Kühen (Studie 3), die aus der SEGFAM-F_2-Population rekrutiert wurden, eine klare Auftrennung der definierten phänotypischen Gruppen anhand der ersten und zweiten Hauptkomponente, die gemeinsam in beiden Geschlechtern jeweils mindestens 38 % der Varianz erklären. Auch über ein hierarchisches Clustering nach Ward konnte die Gruppierung der Tiere anhand von Metabolitenprofilen in Geschlechter und Stoffwechseltypen nachgewiesen werden (siehe Abbildung 5, unveröffentlicht). Da jedoch über die Metabolitenprofile von Metabolon keine absoluten Werte bzw. Plasmakonzentrationen der einzelnen Metabolite verfügbar waren, sondern nur die Werte für die Fläche unter der Kurve, lassen sich keine absoluten Daten für Metabolitenprofile erstellen, nach denen Tiere aus anderen Studien sich analog dem einen oder anderen Stoffwechseltypen zuordnen lassen. Basierend auf diesem Datensatz und den Auswertungen ist folglich keine Erstellung eines allgemein übertragbaren Gruppenmetabolitenprofils möglich. Obwohl beide Geschlechter sich deutlich über ihre Metabolienprofile in nach Nährstoffverteilung divergierende Gruppen auftrennen ließen, waren dennoch mehrheitlich verschiedene Metabolite für die Separierung der Tiere in Gruppen ausschlaggebend. Dies zeigte sich darin, dass Bullen und Kühe mehrheitlich für verschiedene Metabolite differente Abundanzen im Plasma aufwiesen (Studien 2 und 3). Lediglich folgende fünf Metabolite, die sowohl in Bullen als auch Kühen DA waren, waren in den geschlechtsspezifischen Pathway Enrichment Analysen vertreten: 4-Hydroxyglutamat, 2-Oxoarginin, Gamma-Glutamylglycin, Gamma-Glutamylglutamat und 10-Heptadecanoat (17:1n7). Bei Kühen (Studie 3) fiel auf, dass besonders viele Aminosäuren DA waren, während das bei Bullen (Studie 2) weniger und für andere Aminosäuren der Fall war.

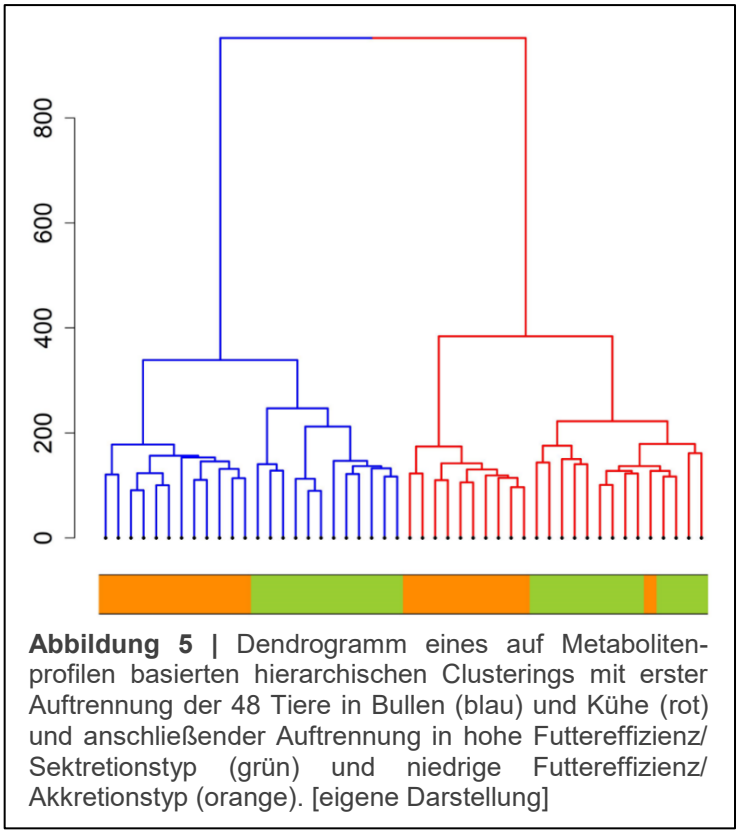

Abbildung 5 | Dendrogramm eines auf Metaboliten-profilen basierten hierarchischen Clusterings mit erster Auftrennung der 48 Tiere in Bullen (blau) und Kühe (rot) und anschließender Auftrennung in hohe Futtereffizienz/ Sektretionstyp (grün) und niedrige Futtereffizienz/ Akkretionstyp (orange). [eigene Darstellung]

In einer PCA basierend auf dem hepatischen Gesamttranskriptom war die Aufsplittung der Kühe in den Sekretionstypen (SEC) und den Akkretionstypen (ACC) in Studie 3 deutlich ersichtlich, während bei den Bullen in Studie 2 keine so eindeutige Auftrennung in Stoffwechselgruppen beobachtet werden konnte (Daten nicht publiziert). Dass die beiden Stoffwechselgruppen in den Kühen von deutlicheren Unterschieden geprägt waren, war auch in der substanziell höheren Anzahl von DE Loci in der Leber reflektiert. Während Kühe in der Leber über 2.114 signifikant (FDR ≤ 0,05) DE Loci verfügten, wovon 247 als lncRNA klassifiziert waren, konnten bei den Bullen lediglich 745 DE Loci detektiert werden (davon 219 lncRNAs), obwohl bei den Bullen sogar ein niedriges Signifikanzniveau von FDR ≤ 0,1 angenommen wurde. Insgesamt schien damit die Leberfunktion bei den untersuchten laktierenden Kühen beider Gruppen stärker differenziell reguliert zu sein als bei den Bullen. Es sei weiterhin vermerkt, dass Kühe beider Stoffwechselgruppen sowohl im IMF als auch im CFC über dem Niveau der Bullen der entsprechenden Stoffwechselgruppe lagen (siehe Studie 1), wobei die Differenzen der Mittelwerte der Stoffwechselgruppen von Bullen und Kühen vergleichbar waren.

7.2. Erstellung eines lncRNA-Katalogs für das Rind

Gegenwärtig existieren nur wenige wissenschaftliche Studien zu lncRNAs beim Rind, daher war das erste Ziel der vorliegenden Arbeit die Erstellung eines umfangreichen Katalogs für lncRNAs für das Rind, der die Basis für alle folgenden Analysen darstellte. Kosinska-Selbi und Kollegen listeten insgesamt 19 Publikationen für den Zeiraum von 2012 bis 2019 (exklusive der hier präsentierten Studien 1 und 2), wovon acht rein explorativer Natur sind und nur 11 eine DEA zwischen phänotypisch differenten Gruppen beinhalten (Kosinska-Selbi et al., 2020). Die in der hier vorliegenden Arbeit vorgestellten Studien stellen eine wertvolle Ergänzung zu den durch Kosinska-Selbi zusammengetragenen lncRNA-Untersuchungen am Rind dar. In der Vergangenheit konzentrierte sich etwa die Hälfte der Untersuchungen auf das Milchdrüsen- oder Muskelgewebe, wohingegen die Leber als zentrales Stoffwechselorgan in diesen Studien unterrepräsentiert war. Im Vergleich zu den angesprochenen Publikationen basieren unsere Arbeiten hinsichtlich der beprobten Tiere mit Abstand auf dem bis dato größten Datensatz, der zur Generierung der *de-novo*-Annotation und Identifikation von neuen lncRNAs genutzt wurde. Darüber hinaus zeichnet sich der Datensatz durch eine hohe Sequenzierdichte von paired-end Reads aus, was besonders einer Annotation von vergleichsweise niedrig exprimierten Loci wie lncRNAs zugutekommt. Insgesamt wurden 204 Proben genutzt, die aus vier Geweben mit je 48 (Jejunum) bzw. 52 (Leber, Muskel, Pansen) biologischen Replikaten stammen, wobei zudem beide Geschlechter berücksichtigt wurden. Gerade durch die häufig dokumentierte, gewebsspezifische Expression von lncRNAs ist es unumgänglich, sukzessive alle Gewebe eines Organismus zu untersuchen, um zu einem holistischen lncRNA-Atlas zu gelangen.

Basierend auf einer projektspezifischen Annotation wurden unter Zuhilfenahme der bovinen Referenzgenome UMD3.1 und ARS-UCD.1.2 in den vorliegenden drei Studien lncRNA-Kataloge erstellt. Die Annotation aus Studie 1 beinhaltete 7.646 lncRNA-Transkripte (3.287 Loci) und die Annotation in Studie 2 und 3 beinhaltete 6.161 lncRNA-Transkripte (3.590 Loci). Die Einführung eines neuen Referenzgenoms während der Projektbearbeitung bedingte, dass die Ergebnisse aus der ersten Studie sich nur mit erweitertem Aufwand direkt mit den Ergebnissen aus der zweiten und dritten Studie vergleichen lassen. Mit je 3.287 bzw. 3.590 vorhergesagten lncRNA-Loci, die in unseren Studien jeweils 10 %,15 % (Studie 1) bzw. 11,65 % (Studie 2 und 3) aller in der projektspezifischen Transkriptannotation verzeichneten Genorte ausmachten, zeigt sich der substanzielle Umfang dieser Klasse von RNAs, die bisher in molekulargenetischen Studien zum Rind weitesgehend unberücksichtigt geblieben sind. Aus der unzureichenden Anzahl der Studien ergibt sich auch die Problematik der funktionalen, biologischen Annotation jeder einzelnen lncRNA. In diesem Sinne haben die systembiologischen, netzwerkbasierten Ansätze einen sehr guten ersten Anhaltspunkt geliefert. Aufgrund der Aktualität des neuen Referenzgenoms ARS-UCD.1.2 wird im Folgenden auf den darauf aufbauenden lncRNA-Katalog (Studien 2 und 3) Bezug genommen.

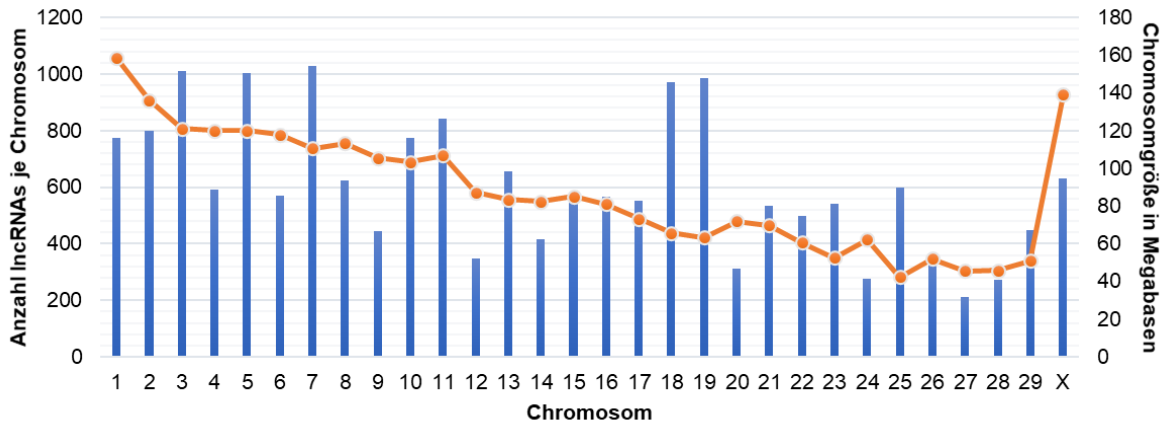

Abbildung 6 | Anzahl der identifizierten lncRNA-Loci je Chromosom (blaue Balken, primäre Y-Achse) neben der Chromsomengröße in Megabasen (orange, sekundaäre Y-Achse). [eigene Darstellung; Daten aus Studien 2 und 3]

Unter Ausschluss der lncRNA-Loci, die sich auf Contigs[7] (N = 138) und im mitochondrialen Genom (N = 12) befanden, verteilen sich die identifizierten lncRNA-Loci relativ gleichmäßig über die Autosomen und das X-Chromosom (siehe Abbildung 6), wobei eine Pearson-Korrelation mittlerer Stärke (r = 0,60, p = 3,147-E04) zwischen der Größe des jeweiligen Chromosoms[8] und der darauf gefundenen Anzahl von lncRNA-Loci besteht. Auf den

[7] Ein Contig bezeichnet einen Abschnitt zusammenhängender und sich überlappender Reads, die in der Genomassemblierung keine chromosomale Zuordnung erfahren haben.

[8] Die Länge der Chromosomen in Basenpaaren wurde dem Bericht zur ARS-UCD.1.2-Assembly auf NCBI entnommmen (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/#/st, Zugriff am 29.11.2020)

Chromosomen 18, 19, 23 und 25 finden sich jeweils über 10 lncRNA-Transkripte je Megabase, was für die Autosomen die höchsten Quotienten darstellt und sich mit Ergebnissen von Weikard et al (2013) zu lncRNAs in der Haut von Rindern deckt.

Es kann davon ausgegangen werden, dass diese Transkripte in Teilen eine Erweiterung der im November 2020 in der NONCODE-Datenbank (Version 6) für das Rind verzeichneten 23.515 lncRNA-Transkripte bzw. 22.227 lncRNA-Loci darstellen. Neben den vorangegangenen Arbeiten der FBN-Arbeitsgruppe der Genomphysiologie (Weikard et al., 2013, 2018), in denen lncRNAs in Haut und Jejunum identifiziert wurden, haben die Studien 1 bis 3 damit zur Erweiterung des bovinen lncRNA-Kataloges beigetragen. Bei der grundlegenden Charakterisierung der lncRNAs konnte festgestellt werden, dass das Expressionsniveau in allen Geweben unter dem mittleren Expressionsniveau der mRNAs lag (siehe Abbildung 7), was sich mit Angaben in der Fachliteratur zu anderen domestizierten Spezies deckt (Le Béguec et al., 2018; Muret et al., 2017).

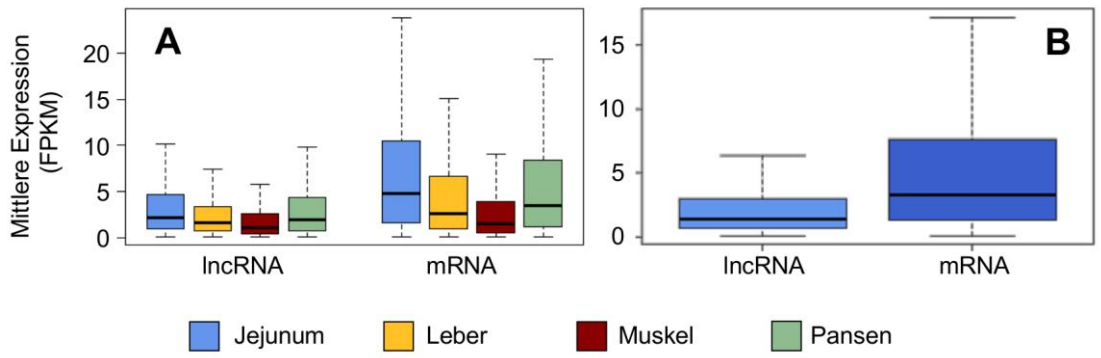

Abbildung 7 | Whisker-Boxplot ohne Ausreißer: Die mittlere Expression der lncRNAs rangiert jeweils in allen (A) untersuchten Geweben (Jejunum, Leber, Muskel, Pansen) sowie im Durchschnitt (B) aller Gewebe unter dem Niveau der mRNAs. [eigene Darstellung; Daten aus Studie 1]

7.3. Identifikation von lncRNAs mit genregulatorischem Potenzial

Ein weiteres Ziel dieser Arbeit war, nach ihrer Identifikation, lncRNAs mit hohem genregulatorischen Potenzial in Bezug auf die Zielmerkmale zu identifizieren. Dies wurde in Studien 1 und 2 durch die Anwendung der RIF-Analyse (Reverter et al., 2010) realisiert. Anschließend erfolgte mittels der PCIT die Bestimmung von signifikanten paarweisen Korrelationen zwischen Loci, die unabhängig von der Expression dritter Gene bestanden (Reverter & Chan, 2008). Dieser kombinierte Ansatz von PCIT und RIF wurde bereits erfolgreich zur Aufdeckung von regulatorischen Netzwerken und genomischen transkriptionsmodulierenden Regulatoren verwendet (Canovas et al., 2014; Nguyen et al., 2018; Perez-Montarelo et al., 2012). In Studien 1 und 2 erhielten jeweils 240 und 238 lncRNAs eine signifikante RIF-Punktzahl und somit die Zuschreibung eines hohen regulatorischen Potenzials.

Da lncRNAs im Vergleich zu proteinkodierenden Genen über eine geringe Sequenzkonservierung verfügen (Derrien et al., 2012), aber eine gewisse syntenische bzw. funktionale Konservierung zwischen Spezies vorhanden ist (Hezroni et al., 2015; Owens et al., 2019; Ulitsky et al., 2011), liegt die Annahme nahe, dass die strukturelle Konservierung in dieser RNA-Klasse von höherer Bedeutung ist als die primäre Struktur (Zampetaki et al., 2018). Die Ergebnisse in Studien 1 und 2 aus der RIF-Analyse, insbesondere die RIF2-Scores, deuten aber daraufhin, dass sich zwischen phänotypischen Gruppen die Korrelationsrichtung (positiv, negativ) einer lncRNA mit einem anderen Gen stark verändert, was auf ein verändertes Bindungsverhalten zurückzuführen sein könnte. Für Transkriptionsfaktoren ist bereits nachgewiesen, dass Sequenzunterschiede ihr Bindungsverhalten verändern können (Johnston et al., 2019). Das legt nahe, dass bei einigen lncRNAs das Bindungsverhalten zu anderen Transkripten bzw. DNA-Abschnitten durch Polymorphismen und andere Sequenzänderungen moduliert wird, was in dieser Dissertation jedoch nicht weiter beleuchtet wurde.

Hinsichtlich der Ergebnisse der RIF-Analysen ist hervorzuheben, dass lediglich 10 % der lncRNAs mit einer signifikanten RIF-Punktzahl in Studie 1 auch zwischen den Stoffwechselgruppen DE waren. Darin zeigt sich ein Vorteil der Methode, da die Bestimmung einer differenziellen Expression von lncRNAs häufig aufgrund ihres sehr niedrigen Expressionsniveaus behindert wird. Jedoch erwies sich die Ergebnisinterpretation in einem gewebeübergreifenden Ansatz als kompliziert, da lncRNAs oft gewebsspezifisch exprimiert werden (Cabili et al., 2011; Ulitsky et al., 2011; Washietl et al., 2014). Zudem ergab die RIF-Analyse, dass nur wenige lncRNAs in mehr als einem Gewebe eine signifikanten RIF-Punktzahl vorweisen konnten und eine lokale regulatorische Funktion u.U. naheliegender ist. Diese Erkenntnisse waren mit ausschlaggebend für die Entscheidung, die Folgestudien 2 und 3 auf ein Gewebe (Leber) zu beschränken. Dies geschah ungeachtet der Tatsache, dass neben der Leber als zentrales Stoffwechselorgan auch die Expressionsprofile anderer Gewebe einen wichtigen Beitrag zum Phänotypen leisten können. Kashi und Kollegen, die aufzeigten, dass in der Mehrheit der bisherigen Studien zu lncRNAs Gewebeproben untersucht wurden, regen sogar dazu an, in Zukunft mehr Single-Cell-Analysen durchzuführen (Kashi et al., 2016).

Trotzdem unterstreichen die lncRNAs mit signifkanter RIF-Punktzahl, die eine sehr hohe Vernetzung in den Koexpressionsnetzwerken aufwiesen, dass die RIF-Metriken gut geeignet sind, um potenziell zentrale Regulatoren auf transkriptioneller Ebene zu isolieren, wie beispielsweise lncRNA *MSTRG.4802* in Studie 2. Die Kombination mehrerer Kriterien wie die RIF-Metrik, eine differenzielle Expression sowie die Kolokalisation mit einem QTL, der mit dem untersuchten Merkmal (RFI) assoziiert war, stellt einen soliden Ansatz dar, der auch zukünftig zur Priorisierung von unbekannten lncRNAs vielversprechend ist. Seit der Einführung von

GWAS (Risch & Merikangas, 1996) wurde dieser Ansatz vielfach zur Identifikation von Genomregionen und kausalen Polymorphismen genutzt, die verschiedensten Merkmalen zugrunde liegen. Traditionell wird anschließend im QTL-nahen Genomabschnitt nach annotierten Genen gesucht, die sich auch aufgrund vorheriger Erkenntnisse als Kandidatengene anbieten. Ergebnisse von Meta-Analysen basierend auf GWAS zur bovinen Körpergröße (Bouwman et al., 2018), Milchdrüsenmorphologie (Pausch et al., 2016) und weiteren ökonomisch bedeutsamen Merkmalen (Fang & Pausch, 2019) zeigten jedoch, dass wiederholt hochsignifikante QTL außerhalb von proteinkodierenden Genen im intergenischen bzw. nichtkodierenden Bereich gefunden wurden. Die in Studie 2 beispielhaft näher untersuchte lncRNA *MSTRG.4802* stützt, zusammen mit weiteren 83 lncRNAs, die mit einem QTL für RFI zusammenfielen, die These, dass QTL nicht notwendigerweise an proteinkodierende Gene geknüpft sind und eine eingehendere Untersuchung der QTL-Regionen speziell auf nichtkodierende Elemente angebracht ist. Der kombinierte Ansatz von RIF und PCIT zur Identifikation regulatorischer lncRNAs wurde kürzlich durch Alexandre und Kollegen in einer Studie zur Futterefffizienz bei Nellorerindern übernommen. Auch in dieser Studie konnten lncRNAs mit signifikanter RIF-Punktzahl in QTL-Regionen für Futtereffizienz gefunden werden (Alexandre et al., 2020).

7.4. LncRNAs mit potenziell merkmalsmodulierendem Einfluss und ihre funktionelle Annotation

Eine weitere Zielstellung war die Vorhersage möglicher biologischer Funktionen ausgewählter lncRNAs, die ein hohes regulatorisches Potenzial zeigten oder anderweitig eng mit Stoffwechselprozessen verbunden waren. Dafür war die Identifikation von DE lncRNAs zwischen Stoffwechseltypen fundamental. Insgesamt waren in den drei Studien jeweils 238, 219 und 247 lncRNAs DE (auf Lokusebene berechnet). Insbesondere hinsichtlich der differentiellen Expression zwischen Stoffwechselgruppen konnte in Studie 1 eine gewisse Gewebsspezifität beobachtet werden, die auch unter Berücksichtigung des Geschlechts standhielt. Der Fakt, dass im Pansen über Geschlechter hinweg keine DE-lncRNAs ausfindig gemacht werden konnten, legt nahe, dass in diesem Gewebe der Einfluss des sehr unterschiedlichen Fütterungsregimes der Geschlechter eine besondere Rolle spielt. So wurden Kühe mit einer totalen Mischration gefüttert, wohingegen Bullen eine fast ausschließlich kraftfutterbasierte Ration erhielten. Der Umstand, dass beide Geschlechter einem unterschiedlichen Fütterungsregime unterlagen, bedingt allgemein einen Störfaktor, d.h. Confounder, bei den statistischen Auswertungen, da die Effekte von Geschlecht und Fütterung sich nicht trennen lassen.

Interessant ist dabei, dass vier der DE lncRNA-Loci aus Studie 2 auch zwischen zwei Gruppen von Nellorerindern (*Bos indicus*) mit divergenter Futtereffizienz als DE detektiert wurden, wobei eine Übereinstimmung der *Bos indicus* und *Bos taurus* Sequenzen von 83 % bis 90 %

festgestellt wurde (Alexandre et al., 2020), die geringer ausfällt, als zunächst zu erwarten gewesen wäre. Hervorzuheben ist weiterhin, dass Alexandre und Kollegen insgesamt vier anderen lncRNAs, die eine Sequenzübereinstimmung von 82 % bis 100 % zu lncRNAs mit signifikanten RIF-Wert in unserer zweiten Studie aufwiesen, in der Leber der untersuchten Nellorerinder ebenfalls einen signifikanten RIF-Wert bescheinigen konnten. Bei Verwendung einer ähnlichen Methodik kristallisieren sich für *Bos indicus* und *Bos taurus* damit zum Teil dieselben lncRNA-Loci als hepatische Schlüsselregulatoren für die Futtereffizienz heraus.

In Studie 1 und 2 wurde zur funktionalen Charakterisierung von Schlüssel-lncRNAs eine Ingenuity-Pathway-Analyse (Kramer et al., 2014) basierend auf den mit der jeweiligen lncRNA korrelierten Genen und Metabolite durchgeführt (Ansatz 1 in Abbildung 4). Ein Schwachpunkt dieses Ansatzes war, dass eine funktionale Charakterisierung über Enrichmentanalysen schwer möglich oder womöglich fehlerbehaftet war, wenn eine unzureichende Anzahl von korrelierten Genen und Metaboliten vorlag. Im Handbuch für die Erstellung von Genlisten für GSEA wird beispielsweise eine minimale Anzahl von 15 und ein Maximum von 500 Genen je Analyse empfohlen (The Broad Institute, 2009) und Mooney und Kollegen sprechen sich für Gensetgrößen von 10 bis 200 Genen aus (Mooney & Wilmot, 2015). Die von uns in Studie 1 und 2 verwendete IPA-Software empfiehlt als ideale Setgröße in Genexpressionsstudien eine Anzahl von 200 bis 3000 Genen (QIAGEN, 2016). Zudem beeinflusst die Größe des Pathways, d.h. die Anzahl der involvierten Gene, die Ergebnisse einer Enrichmentanalyse, wobei Ergebnisverzerrungen vorrangig bei kleinen Pathways auftreten (Nguyen et al., 2019). Ein Nachteil bei den Pathway-Enrichment-Analysen in IPA ist, dass diese Plattform nur eine begrenzte Auswahl von Pathways einbezieht, wobei zudem vorrangig Informationen von Mensch und dem Modelltier Maus einfließen und ein Fokus auf wichtigen (patho-) physiologischen Abläufen liegt. In der Tat gestaltete sich die Ergebnisinterpretation der Enrichment-Analysen, die wir mit IPA durchgeführt haben, gelegentlich schwierig, wie z.B. im Falle von *MSTRG.4740* (Studie 1). Obwohl für diese hochvernetzte lncRNA (über 100 Korrelationspartner) eine ausreichende Anzahl signifikant und stark korrelierter Gene und Metbabolite vorlag, ergab die Pathway-Enrichment-Analyse eine signifikante Anreicherung für tRNA-Charging, wobei dies nur auf korrelierten Metaboliten (essenziellen Aminosäuren) beruhte und nicht durch Gene aus dem entsprechenden biologischen Prozess gestützt wurde. Hier haben wir die Ergebnisse daher vorsichtig als Involvierung in den Aminosäurestoffwechsel gedeutet. Auch Alexandre et al. konnte für hepatische Schlüssel-lncRNAs bei Bullen divergenter Futtereffizienz eine Anreicherung für Stoffwechsel essentieller Aminosäuren (Valin-, Leucin- und Isoleucinabbau) feststellen (Alexandre et al., 2020).

In Studie 3 wurde dagegen ein anderer Ansatz zur funktionalen Charakterisierung der lncRNAs verfolgt (Ansatz 2 in Abbildung 4). Während in Studie 1 und 2 lncRNAs mit hohem

regulatorischen Potenzial und hoher Vernetzung in Korrelationsnetzwerken identifiziert wurden und anhand ihrer Korrelationspartner einem Funktionskreis zugeschrieben wurden, haben wir in Studie 3 zunächst generell aktivierte Schaltkreise identifiziert und anschließend nach Schlüsselgenen und -metaboliten gesucht und diese nachfolgend hinsichtlich ihrer Korrelation mit lncRNAs untersucht. Der Vorteil beim in Studie 3 verfolgten Ansatz ist, dass die einzelnen lncRNAs nicht zwangsläufig Einfluss auf eine große Anzahl von Genen oder Metaboliten haben müssen, sondern auch einzelne und starke Korrelationsbeziehungen eine hohe Gewichtung erfahren. Gerade vor dem Hintergrund der natürlichen Antisense-Transkripte (NATs), die lokal in *cis* auf ihr Wirtsgen wirken, ist dieser Ansatz sinnvoll. Umso erstaunlicher ist, dass auch im Ansatz von Studie 1 und 2 viele NATs gefunden wurden, die neben einer hohen (positiven) Korrelation mit ihrem Wirtsgen auch Verbindungen zu vielen weiteren Loci aufwiesen. Sofern in Studie 2 eine signifikante Korrelation zwischen einer Schlüssel-lncRNA und einem gegenüberliegenden Gen vorlag, war diese in über 95 % der Fälle positiv. Dies deckt sich mit Ergebnissen anderer Studien, wonach die große Mehrheit der Korrelationen zwischen nichtkodierenden Elementen und antisense-Partnergenen positiv war (Grigoriadis et al., 2009) und insbesondere starke negative Korrelationen nur sehr selten beoachtet wurden (Wenric et al., 2017). Die Beeinträchtigung von antisense gelegenen RNAs vermag die Expression der mRNAs, die vom sense-Strang transkribiert werden, verändern. Daraus ergibt sich die Annahme, dass Antisensetranskription zur Kontrolle des übrigen Transkriptoms beim Säugetier beitragen kann (Katayama et al., 2005).

Eine wertvolle Ergänzung bei der Funktionsvorhersage stellte die Integration einer weiteren Informationsebene in Form metabolomischer Daten dar. Die Einbeziehung von Metaboliten bietet sich explizit an, da ihre Funktion und Wirkung in Schaltkreisen verhältnismäßig gut und speziesübergreifend definiert ist.

Auch wenn in den Studien 2 und 3 aufgezeigt wurde, dass nur wenige Loci und Metabolite in beiden Geschlechtern gleichermaßen DE waren, indizieren die Enrichmentanalysen aus IPA (Studie 2) und MetaboAnalyst (Studie 3) jeweils, dass in beiden Geschlechtern der PPARα bzw. PPAR-Signalweg signifikant angereichert war. Die Nutzung verschiedener Plattformen zur Enrichmentanalysse erschwert hierbei den direkten Vergleich. In Bullen konnte über die Auswertung von DE Genen eine signifikante Anreicherung für den PPARα/RXRα Signalweg festgestellt werden ($p \leq 0.01$), wobei eine Herabregulierung dieses Schaltkreises für hocheffiziente Bullen vorlag ($-\log_{10}(p) = 6.12$, z-score[9] $= -0.707$). In Studie 3 zeigte sich über die Enrichmentanalyse eine Anreicherung des PPAR-Signalwegs in der Leber von Kühen

[9] Ein positiver z-score gibt eine Heraufregulierung eines Pathways an und ein negativer z-score eine Herabregulierung. Die Stärke des z-scores bezieht sich auf die Abweichung des Ranges der beobachteten Anreicherung vom Rang der erwarteten Anreicherung eines Pathways (Chen et al., 2013).

($-\log_{10}(p) = 8.02$, $PI^{10} = 1.68$). Zusätzlich deuten die Ergebnisse aus Studie 3 darauf hin, dass der Schaltkreis der Proteinprozessierung im Endoplasmatischen Retikulum in der Leber mit differenziell zwischen Kühen verschiedener Stoffwechseltypen exprimtierten Genen angereichert ist ($-\log_{10}(p) = 11.93$, $PI = 0.24$). Aus Studie 2 lässt sich nicht nachvollziehen, ob dies auch bei Bullen u.U. der Fall war, da hier gezielt ribosomale Gene zu Beginn herausgefiltert wurden, die trotz einer ribosomalen Depletion bei der Aufbereitung der Bibliotheken vor der RNA-Seq sequenziert wurden. Während in Kühen basierend auf allen DE Genen und DA Metaboliten weiterhin eine Aktivierung der Argininbiosynthese ($-\log_{10}(p) = 6.88$, $PI = 1.8$) vermutet werden konnte, ließ sich diese bei Bullen nicht beobachten. In hocheffizienten Bullen konnte dagegen eine Hochregulierung der NRF2-geleiteten oxidativen Stressantwort detektiert werden ($-\log_{10}(p) = 2.83$, z-score = 0.447).

Die zweite Studie lieferte zudem Hinweise darauf, dass lncRNAs bei Bullen unter Umständen in die Glukoneogenese involviert sind (durch Korrelation mit *G6PC*, *PCK1*, *FBP1* und Glycerol), auch wenn keine signifikante Anreicherung für diesen Pathway festgestellt werden konnte. Bei den Kühen (Studie 3) wiesen die signifkant höheren Expressionswerte von *PC* und *PCK1* in SEC Kühen ebenfalls darauf hin, dass die Glukoneogenese in Kühen des Sektretionstyps besonders aktiv war. Während für Kühe divergenter Stoffwechseltypen keine vergleichbaren Studien zu lncRNAs vorliegen, indizieren Ergebnisse zur differenziellen Expression von lncRNAs in der Milchdrüse in verschiedenen Laktationsstadien, dass Lipid- und Glucosestoffwechsel zu den angesprochenen Schaltkreisen zählen (Zheng et al., 2018). Auch bei Mäusen konnten in der Leber bereits lncRNAs mit Schlüsselfunktion im Lipid- und Glucosestoffwechsel identifiziert werden (Pradas-Juni et al., 2020; Yang et al., 2016).

Insgesamt ließ sich im Vergleich von Studie 2 und 3 konstatieren, dass nur wenige DE Gene, die bei Kühen (FDR ≤ 0.05) für die Enrichmentanalyse genutzt wurden, auch bei Bullen signifkant (FDR ≤ 0.1) DE waren und einer Analyse zugeführt wurden. Der direkte Vergleich wird jedoch durch die unterschiedlich angesetzten Signifikanzniveaus erschwert. Zu den Genen, die in beiden Geschlechtern in der Leber DE waren, zählen beispielsweise *P2RY1*, *MAT2A*, *SLC45A3*, *ADIPOR2*, *ANGPTL4* und *GK*, die insbesondere in Studie 3 ausführlicher beleuchtet wurden. Die lncRNA-Transkripte, die in beiden Geschlechtern DE sind, sind v.a. durch ihre Eigenschaft als NATs zu anderen Loci gekennzeichnet, die durch FEELnc als positionelles Partnergen vorhergesagt wurden. Jedoch stand keines dieser lncRNA-Gen-Paare im Fokus der zweiten oder dritten Studie.

Zusammenfassend lässt sich festhalten, dass im Rahmen dieser Arbeit der Katalog von im Rind exprimierten lncRNAs in großem Maß erweitert werden konnte. Die Integration von

[10] PI = Pathway Impact. Der Pathway Impact ist die Summe des Wichtigkeitsmaßes der Metaboliten, die aus einem Datenset dem Pathway zugeordnet werden konnten, und die durch die Summe der Wichtigkeitsmaße aller Metaboliten eines Pathways normalisiert wurde (J. Xia et al., 2011)

Phänotyp-, Metabolom- und Transkriptomdaten hat sich als geeignet und mehrwertbringend für die funktionale Charakterisierung von neuen, unbekannten lncRNAs erwiesen. Es konnte festgestellt werden, dass lncRNAs eine mögliche, regulatorische Funktion in einer Vielzahl von biologischen Prozessen im Stoffwechsel von Bullen und Kühen mit unterschiedlicher Nährstoffverteilung erfüllen. Sowohl in Geschlechtern, Stoffwechselgruppen als auch Geweben lassen die Ergebnisse eine Feinabstimmung der Genregulation duch lncRNAs vermuten. Insbesondere lncRNAs, die NATs zu proteinkodierenden Genen sind, wiesen in der Regel eine starke, positive Koexpression mit selbigen auf, was eine stabilisierende Wirkung der lncRNA auf die jeweilige mRNA des Wirtgens vermuten lässt. Für weiterführende Studien bietet sich eine eingehendere Untersuchung der lncRNAs auf Sequenz- und Spleißvarianten und deren Einfluss auf die Merkmalsmodulierung an. Für nichtkodierende Varianten wurde bereits ein Beitrag zur Ausprägung diverser Merkmale beim Rind festgestellt (Xiang et al., 2019), wodurch sich solche Erkenntnisse zur Nutzung für die züchterische Bearbeitung, z.B. in Form der genomischen Selektion eignen. Vor einer möglichen Implementierung in genomische, züchterische Anwendungen müssen die Erkenntnise jedoch an weiteren Tieren aus reingezüchteten Populationen validiert werden, da die Erkenntnisse der vorliegenden Arbeit in einer Kreuzungspopulation gewonnen wurden. Die erwähnten Sequenzübereinstimmungen ausgewählter lncRNAs mit *Bos indicus* lassen jedoch ein ausreichendes Konservierungsmaß in anderen taurinen Rassen erwarten.

7.5. Schlussfolgerungen und Ausblick

Die vorgelegten drei Studien haben zur Erweiterung des bovinen lncRNA-Kataloges beigetragen und geben erste Aufschlüsse über die Rolle von lncRNAs im Stoffwechsel von Rindern beider Geschlechter. Mit den hier präsentierten Ergebnissen liegt eine Auswahl von Kandidaten-lncRNAs vor, die in weiterführenden Analysen tiefergehend hinsichtlich ihres Bindungsverhaltens, der Regulation von Genexpression und der Modulation von Prozessen, die mit der divergenten Ausprägung des Merkmalskomplexes der Nährstoffverteilung bzw. des jeweiligen Stoffwechseltyps assoziiert sind, untersucht werden können.

Durch eine umfassende Analyse der Transkriptomdaten wurden lncRNAs mit gewebsspezifischer und differenzieller Expression zwischen Stoffwechselgruppen bzw. mit hohem genregulatorischen Potenzial in Jejunum, Leber, Pansen und Skelettmuskel von Bullen und Kühen identifiziert. Die Integration von Metabolom- und Transkriptomdaten mit Phänotypen deutet auf eine Beteiligung von lncRNAs in zahlreichen biologischen Schaltkreisen hin, wie den Aminosäure-, Energie- und Fettstoffwechsel sowie die Proteinsynthese, die Argininbiosynthese, den Zitrat- und Harnstoffzyklus, die Glukoneogenese und den PPAR-Signalweg. Insbesondere die Einarbeitung von Metabolomdaten erwies sich als wertvolle Ergänzung zu proteinkodierenden Genen für eine funktionale Charakterisierung unbekannter lncRNAs. Die Ergebnisse zeigen, dass lncRNAs mit zahlreichen biologischen Prozessen, die die Nährstoffverteilung beim Rind auf molekularer Ebene beeinflussen, verknüpft sind. Daraus lässt sich schlussfolgern, dass lncRNAs ein bisher unzulänglich beachtetes Glied bei der Stoffwechselregulation sind, deren weitere Erforschung zum besseren Verständnis des Stoffum- und -ansatzes beim Rind beitragen wird.

In weiterführenden Studien sollten hier lncRNAs, die in dieser Arbeit als regulatorische Schlüsselelemente herauskristallisiert wurden, einer funktionalen Validierung unterzogen werden. Sowohl eine ChIRP-Analyse zur Aufdeckung von interaktiven Bindungsstellen der lncRNAs im Genom oder Proteom als auch Knock-Down Studien in Zellkultur durch Antisense-Oligonukleotide bieten beispielsweise die Möglichkeit, die Koregulation von lncRNAs mit kodierenden Genen und deren Mechanismen eingehender zu untersuchen. Zur Komplettierung des bovinen lncRNA-Kataloges, sind ergänzende Transkriptomstudien in weiteren Geweben notwendig. Langfristig ist eine Integration von validierten lncRNA-Varianten, die die Nährstoffverteilung divergenter Stoffwechseltypen modulieren, in die genomische Selektion beim Milch- und Fleischrind denkbar, um so die Resseourceneffizienz und tierische Produktion zu optimieren.

8. Literaturverzeichnis

Ajmone-Marsan, P., Garcia, J. F., & Lenstra, J. A. (2010). On the origin of cattle: How aurochs became cattle and colonized the world. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*(4), 148–157. https://doi.org/10.1002/evan.20267

Alexander, L. J., MacNeil, M. D., Geary, T. W., Snelling, W. M., Rule, D. C., & Scanga, J. A. (2007). Quantitative trait loci with additive effects on palatability and fatty acid composition of meat in a Wagyu-Limousin F2 population. *Animal Genetics*, *38*(5), 506–513. https://doi.org/10.1111/j.1365-2052.2007.01643.x

Alexandre, P. A., Reverter, A., Berezin, R. B., Porto-Neto, L. R., Ribeiro, G., Santana, M. H. A., Ferraz, J. B. S., & Fukumasu, H. (2020). Exploring the regulatory potential of long non-coding rna in feed efficiency of indicine cattle. *Genes*, *11*(9), 1–22. https://doi.org/10.3390/genes11090997

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., & Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, *160*(4), 595–606. https://doi.org/10.1016/j.cell.2015.01.009

Antonov, I. V., Mazurov, E., Borodovsky, M., & Medvedeva, Y. A. (2019). Prediction of lncRNAs and their interactions with nucleic acids: Benchmarking bioinformatics tools. *Briefings in Bioinformatics*, *20*(2), 551–564. https://doi.org/10.1093/bib/bby032

Arthur, J. P. F., & Herd, R. M. (2008). Residual feed intake in beef cattle. *Revista Brasileira de Zootecnia*, *37*, 269–279. https://doi.org/10.1590/S1516-35982008001300031

Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., Goyache, F., Royo, L. J., Conti, S., Lari, M., Martini, A., Ouragh, L., Magid, A., Atash, A., Zsolnai, A., Boscato, P., Triantaphylidis, C., Ploumi, K., Sineo, L., … Bertorelle, G. (2006). The origin of European cattle: Evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8113–8118. https://doi.org/10.1073/pnas.0509210103

Bhan, A., Soleimani, M., & Mandal, S. S. (2017). Long noncoding RNA and cancer: A new paradigm. *Cancer Research*, *77*(15), 3965–3981. https://doi.org/10.1158/0008-5472.CAN-16-2634

Blake, R. W., & Custodio, A. A. (1984). Feed Efficiency: A Composite Trait of Dairy Cattle. *Journal of Dairy Science*, *67*(9), 2075–2083. https://doi.org/10.3168/jds.S0022-0302(84)81548-9

BMEL (Bundesministerium für Ernährung und Landwirtschaft). (2020). *Rinderbestände*. [Accessed 2020-11-10]. https://www.bmel-statistik.de/landwirtschaft/tierhaltung/rinderhaltung/%0ASJT-3100920-2019.xlsx%0A

Bolha, L., Ravnik-Glavač, M., & Glavač, D. (2017). Long Noncoding RNAs as Biomarkers in Cancer. *Disease Markers*, *2017*, 7243968. https://doi.org/10.1155/2017/7243968

Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H. F., Avner, P., & Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature*, *351*(6324), 325–329. https://doi.org/10.1038/351325a0

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., Sahana, G., Govignon-Gion, A., Boitard, S., Dolezal, M., Pausch, H., Brøndum, R. F., Bowman, P. J., Thomsen, B., Guldbrandtsen, B., Lund, M. S., Servin, B., Garrick, D. J., Reecy, J., … Hayes, B. J. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics*, *50*(3), 362–367. https://doi.org/10.1038/s41588-018-0056-5

Brannan, C. I., Dees, E. C., Ingram, R. S., & Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Molecular and Cellular Biology*, *10*(1), 28–36. https://doi.org/10.1128/mcb.10.1.28

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., & Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, *351*(6324), 329–331. https://doi.org/10.1038/351329a0

Brown, T. (2002). Genomes 2, Chapter 7 Understanding a Genome Sequence. In *NCBI Bookshelf* (2nd ed.). Wiley-Liss. https://www.ncbi.nlm.nih.gov/books/NBK21136/

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development*, *25*(18), 1915–1927. https://doi.org/10.1101/gad.17446611

Canovas, A., Reverter, A., DeAtley, K. L., Ashley, R. L., Colgrave, M. L., Fortes, M. R. S., Islas-Trejo, A., Lehnert, S., Porto-Neto, L., Rincon, G., Silver, G. A., Snelling, W. M., Medrano, J. F., & Thomas, M. G. (2014). Multi-Tissue Omics Analyses Reveal Molecular Regulatory Networks for Puberty in Composite Beef Cattle. *Plos One*, *9*(7), 17. https://doi.org/10.1371/journal.pone.0102551

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Cheng, J. T., Wang, L., Wang, H., Tang, F. R., Cai, W. Q., Sethi, G., Xin, H. W., & Ma, Z. (2019). Insights into Biological Role of LncRNAs in Epithelial-Mesenchymal Transition. *Cells,* 8(10). https://doi.org/10.3390/cells8101178

Chong, J., Wishart, D. S., & Xia, J. (2019). Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics*, *68*(1), e86. https://doi.org/10.1002/cpbi.86

Chu, C., Quinn, J., & Chang, H. Y. (2012). Chromatin Isolation by RNA Purification (ChIRP). *Journal of Visualized Experiments : JoVE*, *61*, 3912. https://doi.org/10.3791/3912

Darbellay, F., & Necsulea, A. (2020). Comparative Transcriptomics Analyses across Species, Organs, and Developmental Stages Reveal Functionally Constrained lncRNAs. *Molecular Biology and Evolution*, *37*(1), 240–259. https://doi.org/10.1093/molbev/msz212

Darnell, J. E., Lodish, H., Berk, A., Zipursky, L., Matsudaira, P., & Baltimore, D. (2000). *Molecular Cell Biology* (4th ed.).

de Sá, P. H. C. G., Guimarães, L. C., das Graças, D. A., de Oliveira Veras, A. A., Barh, D., Azevedo, V., da Costa da Silva, A. L., & Ramos, R. T. J. (2018). Next-generation sequencing and data analysis: Strategies, tools, pipelines and protocols. In D. Barh & V. B. T.-O. T. and B.-E. Azevedo (Eds.), *Omics Technologies and Bio-engineering: Towards Improving Quality of Life* (Vol. 1, pp. 191–207). Academic Press. https://doi.org/10.1016/B978-0-12-804659-3.00011-7

Demasius, W., Weikard, R., Hadlich, F., Muller, K. E., & Kuhn, C. (2013). Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle. *Veterinary Research*, 44*(*93). https://doi.org/10.1186/1297-9716-44-93

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigo, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/gr.132159.111

Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. *Trends in Genetics*, *30*(4), 121–123. https://doi.org/10.1016/j.tig.2014.01.004

Doublet, A. C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., Laloë, D., & Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetics Selection Evolution*, *51*(1), 52. https://doi.org/10.1186/s12711-019-0495-1

Eberlein, A., Kalbe, C., Goldammer, T., Brunner, R. M., Kuehn, C., & Weikard, R. (2010). Analysis of structure and gene expression of bovine CCDC3 gene indicates a function in fat metabolism. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, *156*(1), 19–

25. https://doi.org/10.1016/j.cbpb.2010.01.013

Egger-Danner, C., Cole, J. B., Pryce, J. E., Gengler, N., Heringstad, B., Bradley, A., & Stock, K. F. (2014). Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal*, *9*(2), 191–207. https://doi.org/10.1017/S1751731114002614

Encyclopædia Britannica. (2018). *Livestock: Cattle*. [Accessed 2020-11-26]. https://www.britannica.com/animal/cattle-livestock

Falker-Gieske, C., Blaj, I., Preuß, S., Bennewitz, J., Thaller, G., & Tetens, J. (2019). GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs. *G3: Genes|Genomes|Genetics*, *9*(9), 2823 LP – 2834. https://doi.org/10.1534/g3.119.400452

Fang, Z. H., & Pausch, H. (2019). Multi-trait meta-analyses reveal 25 quantitative trait loci for economically important traits in Brown Swiss cattle. *BMC Genomics*, *20*(1), 695. https://doi.org/10.1186/s12864-019-6066-6

Felius, M., Beerling, M. L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., & Lenstra, J. A. (2014). On the history of cattle genetic resources. *Diversity,* 6(4), 705–750. https://doi.org/10.3390/d6040705

Ferguson, L. R. (2010). Meat and cancer. *Meat Science*, *84*(2), 308–313. https://doi.org/10.1016/j.meatsci.2009.06.032

Fickett, J. W. (1982). Recognition of protein coding regions in DNA sequences. *Nucleic Acids Research*, *10*(17), 5303–5318. https://doi.org/10.1093/nar/10.17.5303

Friedrich, J., Brand, B., Ponsuksili, S., Graunke, K. L., Langbein, J., Knaust, J., Kühn, C., & Schwerin, M. (2016). Detection of genetic variants affecting cattle behaviour and their impact on milk production: A genome-wide association study. *Animal Genetics*, *47*(1), 12–18. https://doi.org/10.1111/age.12371

Grigoriadis, A., Oliver, G. R., Tanney, A., Kendrick, H., Smalley, M. J., Jat, P., & Neville, A. M. (2009). Identification of differentially expressed sense and antisense transcript pairs in breast epithelial tissues. *BMC Genomics*, *10*, 324. https://doi.org/10.1186/1471-2164-10-324

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M., & Snell, R. (2002). Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, *12*(2), 222–231. https://doi.org/10.1101/gr.224202

Grobet, L., Martin, L. J. R., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Ḿnissier, F., Massabanda, J., Fries, R., Hanset, R., & Georges, M. (1997). A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genetics*, *17*(1), 71–74. https://doi.org/10.1038/ng0997-71

Gudenas, B. L., Wang, J., Kuang, S. Z., Wei, A. Q., Cogill, S. B., & Wang, L. J. (2019). Genomic data mining for functional annotation of human long noncoding RNAs. *Journal of Zhejiang University: Science B*, *20*(6), 476–487. https://doi.org/10.1631/jzus.B1900162

Hammon, H. M., Bellmann, O., Voigt, J., Schneider, F., & Kühn, C. (2007). Glucose-dependent insulin response and milk production in heifers within a segregating resource family population. *Journal of Dairy Science*, *90*(7), 3247–3254. https://doi.org/10.3168/jds.2006-748

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*, *11*(7), 1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023

Hezroni, H., Ben-Tov Perry, R., Meir, Z., Housman, G., Lubelsky, Y., Ulitsky, I. (2017). A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Biology*, *18*(1), 1-162. https://doi.org/10.1186/s13059-017-1293-0

Hocquette, J. F., Gondret, F., Baéza, E., Médale, F., Jurie, C., & Pethick, D. W. (2010). Intramuscular fat content in meat-producing animals: development, genetic and nutritional control, and identification of putative markers. *Animal*, *4*. https://doi.org/10.1017/s1751731109991091

Hu, G., Niu, F., Humburg, B. A., Liao, K., Bendi, S., Callen, S., Fox, H. S., & Buch, S. (2018). Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis. *Oncotarget*, *9*(26), 18648–18663. https://doi.org/10.18632/oncotarget.24307

Ibanez-Escriche, N., & Simianer, H. (2016). Animal breeding in the genomics era. *Animal Frontiers*, *6*(1), 4–5. https://doi.org/10.2527/af.2016-0001

Ilves, A., Harzia, H., Ling, K., Ots, M., Soomets, U., & Kilk, K. (2012). Alterations in milk and blood metabolomes during the first months of lactation in dairy cows. *Journal of Dairy Science*, *95*(10), 5788–5797. https://doi.org/10.3168/jds.2012-5617

Jeroch, H., Drochner, W., & Ortwin, S. (2008). *Ernährung landwirtschaftlicher Nutztiere: Ernährungsphysiologie, Futtermittelkunde, Fütterung* (2nd ed.). Verlag Eugen Ulmer, Stuttgart.

Johnston, A. D., Simões-Pires, C. A., Thompson, T. V., Suzuki, M., & Greally, J. M. (2019). Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding. *Nature Communications*, *10*(1), 3472. https://doi.org/10.1038/s41467-019-11412-5

Kashi, K., Henderson, L., Bonetti, A., & Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1859*(1), 3–15. https://doi.org/10.1016/j.bbagrm.2015.10.010

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., … Wahlestedt, C. (2005). Molecular biology: Antisense transcription in the mammalian transcriptome. *Science*, *309*(5740), 1564–1566. https://doi.org/10.1126/science.1112009

Kirchgeßner, M. (1997). *Tierernährung: Vol. 10th edition*. Verlags Union Agrar DLG-Verlag, Frankfurt.

Knaust, J., Hadlich, F., Weikard, R., & Kuehn, C. (2016). Epistatic interactions between at least three loci determine the “rat-tail” phenotype in cattle. *Genetics Selection Evolution*, *48*(1), 12. https://doi.org/10.1186/s12711-016-0199-8

Koch, R. M., Swiger, L. A., Chambers, D., & Gregory, K. E. (1963). Efficiency of Feed Use in Beef Cattle. *Journal of Animal Science*, *22*(2), 486–494. https://doi.org/10.2527/jas1963.222486x

Kogelman, L. J. A., Kadarmideen, H. N., Mark, T., Karlskov-Mortensen, P., Bruun, C. S., Cirera, S., Jacobsen, M. J., Jørgensen, C. B., & Fredholm, M. (2013). An F2 pig resource population as a model for genetic studies of obesity and obesity-related diseases in humans: Design and genetic parameters. *Frontiers in Genetics,* 4(29). https://doi.org/10.3389/fgene.2013.00029

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*, *35*, W345-9. https://doi.org/10.1093/nar/gkm391

Kosinska-Selbi, B., Mielczarek, M., & Szyda, J. (2020). Review: Long non-coding RNA in livestock. *Animal*, *14*(10), 1–11. https://doi.org/10.1017/S1751731120000841

Kramer, A., Green, J., Pollard Jr., J., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, *30*(4), 523–530. https://doi.org/10.1093/bioinformatics/btt703

Krappmann, K., Weikard, R., Gerst, S., Wolf, C., & Kühn, C. (2011). A genetic predisposition for bovine neonatal pancytopenia is not due to mutations in coagulation factor XI. *Veterinary Journal*, *190*(2), 225–229. https://doi.org/10.1016/j.tvjl.2010.10.007

Krappmann, K., Widmann, P., Weikard, R., & Kühn, C. (2012). Variants of the bovine retinoic acid receptor-related orphan receptor C gene are in linkage disequilibrium with QTL for milk production traits on chromosome 3 in a beef × dairy crossbreed population. *Archives Animal Breeding*, *55*(4), 346–355. https://doi.org/10.5194/aab-55-346-2012

Kühn, C., Bellmann, O., Voigt, J., Wegner, J., Guiard, V., & Ender, K. (2002). An experimental approach for studying the genetic and physiological background of nutrient transformation in cattle with respect to nutrient secretion and accretion type. *Archives Animal Breeding*, *45*(4), 317–330. https://doi.org/10.5194/aab-45-317-2002

Kunz, M., Wolf, B., Fuchs, M., Christoph, J., Xiao, K., Thum, T., Atlan, D., Prokosch, H.-U., & Dandekar, T. (2020). A comprehensive method protocol for annotation and integrated functional understanding of lncRNAs. *Briefings in Bioinformatics*, *21*(4), 1391–1396. https://doi.org/10.1093/bib/bbz066

LaRossa, R. A. (2013). Transcriptome. In S. Maloy & K. B. T.-B. E. of G. (Second E. Hughes (Eds.),

Brenner's Encyclopedia of Genetics: Second Edition (pp. 101–103). Academic Press. https://doi.org/10.1016/B978-0-12-374984-0.01553-9

Le Beguec, C., Wucher, V., Lagoutte, L., Cadieul, E., Botherel, N., Hedan, B., De Brito, C., Guillory, A. S., Andre, C., Derrien, T., & Hitte, C. (2018). Characterisation and functional predictions of canine long non-coding RNAs. *Scientific Reports*, *8*, 12. https://doi.org/10.1038/s41598-018-31770-2

Leibniz-Gemeinschaft. (2008). *Stellungnahme zum Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere (FBN) Dummerstorf: Vol. SEN 0077/0*.[accessed 2020-11-17] https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/ARCHIV_downloads/Archiv/Evaluierung/Senatsstellungnahmen/Senatsstellungnahme-FBN-2008.pdf

Lenstra, J. A. (2014). Cattle: Domestication. In C. Smith (Ed.), *Encyclopedia of Global Archaeology* (pp. 1186–1188). Springer New York. https://doi.org/10.1007/978-1-4419-0465-2_2201

Lewin, B. (2000). *Genes VII*. Oxford University Press, Oxford.

Li, A., Zhang, J., & Zhou, Z. (2014). PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*, *15*(1), 311. https://doi.org/10.1186/1471-2105-15-311

Li, Q., Li, Z., Feng, C., Jiang, S., Zhang, Z., & Ma, L. (2020). Multi-omics annotation of human long non-coding RNAs. *Biochemical Society Transactions*, *48*(4), 1545–1556. https://doi.org/10.1042/BST20191063

Li, Y., Xu, C., Xia, C., Zhang, H., Sun, L., & Gao, Y. (2014). Plasma metabolic profiling of dairy cows affected with clinical ketosis using LC/MS technology. *Veterinary Quarterly*, *34*(3), 152–158. https://doi.org/10.1080/01652176.2014.962116

Liao, Q., Liu, C., Yuan, X., Kang, S., Miao, R., Xiao, H., Zhao, G., Luo, H., Bu, D., Zhao, H., Skogerbø, G., Wu, Z., & Zhao, Y. (2011). Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Research*, *39*(9), 3864–3878. https://doi.org/10.1093/nar/gkq1348

Littlejohn, M., Grala, T., Sanders, K., Walker, C., Waghorn, G., MacDonald, K., Coppieters, W., Georges, M., Spelman, R., Hillerton, E., Davis, S., & Snell, R. (2012). Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. *Animal Genetics*, *43*(5), 591–594. https://doi.org/10.1111/j.1365-2052.2011.02293.x

Liu, Y., Albrecht, E., Schering, L., Kuehn, C., Yang, R., Zhao, Z., & Maak, S. (2018). Agouti signaling protein and its receptors as potential molecular markers for intramuscular and body fat deposition in cattle. *Frontiers in Physiology*, *9*(172). https://doi.org/10.3389/fphys.2018.00172

Long, Y., Wang, X., Youmans, D. T., & Cech, T. R. (2017). How do lncRNAs regulate transcription? *Science Advances*, *3*(9), eaao2110–eaao2110. https://doi.org/10.1126/sciadv.aao2110

Luo, Z. Z., Shen, L. H., Jiang, J., Huang, Y. X., Bai, L. P., Yu, S. M., Yao, X. P., Ren, Z. H., Yang, Y. X., & Cao, S. Z. (2019). Plasma metabolite changes in dairy cows during parturition identified using untargeted metabolomics. *Journal of Dairy Science*, *102*(5), 4639–4650. https://doi.org/10.3168/jds.2018-15601

Lutz, V., Stratz, P., Preuß, S., Tetens, J., Grashorn, M. A., Bessei, W., & Bennewitz, J. (2017). A genome-wide association study in a large F2-cross of laying hens reveals novel genomic regions associated with feather pecking and aggressive pecking behavior. *Genetics Selection Evolution*, *49*(1), 18. https://doi.org/10.1186/s12711-017-0287-4

Magistri, M., Faghihi, M. A., St Laurent, G., & Wahlestedt, C. (2012). Regulation of chromatin structure by long noncoding RNAs: Focus on natural antisense transcripts. *Trends in Genetics*, *28*(8), 389–396. https://doi.org/10.1016/j.tig.2012.03.013

Marchese, F. P., Raimondi, I., & Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biology*, *18*(206). https://doi.org/10.1186/s13059-017-1348-2

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, *12*(10), 671–682. https://doi.org/10.1038/nrg3068

Mattick, J. S. (2001). Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Reports*, *2*(11), 986–991. https://doi.org/10.1093/embo-reports/kve230

Melzer, N., Trißl, S., & Nürnberg, G. (2017). Short communication: Estimating lactation curves for highly inhomogeneous milk yield data of an F2 population (Charolais × German Holstein). *Journal of Dairy Science*, *100*(11), 9136–9142. https://doi.org/10.3168/jds.2017-12772

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*, *10*(3), 155–159. https://doi.org/10.1038/nrg2521

Miao, Y., Xu, S.-Y., Chen, L.-S., Liang, G.-Y., Pu, Y.-P., & Yin, L.-H. (2017). Trends of long noncoding RNA research from 2007 to 2016: a bibliometric analysis. *Oncotarget*, *8*(47), 83114–83127. https://doi.org/10.18632/oncotarget.20851

Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., & Baes, C. F. (2017). A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *Journal of Dairy Science*, *100*(12), 10251–10271. https://doi.org/10.3168/jds.2017-12968

Mooney, M. A., & Wilmot, B. (2015). Gene set analysis: A step-by-step guide. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *168*(7), 517–527. https://doi.org/10.1002/ajmg.b.32328

Morlan, J. D., Qu, K., & Sinicropi, D. V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS ONE*, *7*(8), e42882. https://doi.org/10.1371/journal.pone.0042882

Morlando, M., Ballarino, M., & Fatica, A. (2015). Long non-coding RNAs: New players in hematopoiesis and leukemia. *Frontiers in Medicine,* 2(23). https://doi.org/10.3389/fmed.2015.00023

Morrow, D. A. (1976). Fat Cow Syndrome. *Journal of Dairy Science*, *59*(9), 1625–1629. https://doi.org/10.3168/jds.S0022-0302(76)84415-3

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. https://doi.org/10.1038/nmeth.1226

Murakami, Y., Kubo, S., Tamori, A., Itami, S., Kawamura, E., Iwaisako, K., Ikeda, K., Kawada, N., Ochiya, T., & Taguchi, Y. H. (2015). Comprehensive analysis of transcriptome and metabolome analysis in Intrahepatic Cholangiocarcinoma and Hepatocellular Carcinoma. *Scientific Reports*, *5*(1), 16294. https://doi.org/10.1038/srep16294

Muret, K., Klopp, C., Wucher, V., Esquerré, D., Legeai, F., Lecerf, F., Désert, C., Boutin, M., Jehl, F., Acloque, H., Giuffra, E., Djebali, S., Foissac, S., Derrien, T., & Lagarrigue, S. (2017). Long noncoding RNA repertoire in chicken liver and adipose tissue. *Genetics, Selection, Evolution : GSE*, *49*, 6. https://doi.org/10.1186/s12711-016-0275-0

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, *320*(5881), 1344–1349. https://doi.org/10.1126/science.1158441

National Human Genome Research Institute (NHGRI). (2020). *Transcriptome Fact Sheet*. [accessed 2020-11-17]. https://www.genome.gov/about-genomics/fact-sheets/Transcriptome-Fact-Sheet%0Ahttps://www.genome.gov/13014330/transcriptome-fact-sheet/

Nguyen, L. T., Reverter, A., Cánovas, A., Venus, B., Anderson, S. T., Islas-Trejo, A., Dias, M. M., Crawford, N. F., Lehnert, S. A., Medrano, J. F., Thomas, M. G., Moore, S. S., & Fortes, M. R. S. (2018). STAT6, PBX2, and PBRM1 Emerge as Predicted Regulators of 452 Differentially Expressed Genes Associated With Puberty in Brahman Heifers. *Frontiers in Genetics*, *9*(87). https://doi.org/10.3389/fgene.2018.00087

Nguyen, T. M., Shafi, A., Nguyen, T., & Draghici, S. (2019). Identifying significantly impacted pathways: A comprehensive review and assessment. *Genome Biology*, *20*(1), 203. https://doi.org/10.1186/s13059-019-1790-4

Nolte, W., Weikard, R., Albrecht, E., Hammon, H., & Kühn, C. (unveröffentlicht). Metabogenomic analysis to functionally annotate the regulatory role of long non-coding RNAs in the liver of cows with different nutrient partitioning phenotype.

Nolte, W., Weikard, R., Brunner, R. M., Albrecht, E., Hammon, H. M., Reverter, A., & Küehn, C. (2020). Identification and annotation of potential function of regulatory antisense long non-coding RNAs related to feed efficiency in bos taurus bulls. *International Journal of Molecular Sciences*,

21(9). https://doi.org/10.3390/ijms21093292

Nolte, W., Weikard, R., Brunner, R. M., Albrecht, E., Hammon, H. M., Reverter, A., & Kühn, C. (2019). Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.01130

NONCODE. (2020). *An integrated knowledge database dedicated to ncRNAs, especially lncRNAs.* [Accessed 2020-11-10]. http://www.noncode.org/analysis.php

Novais, F. J., Pires, P. R. L., Alexandre, P. A., Dromms, R. A., Iglesias, A. H., Ferraz, J. B. S., Styczynski, M. P. W., & Fukumasu, H. (2019). Identification of a metabolomic signature associated with feed efficiency in beef cattle. *BMC Genomics*, *20*(1), 8. https://doi.org/10.1186/s12864-018-5406-2

Owens, M. C., Clark, S. C., Yankey, A., & Somarowthu, S. (2019). Identifying structural domains and conserved regions in the long non-coding RNA lncTCF7. *International Journal of Molecular Sciences,* 20(19). https://doi.org/10.3390/ijms20194770

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., & Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*, *37*(18), e123–e123. https://doi.org/10.1093/nar/gkp596

Pausch, H., Emmerling, R., Schwarzenbacher, H., & Fries, R. (2016). A multi-trait meta-analysis with imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh cattle. *Genetics Selection Evolution*, *48*(1), 14. https://doi.org/10.1186/s12711-016-0190-4

Pelechano, V., & Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics*, *14*(12), 880–893. https://doi.org/10.1038/nrg3594

Pennacchio, L. A., & Rubin, E. M. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nature Reviews Genetics*, *2*(2), 100–109. https://doi.org/10.1038/35052548

Perez-Montarelo, D., Hudson, N. J., Fernandez, A. I., Ramayo-Caldas, Y., Dalrymple, B. P., & Reverter, A. (2012). Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. *PLoS One*, *7*(9), e46159. https://doi.org/10.1371/journal.pone.0046159

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667. https://doi.org/10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Pradas-Juni, M., Hansmeier, N. R., Link, J. C., Schmidt, E., Larsen, B. D., Klemm, P., Meola, N., Topel, H., Loureiro, R., Dhaouadi, I., Kiefer, C. A., Schwarzer, R., Khani, S., Oliverio, M., Awazawa, M., Frommolt, P., Heeren, J., Scheja, L., Heine, M., … Kornfeld, J. W. (2020). A MAFG-lncRNA axis links systemic nutrient abundance to hepatic glucose metabolism. *Nature Communications*, *11*(1), 644. https://doi.org/10.1038/s41467-020-14323-y

QIAGEN. (2016). *IPA: Maximizing the Biological Interpretation of Gene, Transcript & Protein Expression Data with IPA* (Issue 203, p. [accessed 2020-11-10]). https://chhe.research.ncsu.edu/wordpress/wp-content/uploads/2015/10/IPA-Data-Analysis-training-slides-2016_04.pdf

Ramakrishnaiah, Y., Kuhlmann, L., & Tyagi, S. (2020). Towards a comprehensive pipeline to identify and functionally annotate long noncoding RNA (lncRNA). *Computers in Biology and Medicine*, *127*, 104028. https://doi.org/10.1016/j.compbiomed.2020.104028

Ramilowski, J. A., Yip, C. W., Agrawal, S., Chang, J. C., Ciani, Y., Kulakovskiy, I. V., Mendez, M., Ooi, J. L. C., Ouyang, J. F., Parkinson, N., Petri, A., Roos, L., Severin, J., Yasuzawa, K., Abugessaisa, I., Akalin, A., Antonov, I. V., Arner, E., Bonetti, A., … Carninci, P. (2020). Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Research*, *30*(7), 1060–1072. https://doi.org/10.1101/gr.254219.119

Reverter, A., & Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics*, *24*(21), 2491–2497. https://doi.org/10.1093/bioinformatics/btn482

Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M., & Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics*, *26*(7), 896–904. https://doi.org/10.1093/bioinformatics/btq051

Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. In R. D. Kornberg (Ed.), *Annual Review of Biochemistry,* 81, 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516–1517. https://doi.org/10.1126/science.273.5281.1516

Ritchie, H. (2017). *Meat and Dairy Production - Our World in Data*. Our World in Data. [accessed 2020-11-16] https://ourworldindata.org/meat-production

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., & Alba, M. M. (2014). Long non-coding RNAs as a source of new peptides. *ELife*, *3*, e03523. https://doi.org/10.7554/eLife.03523

Salehi, S., Taheri, M. N., Azarpira, N., Zare, A., & Behzad-Behbahani, A. (2017). State of the art technologies to explore long non-coding RNAs in cancer. *Journal of Cellular and Molecular Medicine*, *21*(12), 3120–3140. https://doi.org/10.1111/jcmm.13238

Schuierer, S., Carbone, W., Knehr, J., Petitjean, V., Fernandez, A., Sultan, M., & Roma, G. (2017). A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC Genomics*, *18*(1), 442. https://doi.org/10.1186/s12864-017-3827-y

Setoguchi, K., Watanabe, T., Weikard, R., Albrecht, E., Kühn, C., Kinoshita, A., Sugimoto, Y., & Takasuga, A. (2011). 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*, *42*(6), 650–655. https://doi.org/10.1111/j.1365-2052.2011.02196.x

Sevane, N., Nute, G., Sañudo, C., Cortes, O., Cañon, J., Williams, J. L., Dunner, S., Checa, M. L., Christensen, M., Crisá, A., Delourme, D., Failla, S., García, D., Gigli, S., Hocquette, J. F., Levéziel, H., Mangin, B., Marchitelli, C., Miranda, D., … Valentini, A. (2014). Muscle lipid composition in bulls from 15 european breeds. *Livestock Science*, *160*(1), 1–11. https://doi.org/10.1016/j.livsci.2013.11.001

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sparber, P., Filatova, A., Khantemirova, M., & Skoblov, M. (2019). The role of long non-coding RNAs in the pathogenesis of hereditary diseases. *BMC Medical Genomics*, *12*(2), 42. https://doi.org/10.1186/s12920-019-0487-6

Statista GmbH. (2015). *Konsum von Milch und Milchprodukten in Deutschland - Statista-Dossier*. [accessed 2020-11-17] https://de.statista.com/statistik/studie/id/11867/dokument/milch-statista-dossier/

Statista GmbH. (2020). *Konsum von Fleisch in Deutschland*. [accessed 2020-11-17] https://de.statista.com/statistik/studie/id/29857/dokument/konsum-von-fleisch-in-deutschland-statista-dossier/

Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H., & Yaspo, M. L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics*, *15*(1), 675. https://doi.org/10.1186/1471-2164-15-675

Swan, A. A., & Kinghorn, B. P. (1992). Evaluation and Exploitation of Crossbreeding in Dairy Cattle. *Journal of Dairy Science*, *75*(2), 624–639. https://doi.org/10.3168/jds.S0022-0302(92)77800-X

The Broad Institute. (2009). *Gene Set Enrichment Analysis GSEA User Guide*. [Accessed 2020-11-10]. https://www.gsea-msigdb.org/gsea/doc/GSEAUserGuideTEXT.htm#_Gene_Sets_and_GSEA

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., & Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, *147*(7), 1537–1550. https://doi.org/10.1016/j.cell.2011.11.055

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings in*

Bioinformatics, *19*(4), 575–592. https://doi.org/10.1093/bib/bbw139

Van Eijndhoven, M. H. T., Hiemstra, S. J., & Calus, M. P. L. (2011). Short communication: Milk fat composition of 4 cattle breeds in the Netherlands. *Journal of Dairy Science*, *94*(2), 1021–1025. https://doi.org/10.3168/jds.2009-3018

Verbeke, W., Pérez-Cueto, F. J. A., Barcellos, M. D. d., Krystallis, A., & Grunert, K. G. (2010). European citizen and consumer attitudes and preferences regarding beef and pork. *Meat Science*, *84*(2), 284–292. https://doi.org/10.1016/j.meatsci.2009.05.001

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., & Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Research*, *41*(6), e74. https://doi.org/10.1093/nar/gkt006

Wang, X., & Kadarmideen, H. N. (2019). Metabolomics analyses in high-low feed efficient dairy cows reveal novel biochemical mechanisms and predictive biomarkers. *Metabolites*, *9*(7), 151. https://doi.org/10.3390/metabo9070151

Washietl, S., Kellis, M., & Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Research*, *24*(4), 616–628. https://doi.org/10.1101/gr.165035.113

Weikard, R., Altmaier, E., Suhre, K., Weinberger, K. M., Hammon, H. M., Albrecht, E., Setoguchi, K., Takasuga, A., & Kühn, C. (2010). Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. *Physiological Genomics*, *42 A*(2), 79–88. https://doi.org/10.1152/physiolgenomics.00120.2010

Weikard, R., Demasius, W., & Kuehn, C. (2017). Mining long noncoding RNA in livestock. *Animal Genetics*, *48*(1), 3–18. https://doi.org/10.1111/age.12493

Weikard, R., Hadlich, F., Hammon, H. M., Frieten, D., Gerbert, C., Koch, C., Dusel, G., & Kuehn, C. (2018). Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. *Oncotarget*, *9*(30), 21052–21069. https://doi.org/10.18632/oncotarget.24898

Weikard, R., Hadlich, F., & Kuehn, C. (2013). Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC Genomics*, *14*(1), 789. https://doi.org/10.1186/1471-2164-14-789

Wenric, S., Elguendi, S., Caberg, J. H., Bezzaou, W., Fasquelle, C., Charloteaux, B., Karim, L., Hennuy, B., Frères, P., Collignon, J., Boukerroucha, M., Schroeder, H., Olivier, F., Jossa, V., Jerusalem, G., Josse, C., & Bours, V. (2017). Transcriptome-wide analysis of natural antisense transcripts shows their potential role in breast cancer. *Scientific Reports*, *7*(1), 17452. https://doi.org/10.1038/s41598-017-17811-2

Widmann, P., Nuernberg, K., Kuehn, C., & Weikard, R. (2011). Association of an ACSL1 gene variant with polyunsaturated fatty acids in bovine skeletal muscle. *BMC Genetics*, *12*. https://doi.org/10.1186/1471-2156-12-96

Widmann, P., Reverter, A., Fortes, M. R. S., Weikard, R., Suhre, K., Hammon, H., Albrecht, E., & Kuehn, C. (2013). A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics*, *14*(1), 17. https://doi.org/10.1186/1471-2164-14-798

Widmann, P., Reverter, A., Weikard, R., Suhre, K., Hammon, H. M., Albrecht, E., & Kuehn, C. (2015). Systems Biology Analysis Merging Phenotype, Metabolomic and Genomic Data Identifies Non-SMC Condensin I Complex, Subunit G (NCAPG) and Cellular Maintenance Processes as Major Contributors to Genetic Variability in Bovine Feed Efficiency. *Plos One*, *10*(4), 22. https://doi.org/10.1371/journal.pone.0124574

William, A., & Simianer, H. (2011). *Tierzucht. Grundwissen Bachelor.* Verlag Eugen Ulmer.

Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., Cirera, S., Fredholm, M., Botherel, N., Leegwater, P. A. J., Le Béguec, C., Fieten, H., Johnson, J., Alföldi, J., André, C., … Derrien, T. (2017). FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Research*, *45*(8), e57. https://doi.org/10.1093/nar/gkw1306

Xia, J., Wishart, D. S., & Valencia, A. (2011). MetPA: A web-based metabolomics tool for pathway

analysis and visualization. *Bioinformatics*, *27*(13), 2342–2344. https://doi.org/10.1093/bioinformatics/btq418

Xia, Y., Xiao, X., Deng, X., Zhang, F., Zhang, X., Hu, Q., & Sheng, W. (2017). Targeting long non-coding RNA ASBEL with oligonucleotide antagonist for breast cancer therapy. *Biochemical and Biophysical Research Communications*, *489*(4), 386–392. https://doi.org/10.1016/j.bbrc.2017.05.136

Xiang, R., Van Den Berg, I., MacLeod, I. M., Hayes, B. J., Prowse-Wilkins, C. P., Wang, M., Bolormaa, S., Liu, Z., Rochfort, S. J., Reich, C. M., Mason, B. A., Vander Jagt, C. J., Daetwyler, H. D., Lund, M. S., Chamberlain, A. J., & Goddard, M. E. (2019). Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), 19398–19408. https://doi.org/10.1073/pnas.1904159116

Xu, C., Xu, Q., Chen, Y., Yang, W., Xia, C., Yu, H., Zhu, K., Shen, T., & Zhang, Z. (2016). FGF-21: promising biomarker for detecting ketosis in dairy cows. *Veterinary Research Communications*, *40*(1), 49–54. https://doi.org/10.1007/s11259-015-9650-5

Yan, K., Arfat, Y., Li, D., Zhao, F., Chen, Z., Yin, C., Sun, Y., Hu, L., Yang, T., & Qian, A. (2016). Structure prediction: New insights into decrypting long noncoding RNAs. *International Journal of Molecular Sciences,* 17(1). https://doi.org/10.3390/ijms17010132

Yang, L., Li, P., Yang, W., Ruan, X., Kiesewetter, K., Zhu, J., & Cao, H. (2016). Integrative Transcriptome Analyses of Metabolic Responses in Mice Define Pivotal LncRNA Metabolic Regulators. *Cell Metabolism*, *24*(4), 627–639. https://doi.org/10.1016/j.cmet.2016.08.019

Yosim, A. E., & Fry, R. C. (2015). Systems Biology in Toxicology and Environmental Health. In R. Fry (Ed.), *Systems Biology in Toxicology and Environmental Health* (1st ed.). Academic Press. https://doi.org/10.1016/B978-0-12-801564-3.00001-8

Zampetaki, A., Albrecht, A., & Steinhofel, K. (2018). Long non-coding RNA structure and function: Is there a link? *Frontiers in Physiology*, *9*:1201. https://doi.org/10.3389/fphys.2018.01201

Zhang, Y., Liu, X. S., Liu, Q. R., & Wei, L. (2006). Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Research*, *34*(12), 3465–3475. https://doi.org/10.1093/nar/gkl473

Zhang, Y. M. (2012). F2 designs for QTL analysis. *Methods in Molecular Biology*, *871*, 17–29. https://doi.org/10.1007/978-1-61779-785-9_2

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D. N., & Perou, C. M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*, *15*(1), 419. https://doi.org/10.1186/1471-2164-15-419

Zheng, X., Ning, C., Zhao, P., Feng, W., Jin, Y., Zhou, L., Yu, Y., & Liu, J. (2018). Integrated analysis of long noncoding RNA and mRNA expression profiles reveals the potential role of long noncoding RNA in different bovine lactation stages. *Journal of Dairy Science*, *101*(12), 11061–11073. https://doi.org/10.3168/jds.2018-14900

Zou, Y., & Xu, H. (2020). Involvement of long noncoding RNAs in the pathogenesis of autoimmune diseases. *Journal of Translational Autoimmunity*, *3*, 100044. https://doi.org/10.1016/j.jtauto.2020.100044

9. Veröffentlichungen

9.1. Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated with Metabolic Efficiency in Cattle.

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A und Kühn C (2019) in *Frontiers in Genetics* 10:1130.

doi: 10.3389/fgene.2019.01130.

frontiers in Genetics

ORIgINAL REsEARC h
published: 22 November 201 9
doi: 10.3389 /fgene.2019.0113 0

Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle

Wietje Nolte [1], ***Rosemarie Weikard*** [1], ***Ronald M. Brunner*** [1], ***Elke Albrecht*** [2], ***Harald M. Hammon*** [3], ***Antonio Reverter*** [4] ***and Christa Kühn*** [1,5]*

[1] *Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany,* [2] *Institute of Muscle Biology and Growth, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany,* [3] *Institute of Nutritional*

Physiology "Oskar Kellner," Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, [4] *Commonwealth*

Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food, Queensland Bioscience Precinct, St Lucia, QLD, Australia, [5] *Faculty of Agricultural and Environmental Sciences, University Rostock, Rostock, Germany*

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
James Reecy
Iowa State University, United States
Kieran G. Meade,
The Irish Agriculture and Food Development Authority, Ireland

***Correspondence:**
Christa Kühn kuehn@fbn-dummerstorf.de

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics
Received: *25 June 2019*
Accepted: *17 October 2019*
Published: *22 November 2019*

Citation:
Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2019) Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle. Front. Genet. 10:1130. doi: 10.3389/fgene.2019.01130

Background: Genomic regions associated with divergent livestock feed efficiency have been found predominantly outside protein coding sequences. Long non-coding RNAs (lncRNA) can modulate chromatin accessibility, gene expression and act as important metabolic regulators in mammals. By integrating phenotypic, transcriptomic, and metabolomic data with quantitative trait locus data in prioritizing co-expression network analyses, we aimed to identify and functionally characterize lncRNAs with a potential key regulatory role in metabolic efficiency in cattle.

Materials and Methods: Crossbred animals (n = 48) of a Charolais x Holstein F_2-population were allocated to groups of high or low metabolic efficiency based on residual feed intake in bulls, energy corrected milk in cows and intramuscular fat content in both genders. Tissue samples from jejunum, liver, skeletal muscle and rumen were subjected to global transcriptomic analysis via stranded total RNA sequencing (RNAseq) and blood plasma samples were used for profiling of 640 metabolites. To identify lncRNAs within the indicated tissues, a project-specific transcriptome annotation was established. Subsequently, novel transcripts were categorized for potential lncRNA status, yielding a total of 7,646 predicted lncRNA transcripts belonging to 3,287 loci. A regulatory impact factor approach highlighted 92, 55, 35, and 73 lncRNAs in jejunum, liver, muscle, and rumen, respectively. Their ensuing high regulatory impact factor scores indicated a potential regulatory key function in a gene set comprising loci displaying differential expression, tissue specificity and loci overlapping with quantitative trait locus regions for residual feed intake or milk production. These were subjected to a partial correlation and information theory analysis with the prioritized gene set.

Results and Conclusions: Independent, significant and group-specific correlations ($|r| > 0.8$) were used to build a network for the high and the low metabolic efficiency group resulting in 1,522 and 1,732 nodes, respectively. Eight lncRNAs displayed a particularly high connectivity (>100 nodes). Metabolites and genes from the partial correlation and information theory networks, which each correlated significantly with the respective lncRNA, were included in an enrichment analysis indicating distinct

affected pathways for the eight lncRNAs. LncRNAs associated with metabolic efficiency were classified to be functionally involved in hepatic amino acid metabolism and protein synthesis and in calcium signaling and neuronal nitric oxide synthase signaling in skeletal muscle cells.

Keywords: *Bos taurus*, metabolic efficiency, co-expression network analysis, long non-coding RNA, Functional Annotation of Animal genome

INTRODUCTION

In recent years the focus of livestock production and farming has shifted in developed countries towards a stronger emphasis on resource efficiency and sustainability (Thornton, 2010). In cattle, energy metabolism, nutrient conversion and efficient use of primary resources are of increasing economic and ecological importance to breeders and consumers. Genomic selection and the use of biomarkers greatly facilitate the improvement of complex phenotypes, e.g. feed efficiency, which remain cost- and time-consuming to measure (Kenny et al., 2018).

Some pivotal gene mutations are known in major livestock production traits, e.g. a meta-analysis on stature in cattle identified *PLAG1* as a major regulator and pointed towards putative causal mutations (Bouwman et al., 2018). In pigs, the scavenger receptor cysteine-rich domain 5 in gene *CD163*, when not being translated, led to resistance to porcine reproductive and respiratory syndrome virus 1 infection (Burkard et al., 2018). Pigs that did not express the receptor protein were susceptible to the infection. For the region between *LCORL* and *NCAPG*, which has been associated with growth or feed efficiency in a number of species (cattle, horse, human), multiple mappings have narrowed down the region of interest but the causal mutation remains unknown (Widmann et al., 2015; Bouwman et al., 2018). A large part of the variation in traits like feed efficiency, growth and carcass traits remains still unexplained (Hardie et al., 2017; Medeiros de Oliveira Silva et al., 2017; Seabury et al., 2017) and genome-wide association studies repeatedly pointed towards quantitative trait loci (QTL) outside protein-coding genes (Ibeagha-Awemu et al., 2016; Seabury et al., 2017; Higgins et al., 2018).

Due to their gene expression regulatory potential, long noncoding RNAs (lncRNAs) have emerged as potential key regulators for diverse biological processes, such as X-chromosomal inactivation and dosage compensation (Brown et al., 1992; Clemson et al., 1996), vernalization/ flowering in plants (Csorba et al., 2014), as well as human cancer biology as reviewed by Serviss et al. (2014).

Recently, lncRNAs have been suggested as therapeutic targets for diabetes and other metabolic diseases because of their involvement in lipid metabolism, adipogenesis and fat deposition (Chen et al., 2018a; Liu et al., 2018; Zeng et al., 2018). In mammals, lncRNAs were further identified as key regulators of energy metabolism and lipogenesis (Yang et al., 2016). In adipocytes, these genomic elements also play an integral part in the insulin-signaling pathway (Degirmenci et al., 2019). A central regulatory role of lncRNAs was furthermore observed in skeletal muscle in myogenesis and muscle cell differentiation: *SYISL* has been shown to regulate myoblast proliferation and fusion and acts in an inhibitory way in myogenic differentiation (Jin et al., 2018), *Irm* enhances myogenic differentiation during myogenesis through the binding to *MEF2D* (Sui et al., 2019), and *lnc-mg* overexpression has directly been linked to muscle hypertrophy in mice, whereas a knock-out led to dystrophy (Zhu et al., 2017). It is likely that lncRNAs contribute significantly to economically important production traits and divergent phenotypes in livestock as well. Since they show little sequence conservation across species and their expression appears to be mainly species specific and spatiotemporal (Ulitsky et al., 2011; Ulitsky and Bartel, 2013), knowledge transfer remains a challenging issue. The identification and functional characterization of lncRNAs needs to be performed for each species, and this fits into one of the major goals of the consortium for the Functional Annotation of Animal Genomes (FAANG, https://www.animalgenome.org/community/FAANG/) that strives to identify and annotate functionally relevant elements in livestock genomes.

Another key feature of lncRNAs is their low expression level compared to protein-coding genes (Derrien et al., 2012), which makes their detection challenging. From transcription factors it is known, that little changes in abundance can however have tremendous consequences if these have high regulatory potential in terms of gene expression (Vaquerizas et al., 2009) and we postulated an analogous phenomenon for lncRNAs. For instance, the knockout of the lowly expressed lncRNA *ßlinc* in mice impaired the correct formation of pancreatic islets and severely changed the glucose homeostasis in adult animals (Arnes et al., 2016). A low and tightly regulated gene expression has implications for differential expression (DE) analyses, because little changes in expression are often not recognized as significant due to lack of power in standard experimental designs. Therefore, other approaches are necessary when aiming to identify and functionally annotate key regulatory lncRNAs. A tested and proven method in the screening for critical transcription factors from gene expression data, which are typically low in abundance but have high regulatory power as reviewed by Vaquerizas et al. (2009), is network co-expression analysis that incorporates the regulatory impact factor (RIF) metrics and a partial correlation and information theory (PCIT) (Reverter et al., 2010; Perez-Montarelo et al., 2012). This approach has previously also led to the identification of regulatory elements associated with puberty (Canovas et al., 2014; Nguyen et al., 2018) and feed efficiency in cattle (Alexandre et al., 2019). We assumed that this rational network approach could also be used as a hypothetical generation tool for the systematic detection of lncRNAs with important regulatory potential.

In this study, we took advantage of a unique F_2 cross-population of meat and dairy cattle breeds (Charolais x Holstein) (Kühn et al., 2002) that has been deeply phenotyped and genotyped.
Earlier studies have shown that in this cross population a gene variant of the *NCAPG* gene is associated with fetal and pubertal growth (Eberlein et al., 2009; Weikard et al., 2010). By integrating quantitative metabolite data with genotype information, this *NCAPG* genotype was found to be associated with plasma arginine levels (Weikard et al., 2010). A systems biology approach, which combined metabolome data, growth-associated phenotypic and genetic information, revealed a functional gene interaction network characterizing the intensive growth phase at the beginning of the pubertal growth interval (Widmann et al., 2013). Potential interaction partners of the *NCAPG* gene were predicted and the functional role of the *NCAPG* gene as a growth regulator linked to the arginine NO metabolism was concluded. A combined phenotype–metabolome–genome analysis was also used to identify genetic switches of associated molecular signaling pathways linked to variance in efficiency of feed conversion (Widmann et al., 2015). This current study on the regulatory role of lncRNAs for metabolic efficiency was aimed to contribute to a more detailed elucidation of the molecular background of this complex physiological trait and help to characterize divergent metabolic types with respect to nutrient partitioning. Therefore, phenotypic information, transcriptomic data from four metabolically relevant tissues and QTL information were used to establish a prioritized gene set that was submitted to the combinational RIF metrics and subsequently to the PCIT algorithm for co-expression network creation. The integration of metabolomic profiles through correlation with transcriptomic data added valuable information for the interpretation of biological functions.

MATERIALS AND METHODS

Design of the study

For this study, we made use of 48 animals (24 bulls, 24 cows) of a F_2-population [SEGFAM (Kühn et al., 2002)] from a Charolais × Holstein cross. The cross population was bred at the Leibniz Institute for Farm Animal Biology in Dummerstorf (Germany) and kept under standardized housing and feeding conditions as previously described (Eberlein et al., 2009; Weikard et al., 2010; Widmann et al., 2011). Males were slaughtered at 18 months of age and females were slaughtered after their second parity at 30 days postpartum. Based on residual feed intake (RFI) in bulls and energy corrected milk yield (ECM_w) in cows as well as intramuscular fat content (IMF) of *M. longissimus dorsi* in both genders, animals were assigned to either of the two groups: high or low metabolic efficiency (**Table 1**). In this study we defined high metabolic efficiency in cattle as the preference to accrete or secrete protein while receiving the same diet as their inefficient conspecifics, which were characterized by a clear tendency to accrete fat instead of protein. In European production systems, those animals are most sustainable and economically efficient producers, which build up protein mass (muscle) with little fat content or, in case of females, secrete high amounts of milk.
Cows were categorized as highly efficient if their milk yield within the 7 days prior to slaughter was above 140 kg energy correct milk (ECM_w) and the carcass fat content (CFC) was less than the average CFC of all cows plus one standard deviation. In contrast, cows were classified as lowly efficient if their milk yield within the last week was between 14 and 40 kg ECM_w and the CFC was above the average CFC of all cows minus one standard deviation. For all cows, the calving interval had to be less than 540 days, the maximum age was 1,510 days and they had to be free of pathological findings with metabolic implications noted after slaughter. Cows that were categorized as highly efficient (high ECM_w) on average had a lower CFC (mean 17.1%, SD 2.7%) and lowly efficient cows (low ECM_w) had a higher CFC (mean 25.9%, SD 3.6%) than the mean of the population (21.8%, SD 5.3%, n = 242). In addition, highly efficient cows had a lower IMF (mean 4.16%, SD 1.60%) and the lowly efficient cows had a higher IMF (mean 6.46%, SD 2.53%) than the mean of the population (5.21%, SD 2.21%, n = 242).
The individual milk volume yield per cow was measured on a daily basis and the milk composition was determined once per week. The trait included in cow selection for this study corresponded to the weekly ECM determined for the 7 days before slaughter (ECM_w). The formula presented by Kirchgeßner (1997) was modified accordingly for the one week interval (F% = milk fat percentage, P% = milk protein percentage): cows, the ECM_w was used as a substitute feature for feed efficiency, because the facilities did not allow for RFI measurement in cows during the time of the experiment.

$$ECM_w = \frac{0.37\,F\% + 0.21\,P\% + 0.95}{3.1} \times MY - 7d$$

TABLE 1 | Sample characteristics.

Metabolic efficiency group	Number of animals	Sex	RFI[1] in last month of life (bulls)	ECM^2_w (cows)	IMF[3] (both sexes)	CFC[4] (both sexes)
			μ[5] (SD[6])	μ (SD)	μ (SD)	μ (SD)
High	25	12 males 13 females	-21.30 (4.44)	190.87 (22.02)	3.46 (1.30)	15.93 (3.16)
Low	23	12 males 11 females	20.83 (4.41)	30.97 (9.18)	5.51 (2.34)	22.93 (4.88)

[1]RFI, residual feed intake; [2]ECM_w, energy corrected milk 7 days before slaughter; [3]IMF, intramuscular fat content (given in percent, measured in M. longissimus dorsi); [4]CFC, carcass fat content; [5]μ, mean; [6]SD, standard deviation.

For bulls, the decisive factor for animal selection was RFI calculated for the last month prior to slaughter. The RFI equals the animals' energy intake while considering the average daily gain and metabolic mid-weight (average body weight of months of life 17 to 18 raised to the power of 0.75) (Archer et al., 1997).

Bulls with a low RFI (at least 1 standard deviation below average) were assigned to the high metabolic efficiency group and bulls with a high RFI (at least one standard deviation above average) were assigned to the low metabolic efficiency group. In their last month of life, all bulls had to have a positive daily weight gain and no less than the population average minus one standard deviation. Bulls that were categorized as highly efficient (negative RFI) on average had a lower CFC (mean 14.2%, SD 3.0%) and lowly efficient bulls (positive RFI) had a higher CFC (mean 20.2%, SD 4.4%) than the population mean (mean 16.5%, SD 4.0%, n = 246). Analogously to cows, highly efficient bulls had a lower IMF (mean 1.71%, SD 1.00%) and the lowly efficient bulls had a higher IMF (mean 4.64%, SD 1.84%) than the population mean (mean 3.67%, SD 1.76%, n = 246).

Plasma Metabolic Profiles

Blood samples were collected from all individuals (n = 48) at slaughter. Plasma samples were sent to Metabolon Inc. (Durham/NC, USA) for the establishment of holistic metabolite profiles that included 640 biochemical compounds and molecules. Metabolites with more than five animals with missing data were excluded. After this filtering step, 490 metabolites remained and missing values were imputed with the minimum measurement, assuming that missing values were due to concentrations below the detection limit. Values were then scaled without centering for each metabolite in R (Core Team 2018) with the scale-function.

All experimental procedures were carried out according to the German animal care guidelines and were approved and supervised by the relevant authorities of the State Mecklenburg-Vorpommern, Germany (State Office for Agriculture, Food Safety and Fishery; LALLF M-V/TSD/7221.3-2.1-010/03).

Sampling, RNA Isolation, Library Preparation, and sequencing

Tissue samples were collected from jejunum mucosa, liver (*Lobus caudatus*), skeletal muscle (*M. longissimus dorsi*), and rumen (*Saccus ventralis,* papillary base) directly after slaughtering and dissection, shock frozen in liquid nitrogen and subsequently stored at -80°C.

For RNA extraction from muscle and rumen, frozen samples (100 mg) were treated with 1 ml TRIzol reagent (Invitrogen, Darmstadt, Germany) and subjected to the Precellys-24 homogenizer (5,500 rpm, 2 × 15 s, lysing kit containing 1.4 mm ceramic beads). For RNA extraction from liver and jejunum, frozen tissue samples were grinded in liquid nitrogen and 30 mg were used for further purification steps. No TRIzol was used for liver and jejunum samples. All samples were then subjected to an on-column-purification step with the NucleoSpin RNA II kit (Macherey & Nagel, Düren, Germany) including a DNase digestion to remove genomic DNA. In addition, the RNA was tested for remaining traces of DNA contamination and, in case of remaining DNA residues, further cleansed according to Weikard et al. (2012).

The RNA concentration and integrity were measured with a Qubit Fluorometer (Invitrogen, Germany) and a 2100 Bioanalyzer Instrument (Agilent Technologies, Germany). Stranded, ribodepleted and indexed libraries were prepared from 1 µg total RNA using the TruSeq Stranded Total RNA Ribo-Zero H/M/R Gold Kit (Illumina, San Diego, USA) and subjected to paired-end sequencing (2 × 100 bp) in a multiplexed design on a HiSeq 2500 Sequencing System (Illumina).

Alignment and Assembly

After quality control of raw sequencing reads with FastQC (Andrew, 2010), adapter and quality trimming were performed with Cutadapt v. 1.16 (Martin, 2011) and Quality Trim v. 1.6.0 (Robinson, 2015), respectively. In Quality Trim the start of sequences was also trimmed (option -s) and the maximum number of N bases was set to 3, while the minimum base quality was set to 15. Reads were then mapped in a guided alignment with HISAT2 v.2.1.0 (Kim et al., 2015) to the bovine reference genome UMD.3.1 [Ensembl annotation release 92 (Frankish et al., 2017)]. After sorting and indexing of BAM files with samtools v.1.6 (Li et al., 2009), samples were individually assembled with Stringtie v.1.3.4d (Pertea et al., 2015) based on the reference genome and annotation used for alignment. Using the individually assembled samples (n = 204) from all four tissues and the bovine reference genome, we built a new merged annotation in Stringtie across tissues, while specifying for minimal transcript coverage across samples of 15 read alignments per exonic base. In addition to the 192 samples (48 animals, four tissues) included in the subsequent steps for DE and network analyses, we also took benefit from rumen, liver and muscle samples of further four individuals from the same experimental herd. These samples were subjected to exactly the same processing steps as the 192. The new merged annotation was used for fragment counting with featureCounts (subread v.1.6.1) (Liao et al., 2014), while allowing for fractional counting and specifying for reverse strandedness.

Long Non-Coding RNA Prediction and Fragment Counting

LncRNAs were identified *in-situ* with FEELnc (Wucher et al., 2017), a bioinformatics tool for lncRNA prediction and annotation, using the merged transcript annotation and the bovine reference genome and annotation UMD3.1 release 92. FEELnc excludes transcripts annotated as protein coding and subsequently keeps transcripts with a minimum length of 200 nt and at least two exons and only monoexonic transcripts with antisense localization. Other monoexonic transcripts were excluded to reduce the number of false positives, which might arise from the mapping of repetitive sequences (Wucher et al., 2017), DNA contamination (Haerty and Ponting, 2015) and in general transcriptional noise (Kern et al., 2018). For those transcripts matching the requirements, the coding potential of remaining transcripts was determined in shuffling mode.

Fragment Count Normalization

For further pipeline steps, except for the DE analysis, fragments per kilobase million (FPKM) were calculated from the featureCounts derived fragment counts. Genes were filtered for a minimal average expression value of 0.2 FPKM in at least one of the four tissues and ribosomal and spliceosomal RNA genes were excluded (Metazoan signal recognition particle RNA, U6 spliceosomal RNA, small nucleolar RNA U6-53). For further analyses of FPKM values performed in this study, a $\log_2$-scale of the data was used (for log transformation a pseudo-count of 0.001 was added).

Prioritized gene List

Gene co-expression networks are a useful tool when trying to deduce the potential biological function of genes, novel loci and non-coding elements (van Dam et al., 2017), assuming the guilt-by-association principle. In order to create meaningful networks that have a targeted focus on our phenotype (metabolic efficiency), we created a set of prioritized genes where genes had to belong to at least one of these four categories: differentially expressed (DE) genes in at least one of the four investigated tissues, tissue-specific (TS) genes, genes harboring a QTL for milk production or RFI (QTL) according to the literature, and predicted lncRNAs. Small nucleolar RNAs (snoRNAs), ribosomal RNAs, spliceosomal RNAs, and Y-RNAs were excluded from the set.

Differential Expression Analysis

A DE analysis for the high and low metabolic efficiency group was performed within tissues and across sexes in R with the package DEseq2 (Love et al., 2014). Fragment counts from featureCounts were used as input and normalization was performed within DEseq2. To exclude very lowly expressed transcripts within a tissue, the minimal fragment count threshold was set to at least 10 fragments for 10 out of 48 individuals. Ribosomal genes were excluded from the analysis and year of slaughter and sex were used as factors in the model. The significance threshold was set to $q < 0.05$ [Benjamini–Hochberg (BH) test].

Tissue Enriched Genes

The expression ($\log_2$-transformed FPKM) of a gene was defined as enriched in a particular tissue, if the abundance in the other three tissues was less than half the average across all tissues and above the average plus one standard deviation in the tissue at hand. Throughout the further course of this study, we refer to these genes as TS.

Genes Harboring a Quantitative Trait Locus

We extracted QTL for milk production traits (MY) and RFI in cattle from the Animal QTL database (Park et al., 2018) and then screened our dataset in Ensembl Biomart (http://asia.ensembl. org/biomart/martview, accession date 28 March 2019) for genes that overlapped with these QTL regions. A physical overlap of the QTL and the gene is needed for a gene hit, while close neighborhood is not sufficient.

Regulatory Impact Factor Analysis

The RIF (Reverter et al., 2010) analysis makes use of two alternative metrics (RIF1 and RIF2) that attribute scores to potential key regulators. The strength of the score depends on the change in correlation between the regulator and its target in two groups or treatments, the level of DE of the target gene, and the general expression level of the target gene. We conducted RIF analyses within tissues and across metabolic efficiency groups to assess the regulatory capacity of lncRNAs in a set of prioritized genes (lncRNA, DE, TS, QTL harboring). Therefore, RIF metrics were calculated within each tissue for a prioritized gene set (including $\log_2$(FPKM) data) that comprised genes which were DE or TS in that tissue, harbored a QTL or were characterized as a lncRNA. Naturally, some of the QTL-genes might have zero expression in one or more of the tissues. To prevent erroneously high RIF scores stemming from low variation in gene expression, an additional filter for expression level was applied (on top of minimal average expression of 0.2 FPKM in at least one tissue). Only genes with abundance above tissue average were kept for the RIF analysis.

A high RIF1 score was assigned to lncRNAs that were consistently co-expressed with abundant target genes in both metabolic efficiency groups. A high RIF2 score was attributed to lncRNAs that displayed the most altered ability to predict the abundance of target genes between groups, meaning that a lncRNA exhibited strong correlation to a target on one condition but none or a reverse correlation in the other. RIF scores were standardized with a z-score. Key regulators (lncRNA) were considered of significant importance and were included in further analyses if they had an absolute RIF1 or RIF2 z-score of ≥ 1.96, meaning that these lncRNAs and their scores were outside the 95% confidence interval, corresponding to a significance level of $p = 0.05$ in a t-test.

Partial Correlation and Information Theory

The PCIT (Reverter and Chan, 2008) tests for significant pairwise correlations between two elements while accounting for all possible three-way combinations in the dataset that include either of the pair elements. Importantly, the PCIT recognizes independent, significant correlations regardless of the strength of correlation. Within the high and low metabolic efficiency groups, the PCIT approach across all tissues was used to investigate for independent correlations of lncRNAs that had significant RIF scores with DE genes, TS genes, and QTL harboring genes.

Results were filtered for significant correlations (minimal correlation strength $|r| > 0.8$) between a lncRNA and another gene that were exclusive for the high or low metabolic efficiency group, meaning that the correlation was significant in one group but not in the other. The visualization was realized in Cytoscape 3.6.1 (Shannon et al., 2003).

Characterization of Key Regulatory Long Non-Coding RNAs

Blast Search Against New Bovine Assembly

Highly connected lncRNAs with more than 100 directly linked nodes (genes) were selected from each network for further scrutiny. Since the prediction of lncRNAs was based on a merged annotation, which was reference guided by UMD3.1, Ensembl release 92, we wanted to investigate the sequence homology

and annotation status of key lncRNAs in the new bovine assembly ARS1.2 annotated in Ensembl release 95. The lncRNA sequences were blasted online with the blastn suite using the MegaBlast algorithm, specifying for high sequence similarity and otherwise default parameters (Altschul et al., 1990) (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed Mai 2019) against the new bovine assembly (ARS-UCD1.2, https://www.ncbi.nlm.nih.gov/assembly/GCA_002263795.2; GenBank accession NKLS00000000.2; https://www.ensembl.org/Bos_taurus/Info/Index). We considered blast hits to indicate high homology if the sequence identity was at least 98% in a region covering at least 200 nucleotides.

Pathway Enrichment Analysis

To assess the possible biological function of high connectivity lncRNAs, we performed a pathway enrichment analysis based on genes identified as correlated ($|r| > 0.8$) in the PCIT analyses and also including blood plasma metabolites that were significantly ($p \leq 0.05$) correlated with the high connectivity lncRNAs. To this end, a pairwise Pearson correlation analysis between bloodplasma metabolites and lncRNA expression in the tissue, where the lncRNA was most abundant, was performed in R with the function rcorr of the Hmisc package (Harrell and Frank, 2019). The list of significantly correlated metabolites ($p \leq 0.05$) and genes (adjacent network nodes with $|r| > 0.8$) were analysed using the Ingenuity Pathway Analysis (IPA: QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) (Kramer et al., 2014). The workflow from group formation and tissue sampling up to the functional characterization of key lncRNAs is visualized for better comprehensibility and clarity in **Figure 1**.

RESULTS

RNA Preparation, sequencing, Alignment, and Mapping

The average RNA integrity (RIN) across the four tissues was 8.22 ± 0.81 (**Table 2**). After quality trimming the average RNA sequencing depth was at 48 million read pairs per sample. A total of 9 out of 192 samples reached less than a 40 million read pair coverage. The alignment of reads with HISAT2 to the bovine reference genome UMD.3.1 (Ensembl release 92) resulted in an average alignment rate of 92.98 ± 9.50%. Compared with the other tissues, rumen scored a distinctly lower rate (78.00 ± 7.75%). The average mapping rate across all samples to the customized annotation, which contained 30,072 loci, was 81.89%. The tissue specific average mapping rate was lowest in rumen, of comparable dimension in jejunum and muscle, and highest in liver.

Long Non-Coding RNA Prediction

Based on the merged annotation, FEELnc predicted 26,740 mRNAs and 7,646 lncRNA transcripts (3,287 loci), out of which 544 were without potential positional interaction partner gene within the default window size of 10,000 to 100,000 nucleotides. Those 7,102 lncRNA transcripts with an assigned potential positional interaction partner were generated by 3,051 loci (**Table 3**, **Supplementary Table 1**). FEELnc distinguishes between intergenic and genic lncRNA with different subtypes (see Wucher et al. (2017) for a graphical explanation). LncRNAs are also classified according to their position to neighboring protein coding genes (interaction partner gene). For intergenic lncRNAs, the best partner gene is closest in terms of distance in base pairs and for genic lncRNAs the best partner gene directly overlaps with it, preferably at an exon. All predicted 7,646 lncRNA transcripts were considered for further computational analyses.

The total of 3,287 lncRNA loci are equally distributed in terms of strandedness (50.6% on the plus strand, 49.41% on the minus strand), and in a locus-based approach (considering the transcript with highest exon number for each locus) the median number of exons per transcript was 3 (average number of exons per transcript: 4.9 ± 8.2). The total exon length geometric mean of the lncRNA loci amounted to 2,201.0 bp.

Prioritized gene List for Co-Expression Analysis

After filtering the 30,072 genes in the merged annotation for minimal expression (average FPKM across all samples >0.2 in at least one tissue) and exclusion of ribosomal and spliceosomal RNA genes, the dataset contained 22,625 genes out of which 2,886 were lncRNAs, meaning that 401 lncRNAs were removed from RIF and subsequent PCIT co-expression analysis due to very low abundance.

Differential Expression Analysis

The DE analysis yielded a total of 2,154 unique significantly ($q < 0.05$) DE genes between the high and low metabolic efficiency group with 496 DE genes in jejunum, 1,286 DE genes in liver, 479 DE genes in muscle, and no significant differences in rumen (**Figure 2A**). Generally, we observed little overlap of differentially expressed loci between tissues. Out of these unique 2,154 DE genes, 238 were predicted to be lncRNAs corresponding to 11.05%. We observed 40 DE lncRNAs in jejunum, 173 DE lncRNAs in liver, 40 DE lncRNAs in muscle, and none in rumen (**Figure 2B**).

Tissue Enriched Genes

We found a total of 930 genes to be tissue-specifically expressed out of the 22,625 genes, which had passed the initial minimal expression threshold (average expression > 0.2 FPKM in at least on tissue). Out of those 930 genes, 279 were TS in jejunum, 283 in liver, 204 in muscle, and 164 in rumen. Thereof, 21.9% were lncRNAs with 42 in jejunum, 65 in liver, 48 in muscle, and 49 in rumen.

Quantitative Trait Locus Harboring Genes

The database AnimalQTL listed 278 QTL for RFI and 1,881 QTL for milk production traits, which were distributed across 1,615 genes out of which 1,064 passed the minimal expression threshold (average expression > 0.2 FPKM in at least one tissue) in our dataset

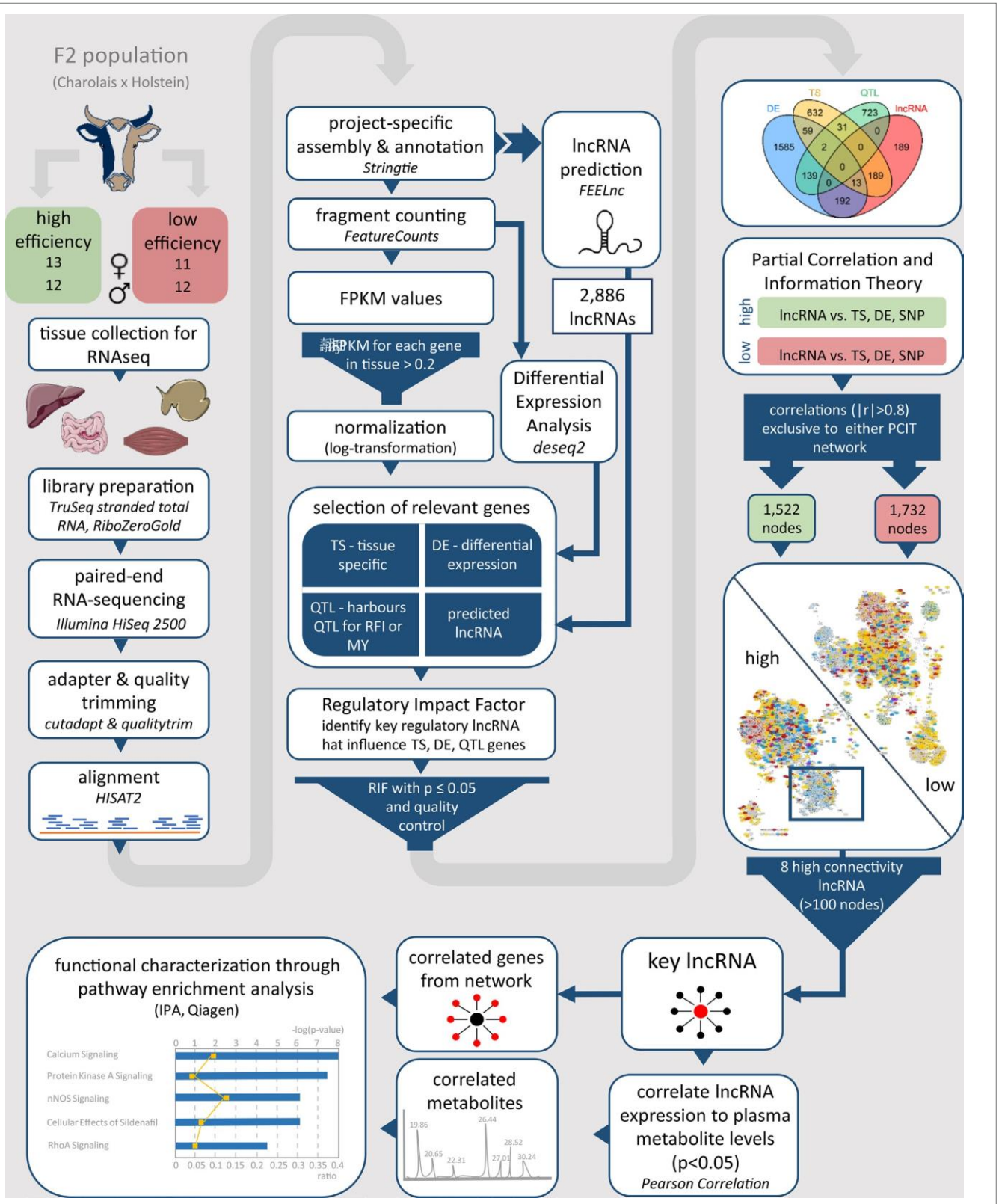

FIGURE 1 | Workflow for the identification and functional characterization of key lncRNAs with regulatory potential in two contrasting biological conditions. The phenotypes under investigation were high and low metabolic efficiency in a Charolais x Holstein cross-population. lncRNA, long non-coding RNA; FPKM, fragments per kilobase transcript length per million reads; TS, tissue specific; DE, differentially expressed; QTL, quantitative trait locus; RFI, residual feed intake; MY, milk production; RIF, regulatory impact factor; PCIT, partial correlation and information theory.

TABLE 2 | Overall and tissue-specific RNA sequencing, alignment, and mapping statistics.

	RIN[1]		Sequencing depth [read pairs]		Alignment to UMD.3.1 [%]		Mapping to project-specific annotation [%]	
	μ[2]	SD[3]	μ	SD	μ	SD	μ	SD
All	8.22	0.81	48,041,209	5,601,638	92.98	9.50	81.89	8.67
Jejunum	8.73	0.44	48,954,376	3,993,201	96.91	0.31	84.99	2.20
Liver	8.00	0.62	50,093,826	5,869,833	98.43	0.20	91.36	1.21
Muscle	7.55	0.85	47,117,156	5,815,843	98.59	0.13	82.42	1.79
Rumen	8.41	0.86	45,999,477	5,587,407	78.00	7.75	69.05	4.67

[1]*RIN, RNA integrity number,* [2]*μ, mean,* [3]*SD, standard deviation.*

Regulatory Impact Factor to select Long Non-Coding RNAs With a Potential Regulatory Effect on Metabolic Efficiency

The input prioritized gene lists filtered for expression level for the tissue specific RIF analysis contained 2,097 loci for jejunum (880 lncRNAs), 1,890 loci for liver (614 lncRNAs), 961 loci for muscle (363 lncRNAs), and 1,458 loci for rumen (755 lncRNAs). RIF scores were then calculated for the lncRNAs in these gene sets.

With a significance threshold of a RIF1 or RIF2 score ≥ 1.96, the tissue specific RIF analyses identified 92 potential key lncRNAs in jejunum, 55 in liver, 35 in muscle, and 73 in rumen. In total 240 unique lncRNAs had a RIF score ≥ 1.96 in at least one tissue and were considered for subsequent PCIT analysis.

Partial Correlation and Information Theory Approach to Identify Long Non-Coding RNA-Associated Co-Expression Networks

For the within-tissue RIF analysis, the sets of DE genes, TS genes, QTL harboring genes and lncRNAs had been filtered for a seizable expression level (abundance above average expression in the respective tissue) to facilitate a reliable calculation of correlation. For the PCIT analysis, a similar filter for minimal expression was applied: abundance above average expression across all samples in at least one tissue when combining DE genes and TS genes from all tissues with the QTL genes and lncRNAs with significant RIF scores. A total of 295 of the 4,049 prioritized loci were excluded due to not meeting this expression limit. The set of prioritized genes that was used for the final PCIT network analysis contained 3,754 unique genes in total. Thereof, 1,990 were DE genes, 895 QTL containing genes, 926 TS genes, and 583 lncRNAs, though some genes belonged to several categories (**Figure 3**, **Supplementary Table 2**).

The PCIT analysis was performed across tissues and results were filtered for significant correlations with a correlation strength $|r| \geq 0.8$, between a lncRNA with significant RIF score and all genes from the prioritized gene list already used for RIF calculation. Furthermore, correlations had to be exclusive to either the high or low metabolic efficiency group. The high and low network contained 1,522 and 1,732 nodes (genes) respectively (**Supplementary Figure 1**, **Supplementary Figure 2**, **Supplementary Table 3**). Six and two lncRNAs showed a high connectivity (>100 nodes) exclusively in one of the two networks, which represent high and low metabolic efficiency, respectively. Thus, these eight lncRNAs stand out as potential regulatory keys for lncRNAs with respect to metabolic efficiency.

Characterization of Key Regulatory Long Non-Coding RNAs in the Networks

Blast Against New Bovine Assembly

The eight lncRNAs characterized by high connectivity for high and low metabolic efficiency in the PCIT analysis were blasted against the new bovine assembly and annotation [ARS-UCD.1.2, National Center for Biotechnology Information (NCBI) release 106] (**Table 4**). If lncRNAs completely overlapped with annotated genes, the respective lncRNA was located on the opposite strand to the annotated gene (e.g. MSTRG.4926 overlapped with *CDH17* on the opposite strand). None of the eight lncRNA loci had yet been annotated as non-coding in the NCBI or the Ensembl genome annotation (ARS-UCD1.2, release 95).

Pathway Enrichment Analysis

The Pearson correlation analysis between blood plasma metabolites and lncRNA expression, which was calculated prior to the pathway enrichment analysis, showed that the eight key lncRNAs were significantly ($p < 0.05$) correlated to very different numbers of metabolites. Correlations ranged from one (MSTRG.18433) to 117 (MSTRG.4740) metabolites, out of which an average of 75% was successfully mapped in the IPA database and used in the subsequent enrichment analyses (**Supplementary Table 4**). The correlation strength ranged from -0.53 to + 0.48 with an average of |0.35|.

Pathway enrichment analysis for each of the eight key lncRNAs with their respective correlated metabolites and genes showed that calcium signaling was the most strongly enriched canonical pathway for half of the key lncRNAs (MSTRG.9051, MSTRG.10337, MSTRG.18433, and MSTRG.19312). The other high ranking canonical pathway hits, i.e. hits with the lowest p-value, were tRNA charging, leukocyte extravasation signaling, caveolar-mediated endocytosis signaling, and T cell receptor signaling (data not shown).

Within the eight lncRNAs with a high connectivity in the PCIT analysis, three loci showed distinct pattern in the pathway enrichment analysis suggesting divergent molecular functions. Inspection of the results showed that the enriched canonical pathways for MSTRG.4740, which was differentially expressed in

TABLE 3 | Characterization of high connectivity long non-coding RNAs from networks specific for high or low metabolic efficiency in cattle.

Identifier		FEElnc Prediction (based on UMD3.1 release 92)			FPKM		Position & Structure			Differential Expression Analysis		
MSTRG	TN	Closest partner gene (gene symbol)	Direction, type, location	Distance bp	Mean	BTA	Exons	Start bp	Strand	Exonic length	Log2 FC	p-value	q-value (BH)	Tissue
MSTRG.4740	1	ENSBTAG00000002062 (TRPA1)	Anti-sense, genic, intronic	0	6.74	14	2	37760114	+	437	0.77	2.10E-04	9.13E-03	Liver
MSTRG.4926	1	ENSBTAG00000021964 (CDH17)	Anti-sense, genic, exonic	0	0.78	14	18	72437085	-	3,321	NA	NA	NA	None
MSTRG.9051	1	ENSBTAG00000004651 (NME1)	Anti-sense, intergenic, downstream	358	2.57	19	2	36225984	-	1,866	NA	NA	NA	None
	2			1,587				36227213						
MSTRG.10337	1	ENSBTAG00000005353 (DES)	Anti-sense, genic, exonic	0	5.37	2	9	108075380	-	2,989	NA	NA	NA	None
MSTRG.17681	1	ENSBTAG00000005726 (HNRNPA2B1)	Anti-sense, intergenic, upstream	9,170	19.80	4	3	70184682	-	1,552	-0.38	8.45E-05	5.04E-03	Liver
MSTRG.18433	1	ENSBTAG00000015828 (FKBP11)	Sense, intergenic, upstream	5,808	8.94	5	2	31038376	+	2,683	NA	NA	NA	None
MSTRG.19098	1	ENSBTAG00000046324 (C-type lectin domain family 2 member D11)	Anti-sense, genic, exonic	0	1.72	5	5	100926473	+	1,570	-0.63	1.27E-06	3.04E-04	Liver
MSTRG.19312	1	ENSBTAG00000009886 (KDELR3)	Anti-sense, genic, exonic	0	25.84	5	3	110665971 110670286	-	5,768	NA	NA	NA	None
	3													

**MSTRG, identifier from project-specific bovine transcriptome annotation; TN, transcript number; FPKM, fragments per kilobase million; BTA, bovine chromosome, bp, base pair; FC, foldchange; BH, Benjamini–Hochberg, NA, not available.*

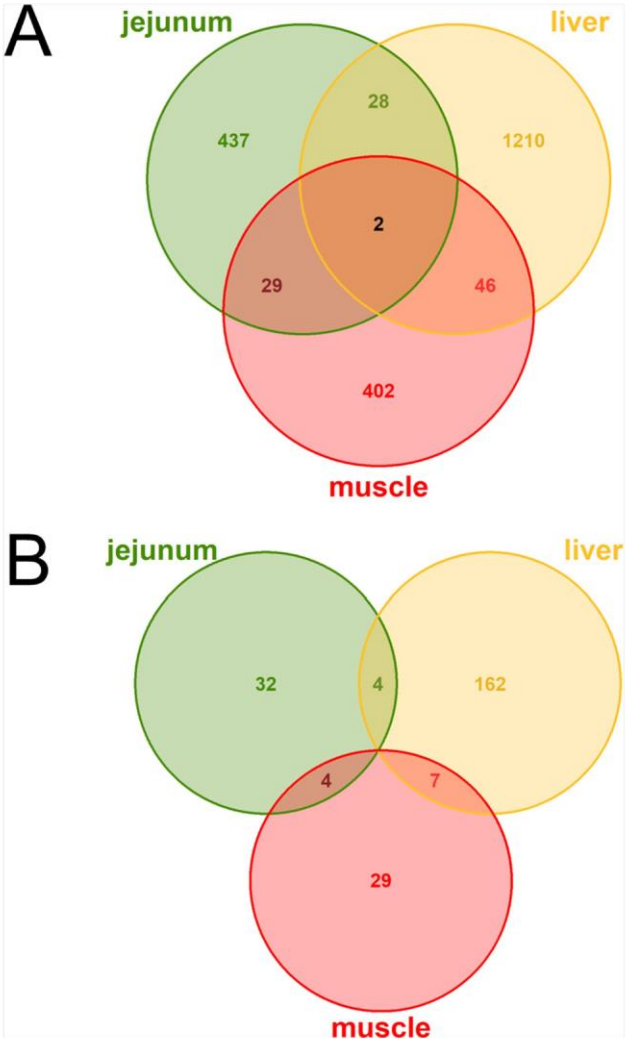

FIGURE 2 | Venn diagrams of **(A)** all loci **(B)** exclusively lncRNAs with differential expression (DE) between high and low metabolic efficiency in cattle. DE analysis was performed within the tissues jejunum, liver, muscle, and rumen. No loci were significantly [q-value (Benjamini–Hochberg) < 0.05] DE in rumen.

liver (**Figure 4**, **Table 3**, **Supplementary Table 5**), were related to amino acid biosynthesis and metabolism, as well as protein synthesis (**Table 5**). MSTRG.17681 (**Figure 5**, **Supplementary Table 5**) which was also differentially expressed in liver, seemed to act very locally in the coatomer subunit of the coat protein I (COPI) in the caveosome. MSTRG.10337, (**Figure 6**, **Supplementary Table 5**) apparently acts specifically in muscle where it was related to several signaling pathways, most strongly to calcium, protein kinase A, neuronal nitric oxide synthase (nNOS), and RhoA signaling (**Table 5**).

DISCUSSION

A major goal of this study was the identification of lncRNAs that hold a potential key regulatory role in metabolic efficiency,

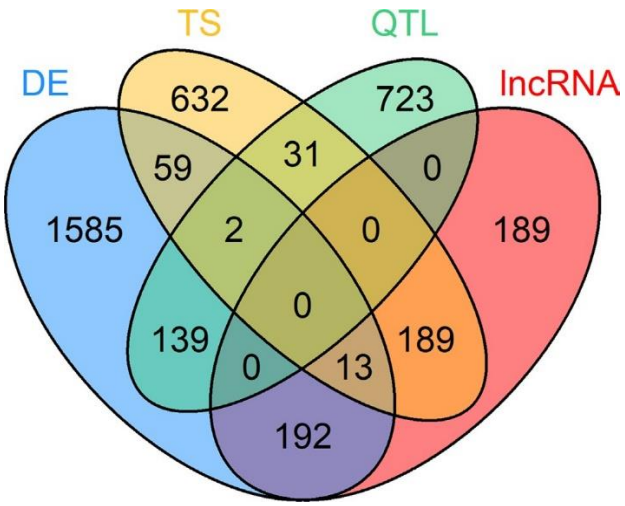

FIGURE 3 | Venn diagram of 3,754 loci selected for co-expression network construction. Loci belonging to at least one of these four categories were considered: differential expression (DE) in at least one tissue, tissue specific (TS) expression, harboring a QTL for residual feed intake and or milk production (QTL) and key regulatory long non-coding (lnc) RNAs [significant ($p < 0.05$) regulatory impact factor score].

which was roughly defined as the animal's ability to direct the energy adsorbed into protein synthesis and use it for muscle mass accumulation or milk secretion. We integrated phenotypic, metabolomics and transcriptomics data from a cattle F_2-population (Charolais × Holstein) in a co-expression network approach to mine for lncRNAs with a regulatory role in metabolic processes. By contrasting animals of high and low metabolic efficiency and by including RNAseq data from four key metabolic tissues in a combined analysis, we identified highly connected hub lncRNAs. Finally, we subjected metabolites and genes, whose plasma levels or transcript abundance significantly correlated with expression levels of the specific, highly connected lncRNA, to the integrative approach for metabolomics and transcriptomics data as offered by the cross-platform IPA (Kramer et al., 2014).

Establishment of a Pipeline Based on Regulatory Impact Factor and Partial Correlation and Information Theory to Establish Co-Expression Networks for Long Non-Coding RNAs and Genes to Predict Their Role in Metabolic Efficiency

Weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) is a frequently applied method to identify co-expression pattern at whole transcriptome level. Recently, Sun et al. (2019) applied this method for mining regulatory signatures of divergent feed efficiency in beef cattle investigating a multi-tissue transcriptome data set. WGCNA has also been used to find hub lncRNAs in a transcriptomic landscape in multiple studies in humans as well as animals (Miao et al., 2016; Tang et al., 2017; Li et al., 2018; Weikard et al., 2018; Wang et al., 2019). To mine for the functional role of lncRNAs of interest via WGCNA, one might select lncRNAs that are strongly correlated with coding neighbor genes (Li et al., 2018) or lncRNAs that were differentially expressed between conditions or

TABLE 4 | BLAST results for eight high connectivity long non-coding RNAs (>100 nodes) in partial correlation and information theory networks with connections exclusive for high or low metabolic efficiency.

lncRNA		BLAST against bovine reference genome (ARS-UCD1.2, release 106)				
Identifier	**Network (connectivity in nodes)**	**Annotated gene with highest sequence homology**	**Identity [%]**	**Query cover [%]**	**E-Value**	**Position of lncRNA relative to homologous gene in ARS-UCD1.2**
MSTRG.4740	Low (147)	mRNA-transient receptor potential cation channel subfamily A member 1 (TRPA1)	100.00	100.00	9.00E-116	Intronic, anti-sense
		ADP-ribosylation factor 4 (ARF4)	98.57	91.00	3.00E-100	Exonic, sense
MSTRG.4926	High (144)	Cadherin-17 precursor (CDH17)	100.00	100.00	0.00E+00	Anti-sense
MSTRG.9051	High (170)	Nucleoside diphosphate kinase A 1 isoform X1 (NME1)	99.72	100.00	0.00E+00	Sense, genic
MSTRG.10337	Low (239)	Desmin (DES)	99.93	100.00	0.00E+00	Exonic, anti-sense
MSTRG.17681	High (120)	39,201 bp at 5' side: alpha-aminoadipic semialdehyde synthase, mitochondrial precursor 88559 bp at 3' side: fez family zinc finger protein 1	98.40	99.00	0.00E+00	Sense, genic
		Chromobox protein homolog 3 isoform X1 (CBX3)	99.00	89.00	0.00E+00	Sense, genic
MSTRG.18433	High (268)	364 bp at 5' side: ADP-ribosylation factor 3; 37831 bp at 3' side: peptidyl-prolyl cis-trans isomerase FKBP11 precursor	99.96	100.00	0.00E+00	Sense, intergenic
MSTRG.19098	High (184)	C-type lectin domain family 2 member D11	100.00	100.00	0.00E+00	Anti-sense, genic
MSTRG.19312	High (212)	ER lumen protein-retaining receptor 3 (KDELR3)	100.00	99.00	0.00E+00	Anti-sense, genic

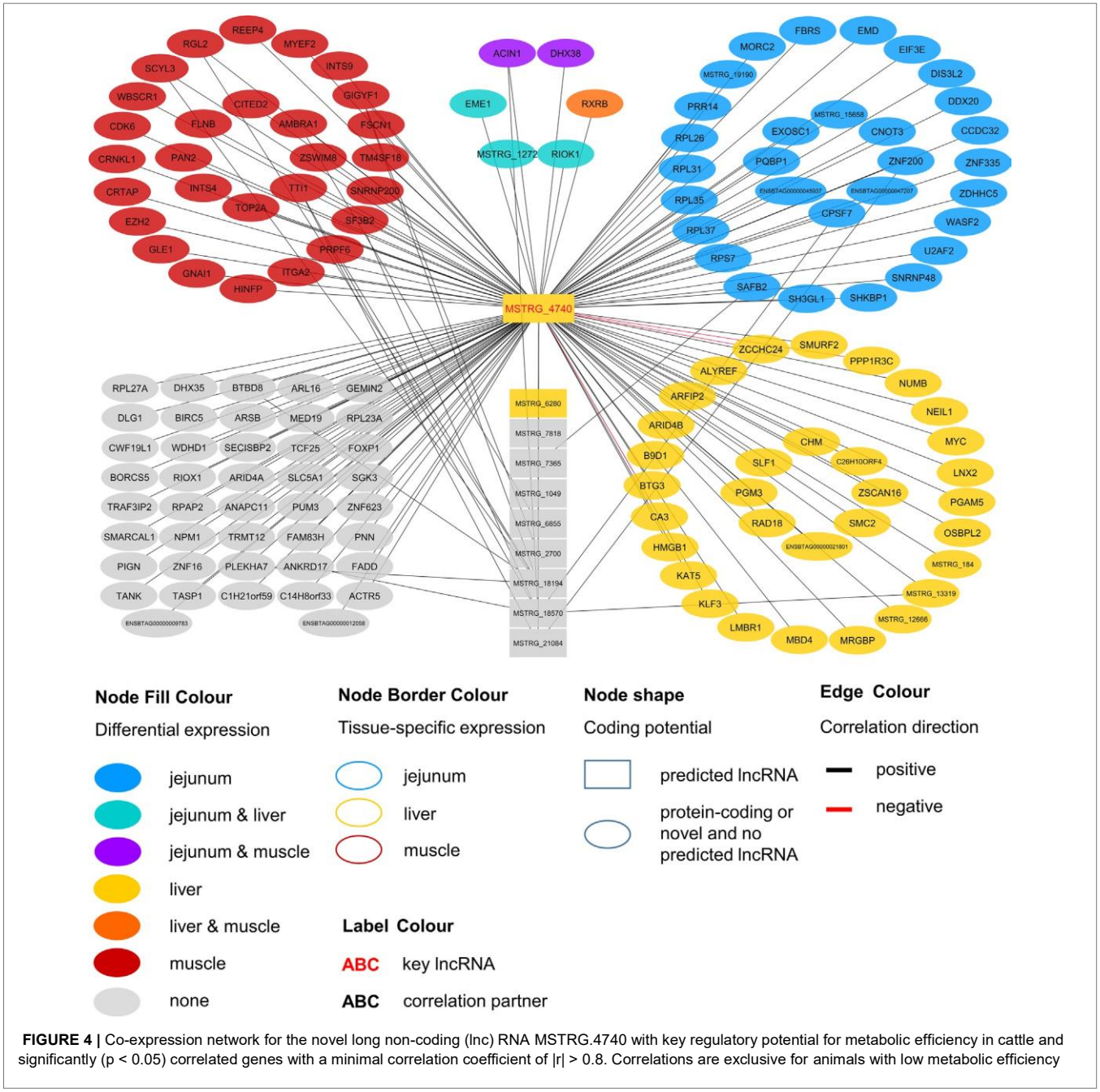

FIGURE 4 | Co-expression network for the novel long non-coding (lnc) RNA MSTRG.4740 with key regulatory potential for metabolic efficiency in cattle and significantly ($p < 0.05$) correlated genes with a minimal correlation coefficient of $|r| > 0.8$. Correlations are exclusive for animals with low metabolic efficiency

The connectivity within a network and the differential wiring between two networks can also serve as a selection criterion (Pellegrina et al., 2017). In our study we present an alternative approach for the selection of lncRNAs of interest, the RIF (Reverter et al., 2010), which has already successfully been applied to transcription factors (TF). In combination with a PCIT (Reverter and Chan, 2008), key regulatory TFs during puberty could be identified in cattle (Cánovas et al., 2014), as well as critical TFs in porcine muscle (Perez-Montarelo et al., 2012). This approach seemed to be particularly applicable for lncRNAs with regard to the expression level as they generally exhibit lower transcript abundance compared with mRNAs (Derrien et al., 2012), as do TFs compared with other coding genes (reviewed by Vaquerizas et al., 2009). We indeed found that only 10% of the unique lncRNAs with a significant RIF-score (n = 240) were also differentially expressed, including three of the eight key hub lncRNAs. LncRNAs were significantly underrepresented in the list of DE loci across all tissues (X^2 test, p = 1.2E-06): while they accounted for 14.85% of all loci in the DE analyses, only 11.05% of the DE loci were classified as lncRNAs. In contrast, the other

TABLE 5 | Top 10 enriched pathways derived from genes and metabolites significantly correlated with key long non-coding RNAs associated with metabolic efficiency

ID	Ingenuity Canonical Pathways	log(p)	Ratio	Molecules
MSTRG.4740	tRNA Charging	5.56E00	8.54E-02	L-valine, L-phenylalanine, L-tryptophan, glycine, L-arginine, L-tyrosine, L-lysine
	EIF2 Signaling	4.13E00	3.83E-02	MYC, RPS7, RPL27A, RPL35, RPL23A, RPL37, RPL26, EIF3E, RPL31
	Glucose and Glucose-1-phosphate Degradation	3.18E00	1.3E-01	D-glucose, PGM3, phosphate
	Tyrosine Biosynthesis IV	2.94E00	2.86E-01	L-phenylalanine, L-tyrosine
	Acetyl-CoA Biosynthesis III (from Citrate)	2.82E00	2.5E-01	phosphate, citric acid
	Glycine Degradation (Creatine Biosynthesis)	2.71E00	2.22E-01	glycine, L-arginine
	Phenylalanine Degradation IV (Mammalian, via Side Chain)	2.68E00	8.82E-02	L-phenylalanine, phenylpyruvic acid, glycine
	Glutathione Biosynthesis	2.53E00	1.82E-01	phosphate, glycine
	Thymine Degradation	2.53E00	1.82E-01	5, 6-dihydrothymine, beta-ureidoisobutyric acid
MSTRG.10337	Calcium Signaling	1.63E01	9.35E-02	TNNT1, CHRNA1, CACNB1, CACNG1, CACNA1S, MYL2, TNNI2, TNNT3,T NNC2, TNNC1, MYL1, ATP2A1, CAMK2A, CASQ1, RYR1, TNNI1, CASQ2, MYL3, ACTA1, CAMK2B
	Protein Kinase A Signaling	7.45E00	3.88E-02	TNNI2, MYL2, MYLPF, MYLK2, PPP1R3A, TTN, MYL1, EPM2A, CAMK2A, PLCB1, RYR1, TNNI1, EYA1, MYL3, CAMK2B, PHKG1
	nNOS Signaling in Skeletal Muscle Cells	6.1E00	1.3E-01	CACNG1, CACNB1, CACNA1S, CHRNA1, RYR1, L-arginine
	Cellular Effects of Sildenafil (Viagra)	6.09E00	6.25E-02	CACNA1S, CACNG1, MYL2, MYLPF, PLCB1, L-arginine, MYL1, MYL3, ACTA1
	RhoA Signaling	4.55E00	5.6E-02	MYL2, MYLPF, MYLK2, TTN, MYL1, MYL3, ACTA1
	Apelin Cardiomyocyte Signaling Pathway	3.7E00	5.00E-02	MYL2, MYLPF, PLCB1, MYL3, MYL1, ATP2A1
	Actin Cytoskeleton Signaling	3.55E00	3.36E-02	MYL2, MYLPF, ACTN3, MYLK2, TTN, ACTA1, MYL3, MYL1
	Regulation of Actin-based Motility by Rho	3.24E00	5.21E-02	MYL2, MYLPF, MYL3, ACTA1, MYL1
	ILK Signaling	3.19E00	3.38E-02	PARVB, MYL2, TNFRSF1A, ACTN3, MYL1, MYL3, ACTA1
	Thrombin Signaling	2.93E00	3.06E-02	CAMK2A, MYL2, MYLPF, PLCB1, MYL1, MYL3, CAMK2B
MSTRG.17681	Caveolar-mediated Endocytosis Signaling	3.56E00	5.48E-02	ARCN1, COPA, COPE, COPB2
	Fatty Acid α-oxidation	2.29E00	8.00E-02	ALDH3A2, ALDH9A1
	Death Receptor Signaling	2.15E00	3.3E-02	PARP10, PARP4, HTRA2
	Histamine Degradation	2.05E00	6.06E-02	ALDH3A2, ALDH9A1
	Oxidative Ethanol Degradation III	2.05E00	6.06E-02	ALDH3A2, ALDH9A1
	G Protein Signaling Mediated by Tubby	2.03E00	5.88E-02	GNG2, GNAQ
	Tryptophan Degradation X (Mammalian, via Tryptamine)	2.00E00	5.71E-02	ALDH3A2, ALDH9A1
	Putrescine Degradation III	2.00E00	5.71E-02	ALDH3A2, ALDH9A1
	Ethanol Degradation IV	1.98E00	5.56E-02	ALDH3A2, ALDH9A1
	NER Pathway	1.96E00	2.8E-02	HIST2H4B, XAB2, RAD23B

loci accounted for 85.25% of all loci in the DE analyses, but had a share of 88.95% in the total of 2,154 differentially expressed unique loci.

In a recent publication, van Dam et al. (2017) reviewed and highlighted the usefulness of gene co-expression networks for the functional classification of genes and novel loci, such as non-coding elements without any known function. Correspondingly Oliveira et al. (2018) successfully applied a co-expression network concept to identify genes and miRNAs regulating IMF in Nellore steers. Besides the preselection of lncRNAs for co-expression networks, it might be advisable to make a knowledge-based preselection also for other genes to be included instead of simply using all expressed genes. The combination of RNA-Seq results with GWAS hits (gene

is an acknowledged procedure to integrate multiple layers of knowledge into a prioritized gene set for co-expression network analysis (Schaefer et al., 2018). In our PCIT analysis, we prioritized genes that appeared to be functionally important from the RNA-Seq analysis [DE loci (2,154) or TS loci (930)] and published GWAS data and selected those for our prioritized gene set to create a stronger focus on bovine metabolic efficiency, accepting however that still unknown, yet important elements might be overlooked. When preparing the prioritized gene set, we noted that the key role of liver in metabolic processes was clearly reflected by the by far highest number of DE loci (1,286) between the two metabolic

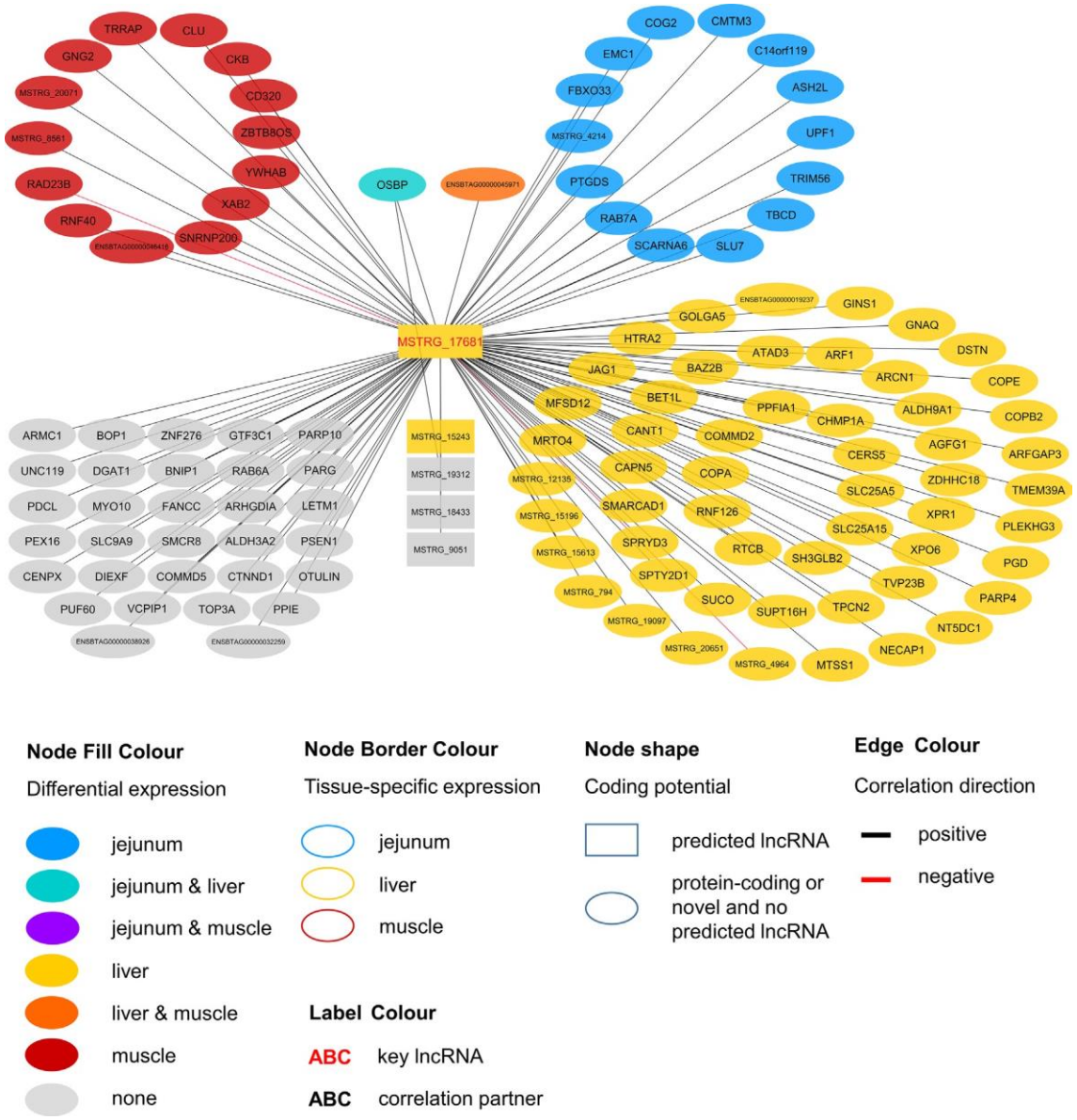

FIGURE 5 | Co-expression network for the novel long non-coding (lnc) RNA MSTRG.17681 with key regulatory potential for metabolic efficiency in cattle and significantly ($p < 0.05$) correlated genes with a minimal correlation coefficient of $|r| > 0.8$. Correlations are exclusive for animals with high metabolic efficiency.

efficiency groups, which was 2.6 fold higher than in jejunum or rumen. For DE loci in the prioritized gene set that was used for the PCIT, we noted that these predominantly (65%) had their highest expression in a different tissue than where they were differently expressed. This underlines that tissue specificity or tissue of highest abundance and DE of loci are indeed different, non-redundant features and that it is recommendable to follow a TS perspective in the beginning of the analysis.

One way to deduce a biological function of lncRNAs is to take a close look at coding genes in their immediate vicinity. This idea has also been implemented in the bioinformatics tool FEELnc for lncRNA prediction and annotation (Wucher et al., 2017), where the potential partner gene is generally assumed to be the closest annotated gene. However, this exclusively focusses on *in-cis* interaction with a narrow frame of impact. However, it has been reported that some lncRNAs execute *in-trans* regulatory tasks by binding directly to distant DNA sites or via RNAprotein interactions (Long et al., 2017) or a direct effect on RNA polymerase II activity (Kornienko et al., 2013).

Another way to infer functionality of unknown genomic elements subsequent to the network construction is to submit correlated coding genes to an enrichment analysis (Chen et al., 2018b), thereby assuming the guilt-by-association principle. Following this approach, we took genes from the prioritized

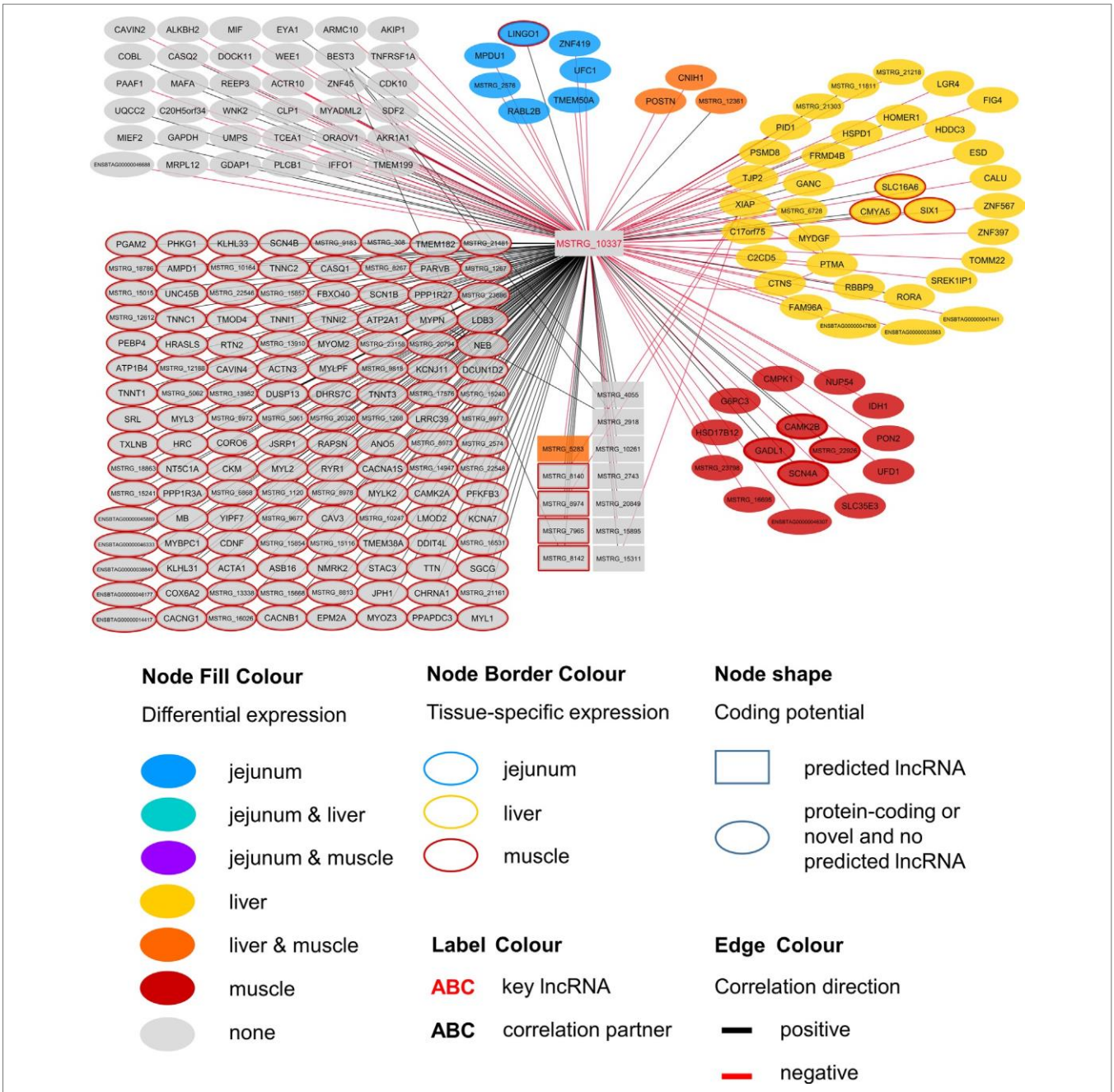

FIGURE 6 | Co-expression network for the novel long non-coding (lnc) RNA MSTRG.10337 with key regulatory potential for metabolic efficiency in cattle and significantly ($p < 0.05$) correlated genes with a minimal correlation coefficient of $|r| > 0.8$. Correlations are exclusive for animals with low metabolic efficiency.

gene set that were correlated with high connectivity lncRNAs of interest. LncRNA partner genes predicted by FEELnc could also be part of the prioritized gene set if they fell into one of the categories (DE, tissue-specificity, QTL-harboring). This was the case for 473 out of 2,741 unique predicted lncRNA interaction partner genes. Thus, 12.6% of the genes that were used as PCIT input (3,754) were very close to or overlapped with a lncRNA.

In addition, we aimed to add a supplementary layer of information to the pathway enrichment analysis and thereby to create further biological depth by using the option to integrate gene expression and metabolic profiles. In a single step this approach facilitates to predict a link between transcriptome activity, the direct functional readout of metabolic activity or physiological status and the functional analysis of lncRNAs. MSTRG.4740, e.g., correlated with plasma levels of 117 metabolites—valuable information that would otherwise be missing from the enrichment analysis. To our knowledge, we here present the first study that integrates metabolomics and transcriptomic data in an enrichment analysis to predict the functional role of lncRNAs

Across-Tissue Candidate Long Non-Coding RNAs for Metabolic Efficiency

LncRNAs were defined as hubs when they were connected to at least 100 other nodes in the high or low efficiency PCIT network. Three of the identified eight hub lncRNAs were exemplarily chosen for a more detailed description of their biological functionality predicted with IPA. These lncRNAs— namely MSTRG.4740, MSTRG.10337, and MSTRG.17681— were hubs in gene groups that showed enrichment for transfer RNA (tRNA) charging (p = 2.78E-06) and EIF2 signaling (p = 7.34E-05), calcium signaling (p = 4.98E-17) and nNOS signaling in skeletal muscle cells (p = 7.88E-07), and calveolar-mediated endocytosis signaling (p = 2.77E-04) and fatty acid oxidation (p = 5.13E-03), respectively.

For MSTRG.4740 an encompassing look at the enriched pathways clearly pointed towards amino acid metabolism and protein synthesis. This lncRNA was DE in liver (adjusted p-value (BH) = 9.13E-03, $\log_2$FC = 1.70) but displayed highest abundance (average FPKM) in jejunum (10.68) and rumen (8.41) and lowest in muscle (1.66) compared to liver (6.23). The DE status in liver suggested biological relevance there. However, the RIF analysis attributed a significant score to MSTRG.4740 in jejunum. The strongest enrichment was for tRNA charging (p = 2.78E-06), which describes the attachment of amino acids to a tRNA before incorporation into a growing polypeptide. According to IPA, the enrichment of this pathway was due to the correlation of MSTRG.4740 expression level with the blood plasma content of six essential or semi-essential amino acids (L-valine, L-phenylalanine, L-tryptophan, L-arginine, L-tyrosine, L-lysine). No non-essential amino acid showed a significant correlation with this lncRNA. The significantly correlated amino acids play integral roles as regulators of metabolism and key body functions, but cannot or only partially be synthesized by bovine animals themselves. Plasma concentration of essential amino acids depends on uptake from the diet, the balance between protein synthesis and degradation in peripheral tissues as well as on the efficiency of transport processes. The enrichment of the tRNA Charging pathway was not backed up by other components in addition to the indicated amino acids (e.g., charged tRNAs themselves). Thus, we restrict our conclusion and suggest that the lncRNA has a close relationship with (semi-) essential amino acid levels, but rather not to tRNA Charging per se. Widmann et al. (2015) reported no significant correlation between plasma amino acids and RFI at the onset of puberty in bulls in the same resource population. However, in the current study we employed adult animals.

Endogenous metabolism and also supply of amino acid have been demonstrated to limit growth or lactation in pigs, cattle and fish as reviewed by Hou et al. (2016). Furthermore, Doelman et al. (2015) showed that an abomasal infusion with essential amino acids leads to increased protein levels of eIF2α and eIF2Bε in the mammary gland in dairy cows. The authors proclaimed a direct link between the eIF2 factor, which is essential for eukaryotic translation initiation and milk protein yield. Interestingly, we found *eIF2Bε* to be DE [q-value (BH) = 0.022, log2FC = 0.204] in liver and to be one of the genes underlying the significant enrichment of the EIF2 Signaling pathway (p = 7.34E-05), which is tightly linked to protein synthesis. Genes encoding for ribosomal proteins of 40S (*RPS7*) or 60S subunits (e.g. *RPL26*, *RPL31*) were significantly correlated with MSTRG.4740, as well as the before mentioned *eIF2Bε*. EIF2 signaling and subsequently *EIF3E* are required for the correct initiation of mRNA translation (Kimball 1999; Walsh and Mohr, 2014).

Considering the presented correlations of MSTRG.4740 with other genes and plasma metabolites, this hub lncRNA seems to be an excellent example of a potential new key regulator in metabolic efficiency through the modulation of translational processes.

In contrast to MSTRG.4740 that seems to act on the broader forefront of translation, MSTRG.17681 appears to have a rather narrow and more targeted function. The first hit in pathway enrichment was calveolar-mediated endocytosis signaling (p = 2.77E-04). Four genes (*COPA*, *COPE*, *COPB2*, *ARCN1*) belonging to this pathway were highly correlated (|r| > 0.8) with this hub lncRNA. We observed significant DE in the liver of divergently efficient animals for MSTRG.17681 (q-value (BH) = 0.0050, $\log_2$FC = 0.766) as well as the respective quartet of genes. *COPA*, *COPE* and *COPB2* are transporters and *ARCN1* encodes the coatomer subunit of the coat protein I (COPI) complex (Tunnacliffe et al., 1996). All genes are allocated to a subunit in the cellular calveolar-mediated endocytosis signaling: the COPI vesicle, which plays a role in intracellular lipid transport (Popoff et al., 2011) and regulates lipid homeostasis (Beller et al., 2008). COPI-vesicle biogenesis is *ARF1*-dependent (Beck et al., 2009), which we found to be DE in liver and to be positively correlated with MSTRG.17681. The *Arf1 GTPase-activating protein* 3 (*ArfGAP3*) that subsequently allows the vesicle to fuse with a target membrane (Beck et al., 2009), was also correlated to MSTRG.17681 and DE in liver.

Considering that COPI-vesicles assist in lipid transport, it seems fitting that we found significant correlations between MSTRG.17681 expression and plasma levels of two saturated fatty acids: caprylate (p = 0.013, r = 0.357) and heptanoate (p = 0.047, r = 0.289). Caprylic acid supplementation in the diet of weaned piglets was observed to lead to a significant increase body weight gain (Marounek et al., 2004). MSTRG.17681 most likely acts predominantly in jejunum, liver, and rumen, where average expression was much higher (31.83, 25.26, and 18.74 FPKM, respectively) compared with the expression in skeletal muscle (3.36 FPKM). We infer that MSTRG.17681 is a key regulator in COPIvesicle functioning and thereby presumably affects lipid levels.

MSTRG.10337 was the third key hub lncRNA with a distinct prediction of biological function. In the network specific for animals of low metabolic efficiency, MSTRG.10337 was co-expressed with 39 genes that were DE in liver, 4 of which were also DE in muscle. Interestingly, the hub lncRNA MSTRG.10337 correlated with *RORA* (*RAR related orphan receptor A*), which was DE in liver. *RORA* is a transcriptional regulator of genes related to lipid metabolism, e.g. *APOA1*, *APOA5*, *APOC3*, and *PRAPRG* (Vu-Dac et al., 1997; Raspe et al., 2001; Sundvold and Lien, 2001; Lind et al., 2005). Although not meeting the threshold

for entering the PCIT network with respect to correlation to MSTRG.10337, we found *APOA1* to be DE in the liver, providing consistency in gene expression and biological interplay with regard to *RORA*. Previously, Krappmann et al. (2012) has attested an association of a *RORC* (*RAR Related Orphan Receptor C*) variant with milk yield, as well as milk fat and protein percentage in our SEGFAM resource population. Furthermore, Zhang et al. (2017) linked both nuclear receptors *RORA* and *RORC* to hepatic lipid and fatty acid metabolism as well as circadian rhythm pathways in a liver-specific depletion experiment in mice.

The most enriched pathways related to MSTRG.10337 are Calcium signaling (p = 4.98E-17) Protein Kinase A (PKA) signaling (p = 3.51E-08), and nNOS signaling in skeletal muscle cells (p = 7.88E-07). These data confirmed findings from an alternative previous network analysis in our resource population, where GWAS results for RFI and metabolomics profiles were merged for bulls in puberty. Widmann et al. (2015) also has identified Protein Kinase A (PKA) signaling and Nitric Oxide signaling to be significantly enriched pathways in IPA analyses. Calcium signaling, Protein Kinase A (PKA) signaling and nNOS signaling in skeletal muscle cells are in biological interplay. Protein kinases are in charge of nNOS phosphorylation on different serine residues and catalyze the hydroxylation of L-arginine (Fleming, 2008). In turn, L-arginine plasma levels were negatively correlated with expression levels of MSTRG.10337 (p=0.038, r=-0.323) in our study. This would fit an inhibitory role of MSTRG.10337 in metabolic efficiency, because of unfavorable effects of arginine depletion in the diet on milk protein synthesis in dairy cows (Tian et al., 2017). The inhibitory effect is underlined by numerous negative correlations of MSTRG.10337 to genes with DE in liver (e.g. *LGR4*, *FIG4*, *ESD*), muscle (e.g. *PON2*, *IDH1*, *NUP54*) and jejunum (e.g. *LINGO1*, *MPDU1*, *UFC1*), as well as QTL harboring genes (e.g. *GAPDH*, *MAFA*, *MYBPC1*), although the exact mode of operation is unclear. The supplementation of arginine has been reported to reduce body fat deposition, improve muscle gain and improve insulin sensitivity and the metabolic profile (Wu et al., 2009), and its availability in the organism is therefore particularly interesting for beef production. In chicken, L-arginine supplementation enhanced lean muscle growth (Castro et al., 2018). However, protein anabolic effects in muscle via dietary arginine supplementation are controversially discussed in other species (Tang et al., 2011). In addition to Calcium and PKA signaling, a third highly enriched pathway for MSTRG.10337 was nNOS signaling. In terms of gene expression, nNOS is not restricted to neuronal cells but is commonly expressed in skeletal muscle and certain vascular smooth muscle cells as well (Fleming 2008), where it is important for tissue integrity and contractile performance (Percival, 2011). After Ca^{2+}-activation, nNOS enzymes produce NO, which affects the autoregulation of blood flow, myocyte differentiation and glucose homeostasis in skeletal muscle cells (Stamler and Meissner, 2001). In a previous study we already suspected a relationship between NO signaling, arginine and growth in cattle (Widmann et al., 2013).

We assume that MSTRG.10337 influences the onset of nNOS activation, because of its correlation to calcium voltage-gated channel genes and *RYR1* (*ryanodine receptor 1*) that encodes a calcium release channel protein (Loy et al., 2011). Co-expression with a large number of muscle specific genes (e.g. *CACNG1, MYLK2, TNNT1, MYL2*) or genes that are DE in muscle (*CAMK2B*) related this hub lncRNA to PKA and nNOS signaling. It might thereby influence phosphorylation, degradation and availability of L-arginine in the muscle cells, but simultaneously perform some regulatory tasks in hepatic lipid metabolism.

CONCLUSIONS

In this study, we were able to identify novel lncRNAs with potential key regulatory function in metabolic efficiency in cattle. Although usually low expression levels of lncRNAs entail difficulties in DE and co-expression analyses, the careful setting of expression thresholds, the use of a-priori knowledge in gene prioritization and the integrated use of RIF metrics and PCIT based co-expression networks have proven to be a valid method for the identification of regulatory hub lncRNAs. The enrichment analysis based on metabolites and gene expression data provided valuable insight into the putative biological functions of yet uncharacterized lncRNAs.

We focused on phenotypic differences and looked at mechanisms or correlations that were exclusive to either metabolic efficiency group. Still, other correlations between lncRNAs and mRNAs might exist simultaneously in both groups, and we propose to take a group transcending approach in a follow-up study. For future work, we suggest to proceed within tissues to get a clearer picture of gene-gene interactions within a tissue, also because we noted that a multi-tissue approach presents its challenges when interpreting pathway enrichment results. The hub lncRNAs, which we identified, can be considered as candidates for further validation studies, *in vitro* or *in vivo*. Kashi et al. (2016) neatly described modern methods to determine where and how lncRNAs act in the cell or organism, such as chromatin isolation by RNA purification (ChIRP) sequencing (Chu et al., 2011).

In conclusion, our study demonstrates that the method we presented is suitable for the identification for key regulatory lncRNAs in a complex phenotype. By carefully adjusting different elements of the procedure, e.g. the tissue under consideration or the choice of priority categories for genes to include in the network analysis, this pipeline allows us to answer targeted biological questions.

DATA AVAILABILITY STATEMENT

The datasets generated for this study have been submitted to the "Functional Annotation of Animal Genomes" (FAANG) initiative database, accession PRJEB34570, and are also available via the European Nucleotide Archive (ENA).

ETHICS STATEMENT

The animal study was reviewed and approved by Animal care and experimental procedures following the guidelines of the German Law of Animal Protection. The protocols were approved by the Animal Protection Board of the Leibniz Institute for Farm Animal Biology as well as by the Animal Care Committee of the State Mecklenburg-Western Pomerania, Germany (State Office for Agriculture, Food Safety and Fishery; LALLF M-V/ Rostock, Germany, TSD/7221.3-2.1-010/03).

AUTHOR CONTRIBUTIONS

WN performed the statistical analyses and investigations, created the visualizations and wrote the original draft. RW and CK performed data collection, generated transcriptomic data, contributed to data analysis and conceptualized and administered the project and supervised WN. AR coded and performed bioinformatics analyses and supervised WN. RB, EA, and HH provided support with sampling and phenotyping of the test animals. All authors contributed to reviewing and editing the manuscript.

FUNDING

This study was funded by the German Research Foundation (DFG–grant numbers: KU 771/8-1 and WE 1786/5-1). WN received a scholarship for doctoral candidates from the German Academic Exchange Service (DAAD) and travel funds from the Graduate Academy of the University of Rostock. The publication of this article was funded by the Open Access Fund of the Leibniz Institute for Farm Animal Biology (FBN).

ACKNOWLEDGMENTS

The authors thank Frieder Hadlich for his support with bioinformatics obstacles of all kind, Marina Naval-Sanchez for her insightful ideas in network analysis, and Simone Wöhl and Bärbel Pletz for their excellent technical work in the lab.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fgene.2019.01130/full#supplementary-material

REFERENCES

Alexandre, P. A., Naval-Sanchez, M., Porto-Neto, L. R., Ferraz, J. B. S., Reverter, A., and Fukumasu, H. (2019). Systems biology reveals NR2F6 and TGFB1 as key regulators of feed efficiency in beef cattle. *Front. Genet.* 10, 230. doi: 10.3389/fgene.2019.00230

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrew, S. (2010). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Archer, J. A., Arthur, P. F., Herd, R. M., Parnell, P. F., and Pitchford, W. S. (1997). Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. *J. Anim. Sci.* 75, 2024–2032. doi: 10.2527/1997.7582024x

Arnes, L., Akerman, I., Balderes, D. A., Ferrer, J., and Sussel, L. (2016). Betalinc1 encodes a long noncoding RNA that regulates islet beta-cell formation and function. *Genes Dev.* 30, 502–507. doi: 10.1101/gad.273821.115

Beck, R., Ravet, M., Wieland, F. T., and Cassel, D. (2009). The COPI system: molecular mechanisms and function. *FEBS Lett.* 583, 2701–2709. doi: 10.1016/j.febslet.2009.07.032

Beller, M., Sztalryd, C., Southall, N., Bell, M., Jäckle, H., Auld, D. S., et al. (2008). COPI Complex is a regulator of lipid homeostasis. *PloS Biol.* 6, e292. doi: 10.1371/journal.pbio.0060292

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., et al. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* 50, 362–367. doi: 10.1038/s41588-018-0056-5

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J. B., et al. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542. doi: 10.1016/0092-8674(92)90520-M

Burkard, C., Opriessnig, T., Mileham, A. J., Stadejek, T., Ait-Ali, T., Lillico, S. G., et al. (2018). Pigs lacking the scavenger receptor cysteine-rich domain 5 of CD163 are resistant to porcine reproductive and respiratory syndrome virus 1 infection. *J. Virol.* 92, e00415–e00418. doi: 10.1128/JVI.00415-18

Canovas, A., Reverter, A., DeAtley, K. L., Ashley, R. L., Colgrave, M. L., Fortes, M. R. S., et al. (2014). Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PloS One* 9, 17. doi: 10.1371/ journal.pone.0102551

Cánovas, A., Reverter, A., Kasey, L., DeAtley, K. L., Ashley, R. L., Colgrave, M. L., et al. (2014). Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PloS One* 9, e102551. doi: 10.1371/journal. pone.0102551

Castro, F. L. S., Su, S., Choi, H., Koo, E., and Kim, W. K. (2018). L-Arginine supplementation enhances growth performance, lean muscle, and bone density but not fat in broiler chickens. *Poultry Sci.* 98, 1716–1722. doi: 10.3382/ps/ pey504

Chen, C., Cui, Q. M., Zhang, X., Luo, X., Liu, Y. Y., Zuo, J. B., et al. (2018a). Long noncoding RNAs regulation in adipogenesis and lipid metabolism: emerging insights in obesity. *Cell. Signalling* 51, 47–58. doi: 10.1016/j.cellsig.2018.07.012

Chen, W., Zhang, X., Li, J., Huang, S., Xiang, S., Hu, X., et al. (2018b). Comprehensive analysis of coding-lncRNA gene co-expression network uncovers conserved functional lncRNAs in zebrafish. *BMC Genomics* 19, 112. doi: 10.1186/s12864-018-4458-7

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* 132, 259–275. doi: 10.1083/ jcb.132.3.259

Core Team, R. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Csorba, T., Questa, J. I., Sun, Q., and Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci.* 111, 16160–16165. doi: 10.1073/pnas.1419030111 Degirmenci, U., Li, J., Lim, Y. C., Siang, D. T. C., Lin, S., Liang, H., et al. (2019). Silencing an insulin-induced lncRNA, LncASIR, impairs the transcriptional response to insulin signalling in adipocytes. *Sci. Rep.* 9, 5608. doi: 10.1038/ s41598-019-42162-5

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111

Doelman, J., Curtis, R. V., Carson, M., Kim, J. J. M., Metcalf, J. A., and Cant, J. P. (2015). Essential amino acid infusions stimulate mammary expression of eukaryotic initiation factor 2Bε but milk protein yield is not increased during an imbalance. *J. Dairy Sci.* 98, 4499–4508. doi: 10.3168/jds.2014-9051

Eberlein, A., Takasuga, A., Setoguchi, K., Pfuhl, R., Flisikowski, K., Fries, R., et al. (2009). 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* 183, 951–964. doi: 10.1534/genetics.109.106476

Fleming, Ingrid. (2008). Chapter 3 "Biology of Nitric Oxide Synthases," in *Microcirculation (Second Edition)*. Eds. Tuma, R. F., Durán, W. N., and Ley, K. (San Diego, USA: Academic Press), 56–80. doi: 10.1016/B978-0-12-374530-9.00003-6

Frankish, A., Vullo, A., Zadissa, A., Yates, A., Thormann, A., Parker, A., et al. (2017). Ensembl 2018. *Nucleic Acids Res.* 46, D754–DD61. doi:10.1093/nar/gkx1098

Haerty, W., and Ponting, C. P. (2015). Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. *RNA* 21, 333–346. doi: 10.1261/rna.047324.114

Hardie, L. C., VandeHaar, M. J., Tempelman, R. J., Weigel, K. A., Armentano, L. E., Wiggans, G. R., et al. (2017). The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. *J. Dairy Sci.* 100, 9061–9075. doi: 10.3168/jds.2017-12604

Harrell, and Frank, E. (2019). Hmisc: Harrell Miscellaneous, R-package version 4.2-0. https://cran.r-project.org/web/packages/Hmisc/index.html

Higgins, M. G., Fitzsimons, C., McClure, M. C., McKenna, C., Conroy, S., Kenny, D. A. et al. (2018). GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Sci. Rep.* 8, 14301–14301. doi: 10.1038/s41598-018-32374-6

Hou, Y. Q., Yao, K., Yin, Y. L., and Wu, G. Y. (2016). Endogenous synthesis of amino acids limits growth, lactation, and reproduction in animals. *Adv. Nutr.* 7, 331–342. doi: 10.3945/an.115.010850

Ibeagha-Awemu, E. M., Peters, S. O., Akwanji, K. A., Imumorin, I. G., and Zhao, X. (2016). High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci. Rep.* 6, 31109. doi: 10.1038/srep31109

Jin, J. J., Lv, W., Xia, P., Xu, Z. Y., Zheng, A. D., Wang, X. J., et al. (2018). Long noncoding RNA SYISL regulates myogenesis by interacting with polycomb repressive complex 2. *Proc. Natl. Acad. Sci.* 115, E9802–E9E11. doi: 10.1073/ pnas.1801471115

Kashi, K., Henderson, L., Bonetti, A., Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochim. Biophys. Acta Gene Regul*. *Mech*. 1859, 3–15. doi:10.1016/j.bbagrm.2015.10.010

Kenny, D. A., Fitzsimons, C., Waters, S. M., and McGee, M. (2018). Invited review: improving feed efficiency of beef cattle - the current state of the art and future challenges. *Animal* 12, 1815–1826. doi: 10.1017/S1751731118000976

Kern, C., Wang, Y., Chitwood, J., Korf, I., Delany, M., Cheng, H., et al. (2018). Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. *BMC Genomics* 19, 684. doi: 10.1186/s12864-018-5037-7

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357. doi: 10.1038/nmeth.3317

Kimball, S. R. (1999). Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.* 31, 25–29. doi: 10.1016/S1357-2725(98)00128-9

Kirchgeßner, M. (1997). Tierernährung. Frankfurt a.M., Germany: Verlags Union Agrar DLG-Verlag, 574.

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., and Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* 11, 59. doi: 10.1186/1741-7007-11-59

Kramer, A., Green, J., Pollard, J. Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703

Krappmann, K., Widmann, P., Weikard, R., and Kühn, C. (2012). Variants of the bovine retinoic acid receptor-related orphan receptor C gene are in linkage disequilibrium with QTL for milk production traits on chromosome 3 in a beef × dairy crossbreed population. *Arch. Anim. Breed.* 55, 346–355. doi: 10.5194/ aab-55-346-2012

Kühn, C., Bellmann, O., Voigt, J., Wegner, J., Guiard, V., and Ender, K. (2002). An experimental approach for studying the genetic and physiological background of nutrient transformation in cattle with respect to nutrient secretion and accretion type. *Arch. Anim. Breed.* 45, 14. doi: 10.5194/aab-45-317-2002

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Li, B. J., Jiang, D. L., Meng, Z. N., Zhang, Y., Zhu, Z. X., Lin, H. R., et al. (2018). Genome-wide identification and differentially expression analysis of lncRNAs in tilapia. *BMC Genomics* 19, 729. doi: 10.1186/s12864-018-5115-x

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078– 2079. doi: 10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., and Shi, W. (2014). Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lind, U., Nilsson, T., McPheat, J., Stromstedt, P. E., Bamberg, K., Balendran, C., et al. (2005). Identification of the human ApoAV gene as a novel RORalpha target gene. *Biochem. Biophys. Res. Commun.* 330, 233–241. doi: 10.1016/j.bbrc.2005.02.151

Liu, P., Jin, L., Zhao, L., Long, K., Song, Y., Tang, Q., et al. (2018). Identification of a novel antisense long non-coding RNA PLA2G16-AS that regulates the expression of PLA2G16 in pigs. *Gene*. 671, 78–84. doi: 10.1016/j. gene.2018.05.114

Long, Y., Wang, X., Youmans, D. T., and Cech, T. R. (2017). How do lncRNAs regulate transcription. *Sci. Adv.* 3, eaao2110. doi: 10.1126/sciadv.aao2110

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Loy, R. E., Orynbayev, M., Xu, L., Andronache, Z., Apostol, S., Zvaritch, E., et al. (2011). Muscle weakness in Ryr1I4895T/WT knock-in mice as a result of reduced ryanodine receptor Ca2+ ion permeation and release from the sarcoplasmic reticulum. *J. Gen. Physiol.* 137, 43–57. doi: 10.1085/jgp.201010523

Marounek, M., Skřivanová, E., and Skřivanová, V. (2004). A note on the effect of caprylic acid and triacylglycerols of caprylic and capric acid on growth rate and shedding of coccidia oocysts in weaned piglets. *J. Anim. Feed Sci.* 13, 269–274. doi: 10.22358/jafs/67411/2004

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

Medeiros de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., Miguel Ferreira de Camargo, G., Matos Ceacero, T., Noely dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-wide association study for carcass traits in an experimental nelore cattle population. *PloS One* 12, e0169860. doi: 10.1371/journal.pone.0169860

Miao, X., Luo, Q., Zhao, H., and Qin, X. (2016). Co-expression analysis and identification of fecundity-related long non-coding RNAs in sheep ovaries. *Sci. Rep.* 6, 39398. doi: 10.1038/srep39398

Nguyen, L. T., Reverter, A., Cánovas, A., Venus, B., Anderson, S. T., Islas-Trejo, A., et al. (2018). STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in brahman heifers. *Front. Genet.* 9, 87. doi: 10.3389/fgene.2018.00087

Oliveira, G. B., Regitano, L. C. A., Cesar, A. S. M., Reecy, J. M., Degaki, K. Y., Poleti, M. D., et al. (2018). Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics* 19, 126. doi: 10.1186/s12864-018-4514-3

Park, C. A., Reecy, J. M., and Hu, Z.-L. (2018). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–DD10. doi: 10.1093/ nar/gky1084

Pellegrina, D.V.da Silva, Severino, P., Barbeiro, H. V., de Souza, H. P., Machado, M. C. C., Pinheiro-da-Silva, F., et al. (2017). Insights into the function of long noncoding rnas in sepsis revealed by gene co-expression network analysis. *Non-Coding RNA* 3, 5. doi: 10.3390/ncrna3010005

Percival, J. M. (2011). nNOS regulation of skeletal muscle fatigue and exercise performance. *Biophys. Rev.* 3, 209–217. doi: 10.1007/s12551-011-0060-9

Perez-Montarelo, D., Hudson, N. J., Fernandez, A. I., Ramayo-Caldas, Y., Dalrymple, B. P., and Reverter, A. (2012). Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. *PloS One* 7, e46159. doi: 10.1371/journal.pone.0046159

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Popoff, V., Adolf, F., Brugger, B., and Wieland, F. (2011). COPI budding within the Golgi stack, Cold Spring Harbor. *Perspect. Biol.* 3, a005231. doi: 10.1101/ cshperspect.a005231

Raspe, E., Duez, H., Gervois, P., Fievet, C., Fruchart, J. C., Besnard, S., et al. (2001). Transcriptional regulation of apolipoprotein C-III gene expression by the orphan nuclear receptor RORalpha. *J. Biol. Chem.* 276, 2865–2871. doi: 10.1074/jbc.M004982200

Reverter, A., and Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. doi: 10.1093/bioinformatics/btn482

Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M., and Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26, 896–904. doi: 10.1093/ bioinformatics/btq051

Robinson, A. (2015). Quality Trim version 1.6.0. https://bitbucket.org/arobinson/ qualitytrim/src/master/

Schaefer, R. J., Michno, J. M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., et al. (2018). Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell* 30, 2922–2942. doi: 10.1105/tpc.18.00299

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* 18, 386–386. doi: 10.1186/ s12864-017-3754-y

Serviss, J. T., Johnsson, P., and Grandér, D. (2014). An emerging role for long non-coding RNAs in cancer metastasis. *Front. Genet.* 5, 234. doi: 10.3389/ fgene.2014.00234

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/ gr.1239303

Stamler, J. S., and Meissner, G. (2001). Physiology of nitric oxide in skeletal muscle. *Physiol. Rev.* 81, 209–237. doi: 10.1152/physrev.2001.81.1.209

Sui, Yutong, Han, Yu, Zhao, Xingyu, Li, Dongsong, and Li, Guangyu (2019). Long non-coding RNA Irm enhances myogenic differentiation by interacting with MEF2D. *Cell Death Dis.* 10, 181–181. doi: 10.1038/s41419-019-1399-2

Sun, H. Z., Zhao, K., Zhou, M., Chen, Y., and Guan, L. L. (2019). Landscape of multi-tissue global gene expression reveals the regulatory signatures of feed efficiency in beef cattle. *Bioinformatics* 35, 1712–1719. doi: 10.1093/ bioinformatics/bty883 (Oxford, England).

Sundvold, H., and Lien, S. (2001). Identification of a novel peroxisome proliferatoractivated receptor (PPAR) gamma promoter in man and transactivation by the nuclear receptor RORalpha1. *Biochem. Biophys. Res. Commun.* 287, 383–390. doi: 10.1006/bbrc.2001.5602

Tang, J. E., Lysecki, P. J., Manolakos, J. J., MacDonald, M. J., Tarnopolsky, M. A., and Phillips, S. M. (2011). Bolus arginine supplementation affects neither muscle blood flow nor muscle protein synthesis in young men at rest or after resistance exercise. *J. Nutr.* 141, 195–200. doi: 10.3945/jn.110.130138

Tang, Z., Wu, Y., Yang, Y., Yang, Y.-C. T., Wang, Z., Yuan, J., et al. (2017). Comprehensive analysis of long non-coding RNAs highlights their spatiotemporal expression patterns and evolutional conservation in Sus scrofa. *Sci. Rep.* 7, 43166. doi: 10.1038/srep43166

Thornton, P. K. (2010). Livestock production: recent trends, future prospects, Philosophical Transactions of the Royal Society of London. *Ser. B Biol. Sci.* 365, 2853–2867. doi: 10.1098/rstb.2010.0134

Tian, W., Wang, H. R., Wu, T. Y., Ding, L. Y., Zhao, R., Khas, E., et al. (2017). Milk protein responses to balanced amino acid and removal of Leucine and Arginine supplied from jugular-infused amino acid mixture in lactating dairy cows. *J. Anim. Physiol. Anim. Nutr.* 101, e278–ee87. doi: 10.1111/jpn.12603 Tunnacliffe, A., Pensotti, V., and Radice, P. (1996). The coatomer protein deltaCOP, encoded by the archain gene, is conserved across diverse eukaryotes. *Mamm. Genome* 7, 784–786. doi: 10.1007/s003359900234

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell*. 154, 26–46. doi: 10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., and Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537–1550. doi: 10.1016/j.cell.2011.11.055

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings Bioinf.* 19, 575–592. doi: 10.1093/bib/bbw139

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263. doi: 10.1038/nrg2538

Vu-Dac, N., Gervois, P., Grotzinger, T., De Vos, P., Schoonjans, K., Fruchart, J. C., et al. (1997). Transcriptional regulation of apolipoprotein A-I gene expression by the nuclear receptor RORalpha. *J. Biol. Chem.* 272, 22401–22404. doi: 10.1074/jbc.272.36.22401

Walsh, D., and Mohr, I. (2014). Coupling 40S ribosome recruitment to modification of a cap-binding initiation factor by eIF3 subunit e. *Genes Dev.* 28, 835–840. doi: 10.1101/gad.236752.113

Wang, C.-H., Shi, H.-H., Chen, L.-H., Li, X.-L., Cao, G.-L., and Hu, X.-F. (2019). Identification of key lncRNAs associated with atherosclerosis progression based on public datasets. *Front. Genet.* 10, 123. doi: 10.3389/fgene.2019.00123

Weikard, R., Altmaier, E., Suhre, K., Weinberger, K. M., Hammon, H. M., Albrecht, E., et al. (2010). Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. *Physiol. Genomics* 42a, 79–88. doi: 10.1152/ physiolgenomics.00120.2010

Weikard, R., Goldammer, T., Brunner, R. M., and Kuehn, C. (2012). Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. *Physiol. Genomics* 44, 728–739. doi: 10.1152/physiolgenomics.00007.2012

Weikard, R., Hadlich, F., Hammon, H. M., Frieten, D., Gerbert, C., Koch, C., et al. (2018). Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. *Oncotarget* 9, 21052–21069. doi: 10.18632/oncotarget.24898

Widmann, P., Nuernberg, K., Kuehn, C., and Weikard, R. (2011). Association of an ACSL1 gene variant with polyunsaturated fatty acids in bovine skeletal muscle. *BMC Genet.* 12, 13. doi: 10.1186/1471-2156-12-96

Widmann, P., Reverter, A., Fortes, M. R. S., Weikard, R., Suhre, K., Hammon, H., et al. (2013). A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics* 14, 798. doi: 10.1186/1471-2164-14-798

Widmann, P., Reverter, A., Weikard, R., Suhre, K., Hammon, H. M., Albrecht, E., et al. (2015). Systems biology analysis merging phenotype, metabolomic and genomic data identifies non-smc condensin i complex, subunit G (NCAPG) and cellular maintenance processes as major contributors to genetic variability in bovine feed efficiency. *PloS One* 10, 22. doi: 10.1371/journal. pone.0124574

Wu, G., Bazer, F. W., Davis, T. A., Kim, S. W., Li, P., Marc Rhoads, J., et al. (2009). Arginine metabolism and nutrition in growth, health and disease. *Amino Acids* 37, 153–168. doi: 10.1007/s00726-008-0210-y

Wucher, V., Legeai, F., Hedan, B., Rizk, G., Lagoutte, L., Leeb, T., et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45, e57. doi: 10.1093/nar/gkw1306

Yang, L., Li, P., Yang, W., Ruan, X., Kiesewetter, K., Zhu, J., et al. (2016). Integrative transcriptome analyses of metabolic responses in mice define pivotal LncRNA metabolic regulators. *Cell Metab.* 24, 627–639. doi: 10.1016/j.cmet.2016.08.019

Zeng, Y., Ren, K., Zhu, X., Zheng, Z., and Yi, G. (2018). "Chapter one long noncoding RNAs," in *Advances in Lipid Metabolism*. Ed. Gregory, S. (Makowski: Advances in Clinical Chemistry (Elsevier)). doi: 10.1016/bs.acc.2018.07.001

Zhang, Y., Papazyan, R., Damle, M., Fang, B., Jager, J., Feng, D., et al. (2017). The hepatic circadian clock fine-tunes the lipogenic response to feeding through RORalpha/gamma. *Genes Dev.* 31, 1202–1211. doi: 10.1101/gad.302323.117

Zhu, M., Liu, J. F., Xiao, J., Yang, L., Cai, M. X., Shen, H. Y., et al. (2017). Lnc-mg is a long non-coding RNA that promotes myogenesis. *Nat. Commun.* 8, 11. doi: 10.1038/ncomms14718

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

9.2. Identification and Annotation of Potential Function of Regulatory Antisense Long Non-Coding RNAs Related to Feed Efficiency in *Bos taurus* Bulls.

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A und Kühn C (2020) in *International Journal of Molecular Sciences* 21(9). doi: 10.3390/ijms21093292.

International Journal of *Molecular Sciences*

MDPI

Article

Identification and Annotation of Potential Function of Regulatory Antisense Long Non-Coding RNAs Related to Feed Efficiency in *Bos taurus* Bulls

Wietje Nolte[1], Rosemarie Weikard [1], Ronald M. Brunner [1], Elke Albrecht [2], Harald M. Hammon [3], Antonio Reverter [4] and Christa Kühn [1,5,*]

[1] Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; nolte@fbn-dummerstorf.de (W.N.); weikard@fbn-dummerstorf.de (R.W.); brunner@fbn-dummerstorf.de (R.M.B.)
[2] Institute of Muscle Biology and Growth, Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; albrecht@fbn-dummerstorf.de
[3] Institute of Nutritional Physiology "Oskar Kellner", Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; hammon@fbn-dummerstorf.de
[4] Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food, Queensland Bioscience Precinct, St Lucia 4067 QLD, Australia; toni.reverter-gomez@csiro.au
[5] Faculty of Agricultural and Environmental Sciences, University Rostock, 18059 Rostock, Germany
* Correspondence: kuehn@fbn-dummerstorf.de

Received: 14 April 2020; Accepted: 4 May 2020; Published: 6 May 2020

Abstract: Long non-coding RNAs (lncRNAs) can influence transcriptional and translational processes in mammalian cells and are associated with various developmental, physiological and phenotypic conditions. However, they remain poorly understood and annotated in livestock species. We combined phenotypic, metabolomics and liver transcriptomic data of bulls divergent for residual feed intake (RFI) and fat accretion. Based on a project-specific transcriptome annotation for the bovine reference genome ARS-UCD.1.2 and multiple-tissue total RNA sequencing data, we predicted 3590 loci to be lncRNAs. To identify lncRNAs with potential regulatory influence on phenotype and gene expression, we applied the regulatory impact factor algorithm on a functionally prioritized set of loci ($n = 4666$). Applying the algorithm of partial correlation and information theory, significant and independent pairwise correlations were calculated and co-expression networks were established, including plasma metabolites correlated with lncRNAs. The network hub lncRNAs were assessed for potential *cis*-actions and subjected to biological pathway enrichment analyses. Our results reveal a prevalence of antisense lncRNAs positively correlated with adjacent protein-coding genes and suggest their participation in mitochondrial function, acute phase response signalling, TCA-cycle, fatty acid β-oxidation and presumably gluconeogenesis. These antisense lncRNAs indicate a stabilizing function for their *cis*-correlated genes and a putative regulatory role in gene expression.

Keywords: *Bos taurus*; feed efficiency; co-expression network analysis; lncRNA; Functional Annotation of Animal Genomes (FAANG)

1. Introduction

While the functionality of protein-coding genes has been thoroughly explored and scrutinized in the past century—and continues to be—the so-called 'dark matter of the genome' has shifted into focus in the recent decades. Non-coding elements are estimated to cover about 98% of the mammalian genome and to comprise different elements such as microRNAs, small nuclear RNAs, small nucleolar RNAs, transfer RNAs (miRNA, snRNA, snoRNA, tRNA, Encode Project Consortium [1]), the previously discovered circular RNAs (circRNA [2]), as well as long non-coding RNAs (lncRNAs).

In cattle breeding and production, the efficient use of feed by the animal is continuously gaining importance for ecological and economic reasons. In the beef industry, residual feed intake (RFI) as a measure of feed efficiency is usually the measure of choice [3]. Numerous association studies have

aimed to find causative genomic regions and gene variants that drive bovine feed efficiency, but repeatedly quantitative trait locus (QTL) peaks fall outside the protein-coding genes, e.g., [4–7]. This plethora of work suggests that the functional tasks of the non-coding elements affecting feed efficiency need to be addressed in greater detail.

To date, functional annotation of lncRNAs is still not fully comprehensive in human and model animals and even less so in livestock species, although first advances are in progress. Known modes of action of lncRNAs include chromatin-remodelling and chromatin state maintenance, and transcriptional enhancement or repression, e.g., through the binding to transcriptional regulatory factors as reviewed by Long et al. [8] and Marchese et al. [9].

Increasing evidence has shown that lncRNAs are involved in a broad range of pathological and disease conditions and environmental transitions but also in the general regulation of immune and metabolic processes in normal cell and tissue homeostasis, e.g., by acting as signal molecules that mark the regulation of developmental and physiological stages and gene expression. Lu et al. [10] summarized results indicating that lncRNAs are able to reprogram glucose and lipid metabolism in tumor cells by modulating key enzymes of glycolysis, oxidative phosphorylation and pentose phosphate as well as lipid synthesis and degradation pathways. A recent comprehensive overview of lncRNAs involved in lipid metabolism [11] elucidated lncRNAs that are potentially associated with hepatic lipid and glucose metabolism and related to metabolic disorders, such as obesity, cardiovascular diseases and hepatic steatosis. In murine liver, Yang et al. [12] found a lncRNA with a pivotal effect on lipogenesis, which was documented to act through a negative feedback loop relationship with a transcription factor coding gene (*SREBP1c*). Recently, Pradas-Juni et al. [13] identified a transcription factor MAFG-lncRNA (obesity-repressed lincIRS2) axis controlling hepatic glucose metabolism in health and metabolic disease.

LncRNAs are generally categorized as genic or intergenic RNA classes, which can be transcribed as sense- or antisense-oriented with respect to their nearest neighbouring protein-coding gene. Antisense lncRNAs, originating from the complementary strand of protein-encoding genes, comprise a major proportion of lncRNAs in the transcriptome across species, e.g., [14–16]. They commonly link neighbouring or overlapping genes in complex loci into chains of transcriptional units [15]. The genomic arrangement of antisense RNAs and opposite sense genes suggests that they might be part of self-regulating circuits that allow afflicted genes to regulate their own expression [16]. Antisense lncRNAs can act in *cis* as stabilizers [17], thereby increasing the abundance of the respective transcripts and protein of the protected gene [18]. Facilitated through the introduction of stranded library protocols in the 2000s, many genes have been found to overlap with antisense non-coding genomic elements, so-called natural antisense transcripts (NATs).

Although there are examples of lncRNAs with high sequence conservation across mammals, e.g., *MALAT1* [19], there is also evidence for a high level of sequence species-specificity in this RNA class compared with protein-coding genes [20]. For this reason, the identification and characterization of phenotype-influencing lncRNAs in the respective target species and tissue are advisable and one of the declared goals of the global initiative for Functional Annotation of Animal Genomes (FAANG, www.animalgenome.org/community/FAANG/).

While there are a variety of bioinformatics tools at hand for the prediction of long non-coding sequences from transcriptomic data (e.g., PLEK [21], FEELnc [22], PLAR [23], CPC [24], CPAT [25], and CNCI [26]), the functional annotation of novel non-coding loci remains challenging. Common practice is to construct co-expression networks and use correlation partners of lncRNAs for gene and pathway enrichment analyses. This guilt-by-association approach has been applied to non-coding elements, such as miRNAs [27–29] and lncRNAs [30–33].

In a previous study, we applied a combination of the regulatory impact factor (RIF [34]) and a partial correlation and information theory (PCIT [35]) to build correlation networks to predict key regulatory lncRNAs with an implication in metabolic efficiency in crossbred cattle [36]. We integrated phenotypic data, plasma metabolite profiles and transcriptomic data from four tissues (jejunum, liver, skeletal muscle, rumen) and two sexes. However, at that stage, little attention was given to the tissue-specificity of expression data [37] and likely molecular function of lncRNAs. Due to its central role in metabolic processes [38], the liver has repeatedly been the subject of transcriptomic studies, also with regard to bovine feed efficiency [39–43]. Therefore, in the present study, we adapted our analysis

pipeline to the new bovine genome ARS-UCD.1.2 and to a single-tissue approach, where we aim to identify liver lncRNAs with high regulatory potential and a functional relation to feed efficiency.

2. Results

2.1. Alignment and Mapping of RNA Sequencing Data

After quality trimming, the average sequencing depth of the RNA-sequencing libraries was 49.8 million read pairs per sample and the average alignment rate to the reference genome ARS-UCD.1.2 was 98.72% ± 0.26% (Table 1). The mapping of fragments to the project specific merged annotation (Supplement 1), which contained 30,806 loci and 82,628 transcripts after quality filtering, resulted in an average mapping rate of 85.98% ± 1.40%.

Table 1. RNA sequencing, alignment, and mapping statistics.

	Sequencing Depth [Read Pairs]	Alignment to ARS-UCD.1.2 (%)	Mapping to Project-Specific Annotation (%)
Mean	49,831,770	98.72	85.98
SD	5,588,004	0.26	1.40

SD = standard deviation.

2.2. Long Non-Coding RNA Prediction

The identification of lncRNAs with FEELnc based on the project-specific merged annotation and the bovine reference genome and annotation yielded a total of 6161 non-coding transcripts (3,590 loci) with a minimal length of 200 nt (results from FEELnc and differential expression analysis and information on structure and position for each transcript are listed in Supplement 2).

Within the default window size (10,000 to 100,000 nt), a total of 19,184 interactions of 3,495 lncRNA loci (out of 3590 loci) with positional partner genes were predicted, while 95 lncRNA loci (corresponding to 202 of the 6161 lncRNA transcripts) remained without a potential positional interaction partner locus. Out of the 3,495 loci with a predicted interaction partner, 1799 lncRNAs were in the sense direction to the predicted partner gene and 1,696 lncRNAs were in the antisense direction to their partner gene. The majority of the lncRNAs with an interaction partner assigned (2955) were classified as genic, meaning that they overlapped with their predicted partner gene in the sense or antisense orientation, and 540 lncRNAs were intergenic. The overall average expression level of the 3590 lncRNA loci was 10.13 FPKM (± 325.21) with a median of 0.26 FPKM.

In a locus-based approach, where we considered the transcript with the highest exon number for each lncRNA locus, we observed that strandedness was equally distributed among the 3590 loci (50.84% on the plus and 49.16% on the minus strand). The average number of exons per locus amounted to 4.52 ± 7.14 (media*n* = 3.00) and the geometric mean of the total exonic length was 1,723.78 nt.

2.3. Differential Metabolite Abundance

Between the groups of high and low efficiency bulls, we found 45 plasma metabolites to be significantly differentially abundant (q (Benjamini–Hochberg) ≤ 0.05 and absolute log-transformed foldchange ($\log_2$FC) ≥ 1). Eighteen metabolites were downregulated, i.e., lower in abundance, in the high efficiency group and 27 were upregulated. The most pronounced differences were found for leukotriene B4 ($q = 6.65 \times 10^{-4}$; $\log_2$FC = 2.40) and isovalerate ($q = 6.65 \times 10^{-4}$; $\log_2$FC = 1.87), which were significantly higher in abundance in highly efficient bulls compared with the low efficiency group (see Figure 1, Supplement 3). The strongest downregulation in the high efficiency group was observed for asparagine ($q = 1.51 \times 10^{-3}$; $\log_2$FC = −3.07) and methionine ($q = 1.83 \times 10^{-3}$; $\log_2$FC = −2.27). Next to these two amino acids (AAs), the AAs glutamine and cysteine were also differentially abundant (q =1.49 × 10^{-3}; $\log_2$FC = −1.71 and $q = 8.75 \times 10^{-3}$; $\log_2$FC = −1.12, respectively).

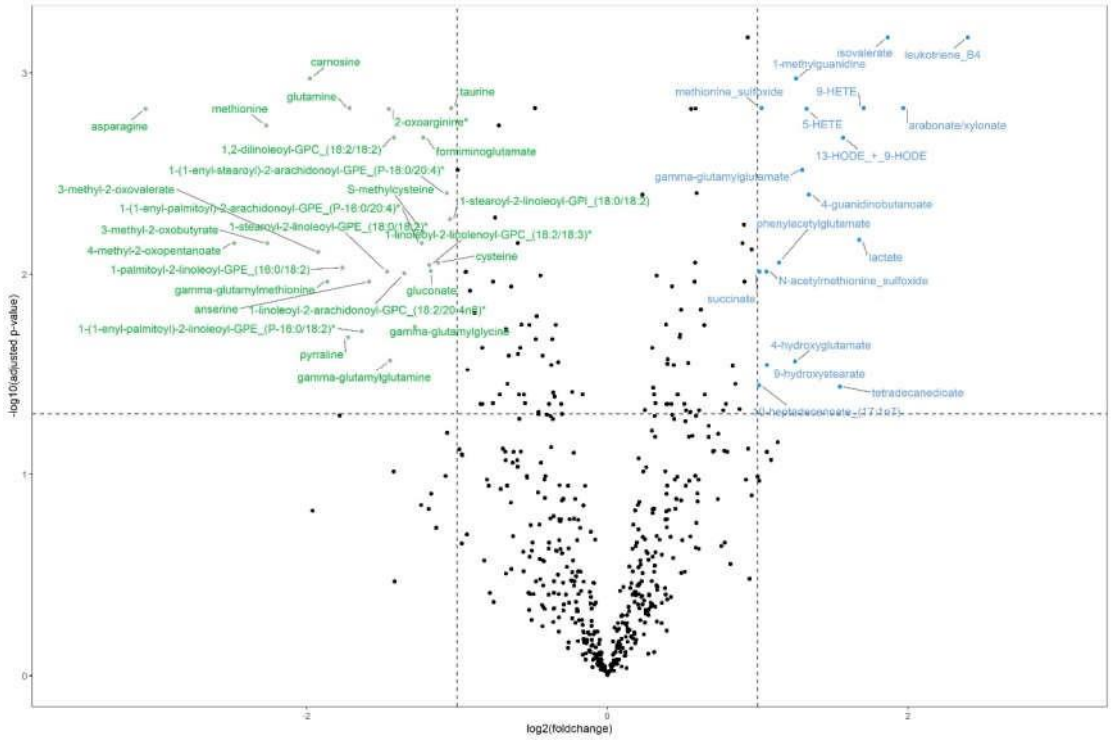

Figure 1. Volcano plot of differentially abundant plasma metabolites for bulls of high (n = 12) and low (n = 13) feed efficiency with upregulation (higher abundance) in highly efficient bulls with blue labels and downregulation (lower abundance) with green labels. Significance threshold (horizontal dotted line) at q (Benjamini-Hochberg) ≤ 0.05 and absolute $\log_2$(foldchange) ≥ 1 (vertical dotted lines).

Plotting of a metabolite based principal component analysis (PCA) showed a clear separation of the bulls in the two efficiency groups (see Figure 2), with the first two components accounting for 38% of the variance in the metabolite abundance.

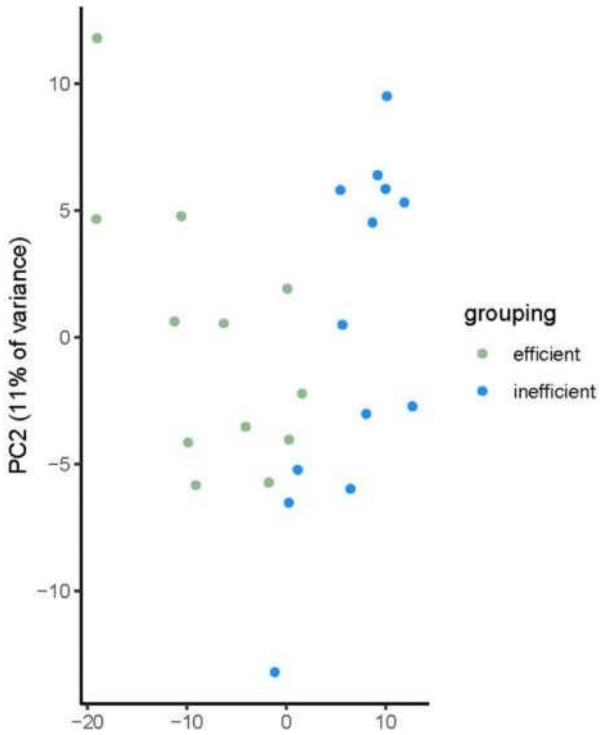

Figure 2. Principal component analysis (PCA) plot for 25 bulls divergent for feed efficiency. Plotting based on plasma metabolite levels (n = 552).

2.4. Set of Prioritized Loci for Co-Expression Network

The AnimalQTL database listed 1573 QTL for RFI that stemmed from SNP array-based studies (manual curation of the complete dataset) and could be remapped to the new bovine reference genome ARS-UCD.1.2. Out of these 1573 QTLs, 1506 had a direct overlap with a locus in our projectspecific merged annotation and no QTL was more than 3 Mb away from the next annotated locus. Finally, 843 of these loci passed the minimal expression threshold and were categorized as QTL locus.

Out of the 745 loci that were significantly differentially expressed (DE) with q (Benjamini–Hochberg) ≤ 0.1 between the high and the low efficiency group, 219 were predicted to be lncRNAs (29.4%), and 84 out of the 843 QTL loci were also predicted to be lncRNAs (10.0%).

In the end, the prioritized loci for the RIF and PCIT analyses contained a total of 4,666 unique loci, including 745 DE loci, 2083 lncRNAs, 2007 protein-coding partner gene loci, and 843 QTL loci (see Figure 3 and Supplement 4). Loci included in the prioritized loci set had to be minimally expressed (>0.1 FPKM in at least six animals of one efficiency group) and could fall into more than one category of the set.

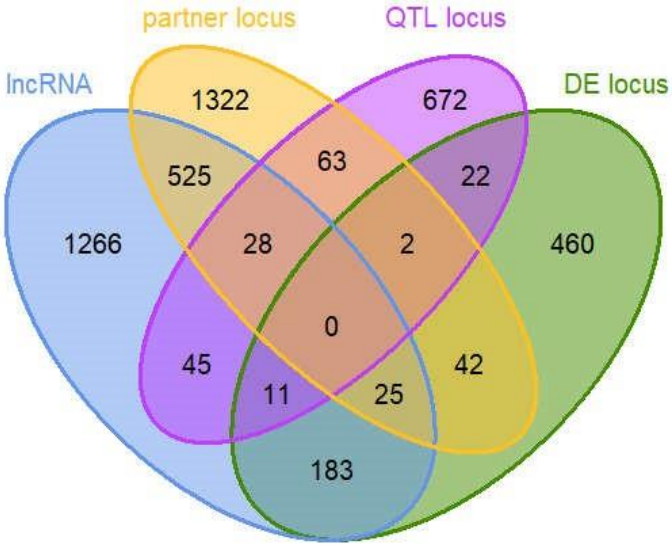

Figure 3. Venn diagram of 4666 loci in a prioritized loci set for co-expression network analysis: loci predicted to be lncRNAs (lncRNA) and their potential positional interaction gene partners (partner locus), loci overlapping with or no farther away than 3 Mb from a quantitative trait locus (QTL) for residual feed intake in cattle (QTL locus), and loci with differential expression (DE locus) between bulls of high and low feed efficiency.

2.5. Regulatory Impact Factor Analysis

Ultimately, 2083 lncRNAs and 3400 unique target loci (loci in the categories partner gene, QTL or DE locus) were submitted to the RIF analysis. In some cases, lncRNAs could be both potential regulators as well as target loci, hence the higher number of target loci. After z-transformation and filtering for lncRNAs with an absolute RIF1 or RIF2 score of ≥ 1.96, 238 lncRNAs were found to be significant and therefore potential key regulators in this dataset. As the two RIF metrics are designed to detect different mechanisms of regulation, the 238 key lncRNAs typically score either very high or very low in RIF1 or RIF2 and have a score around zero in the other metric, which results in a bimodal distribution of accumulated RIF scores (see Figure 4).

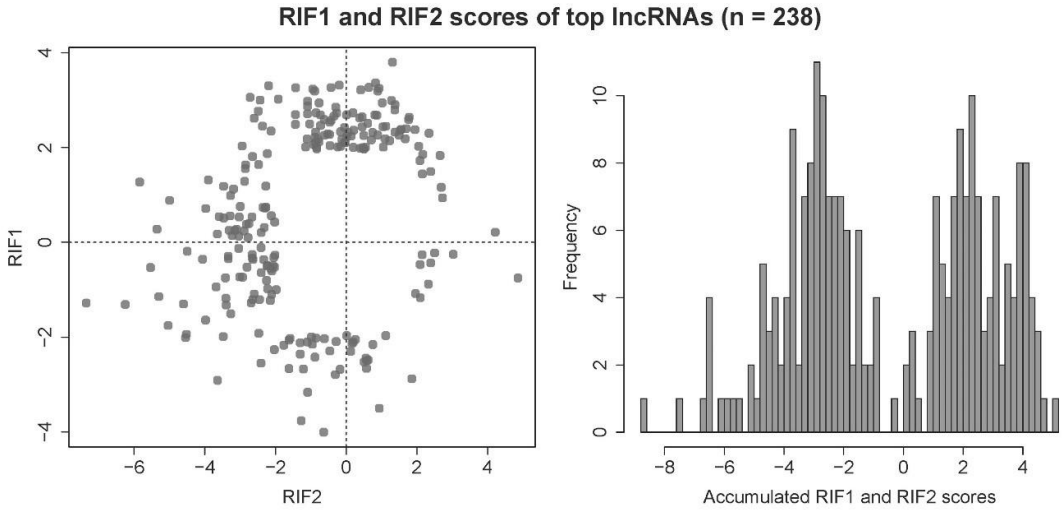

Figure 4. Distribution of scores of the metrics RIF1 and RIF2 from the regulatory impact factor (RIF) analysis for the top potential key regulatory lncRNAs, equalling 238 out of 2083 lncRNAs in the prioritized dataset (absolute z-transformed RIF1 or RIF2 ≥ 1.96).

2.6. Co-Expression Networks Based on Partial Correlation and Information Theory Approach and Detection of Hub LncRNAs

The prioritized loci set (4666 loci) that was used for the RIF analysis was subsequently also submitted to the PCIT algorithm and results were filtered for significant pairwise correlations with a minimal strength of $|r| \geq 0.65$, where one correlation partner had to be a lncRNA with a significant RIF score. This resulted in a total of 16,489 connections including 2299 out of the 4666 loci. After including significant ($p \leq 0.01$) correlations between key lncRNAs and plasma metabolites ($|r| \geq 0.65$) the co-expression networks comprised 2414 nodes with 16,709 edges. With 15,783 edges (94.46%), the vast majority of correlations were positive and only 926 correlations (5.54%) were negative.

Out of the 238 lncRNAs with a significant RIF score in the network, 22 were also categorized as a DE or QTL locus (see Supplement 4). A total of 17 lncRNAs had a network connection with at least 10 annotated genes with an official gene symbol in the bovine genome annotation (Supplement 5). In order to also account for regulatory lncRNAs with high metabolite or exceptionally high gene connectivity, the following additional lncRNAs were selected: five lncRNAs that were correlated with over ten annotated genes and over ten metabolites, and five lncRNAs that showed a connectivity with more than 50 annotated genes. One lncRNA passed both filtering steps (Supplement 5). Finally, 26 hub lncRNAs remained that were of interest regarding their associated interacting networks. These lncRNAs are candidates that probably have a regulatory potential for modulating biological pathways linked to divergent feed efficiency. One of these hub lncRNAs (*MSTRG.16058*) was connected with 14 RNA genes (including snRNAs and snoRNAs), which had escaped filtering. Due to its clear involvement in ribosomal RNA expression and unsuccessful mapping of these genes in Ingenuity Pathway Analysis (IPA), this lncRNA was excluded from further analyses.

2.7. Natural Antisense Transcripts

Out of the 238 lncRNAs with a significant RIF score (key lncRNAs), 237 had a predicted positional interaction partner locus in the FEELnc results. Thereof, 119 lncRNA loci (50%) were in antisense orientation and overlapped with an annotated locus that was termed as the most likely interaction partner (isbest score = 1 according to FEELnc). These lncRNAs can be designated as natural antisense transcripts (NATs). Of these 119 antisense lncRNA–sense partner locus pairs, 44 (18.49%) had a significant correlation in the PCIT analysis (Supplement 6). The vast majority (42 out of 44) were positive correlations and two pairs were negatively correlated. Negative correlations were found for the lncRNA *MSTRG.13915* and *AZGP1* (*Zinc-alpha-2-glycoprotein*, $r = -0.67$) and the lncRNA *MSTRG.5787* and *EPRS* (*Glutamyl-Prolyl-TRNA Synthetase 1*, $r = -0.51$). A total of eight lncRNApartner locus pairs were found to have a positive co-expression with $r > 0.9$. The pairs *MSTRG.5042* and *APOA1* (*Apolipoprotein A*, $r = 0.98$) and *MSTRG.7472* and *HP* (*Haptoglobin*, $r = 0.97$) displayed the strongest correlation coefficients (see Supplement 6). Regarding the ratio of expression levels of the *cis*-interaction partner gene and the corresponding antisense lncRNA, we observed pronounced differences with a minimal expression ratio of 0.21 and a maximum ratio of 392.77 (mean = 40.34, SD = 61.12). Furthermore, the ratio of expression levels is not necessarily dependent on or in a linear relationship with the general expression level of the two loci, which suggests that this observation is more than random noise (Supplement 6). Only in two cases out of the 44 antisense lncRNA-sense partner locus pairs did the antisense lncRNA have a higher expression level than that of the respective *cis*-interaction partner gene and in such cases the expression ratio was comparatively low (below 0.5).

2.8. Characteristics of Key Regulatory Long Non-Coding RNAs in the Co-Expression Network

Out of the 26 hub lncRNAs (see Supplement 4), three coincided with a known QTL for RFI and 16 were differentially expressed between the efficiency groups. Two of these lncRNAs were both DE and overlapped with a QTL: *MSTRG.4802* and *MSTRG.4839*. In addition, we detected two hub lncRNAs, *MSTRG.3808* and *MSTRG.7798*, to be already annotated as lncRNAs in Ensembl release 97 (*ENSBTAG00000048400* and *ENSBTAG00000053946,* respectively). Both lncRNAs, which were included in our co-expression network, were also predicted by FEELnc to be partner loci to other lncRNAs in the dataset.

The screening for potential *cis*-actions of the 26 hub lncRNAs, i.e., a significant PCIT correlation of $|r| \geq 0.65$ with a locus no farther than 1 Mb away, showed that potentially interacting neighbouring loci could be predicted for 18 out of the 26 hub lncRNA loci. With a total of 45 interactions found, each hub lncRNA had 2.5 *cis*-interactions on average.

Again, the highest correlation coefficients between lncRNA and *cis*-interaction partners were found for *MSTRG.5042* and *Apolipoprotein A1* (*APOA1*, r = 0.98) and for *MSTRG.7472* and *Haptoglobin* (*HP*, r = 0.97). Two of the 26 hub lncRNAs stood out because of their strong wiring with plasma metabolites: *MSTRG.4390* and *MSTRG.5042* were significantly correlated ($p \leq 0.01$, $|r| \geq 0.65$) with 44 and 45 metabolites, respectively. Both hub lncRNAs are also correlated with each other (r = 0.80) and shared 24 loci correlation partners and 42 metabolite correlations. Out of these 42 shared metabolite correlation partners, five were differentially abundant between both groups (q (Benjamini-Hochberg) ≤ 0.05 and $|\log_2 FC| \geq 1$): 10-heptadecenoate (17:1n7), 4-hydroxyglutamate, 9-hydroxystearate, succinate and tetradecanedioate (see Supplement 3).

Because of its multi-categorization (hub lncRNA, DE and QTL locus, *cis*-interaction), we selected *MSTRG.4802* (see Figure 5) for a more detailed analysis of its regulatory function with regard to associated biological pathways. Due to their strong intertwining and apparent connection to plasma metabolite levels, *MSTRG.4390* and *MSTRG.5042* have also received more focus. *MSTRG.5042*, along with *MSTRG.7472*, are relevant due to their strong *cis*-interaction with the corresponding antisense oriented protein-coding gene. Table 2 summarizes positional and structural information for these top four potentially regulatory lncRNAs and condenses results from the differential expression analysis, FEELnc application, and the screening for *cis*-interactions. These lncRNAs have been classified by FEELnc as antisense lncRNAs transcribed in the opposite orientation to their partner genes and can be regarded as NATs. The expression levels of the four hub lncRNAs and their antisense proteincoding partner genes are depicted in Figures S1–S4.

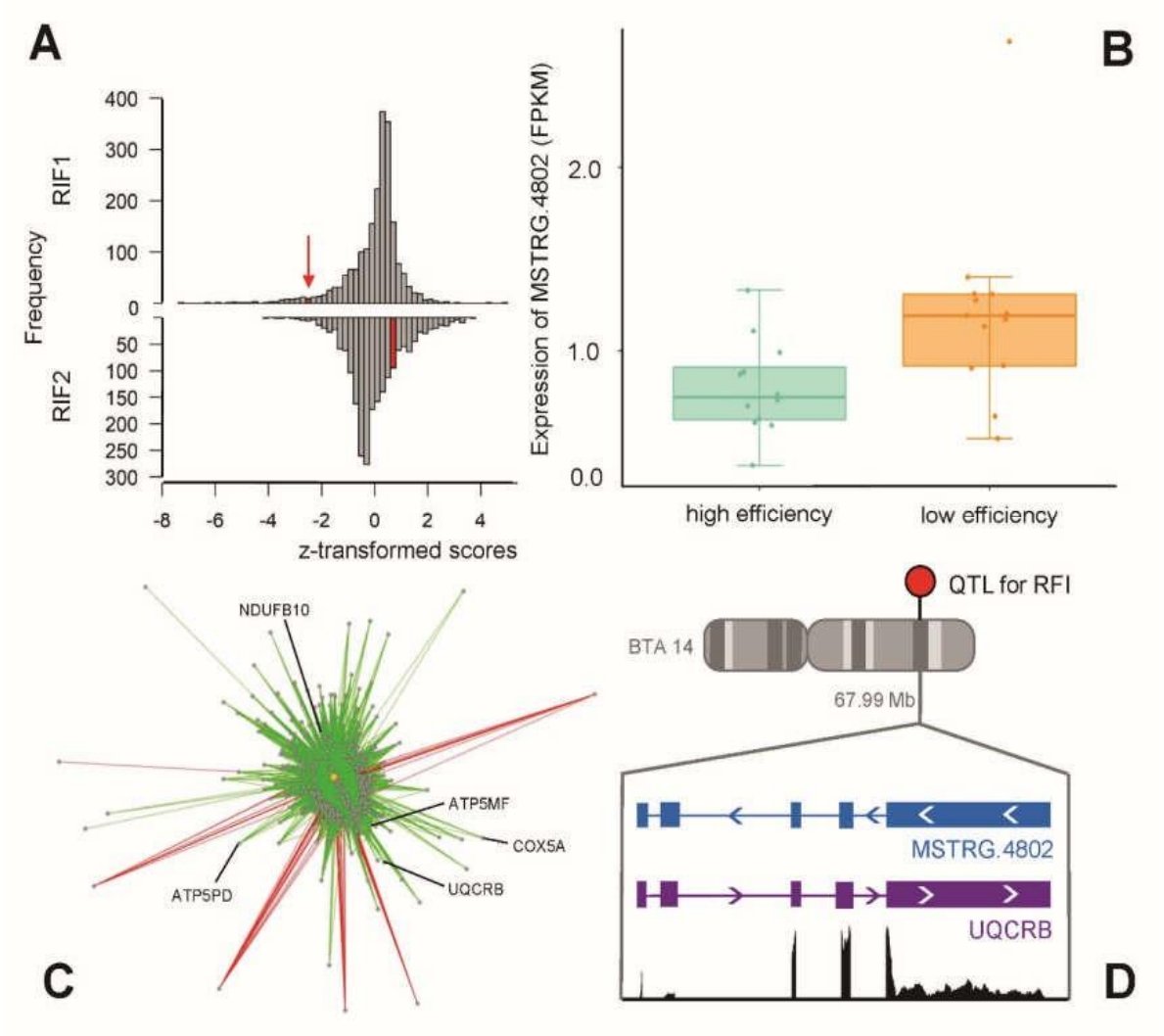

Figure 5. Key lncRNA *MSTRG.4802* with (**A**) a significant RIF1 score, (**B**) differential expression between bulls of high and low feed efficiency, (**C**) high connectivity in a co-expression network, and (**D**) antisense direction to protein-coding gene *UQCRB* on bovine chromosome BTA14 at 67.99 Mb, coinciding with a remapped quantitative trait locus (QTL) for residual feed intake (RFI).

Table 2. Characteristics of four hub lncRNAs with relation to feed efficiency in bulls.

lncRNA	**Position**				**Structure**		**Expression (FPKM [3])**			**Differential Expression Analysis**		
Locus ID	**BTA [1]**	**Start bp [2]**	**End bp**	**Strand**	**Number Exons**	**Exonic Length**	**Mean**	**Mean High Efficiency Group**	**Mean Low Efficiency Group**	**Log_2FC [4]**	***p*-Value**	**Adjusted *p*-Value (BH [5])**
MSTRG.4390	14	518,688	534,106	-	2	20,919	2.586	2.672	2.507	0.0661	0.501	0.796
MSTRG.4802	14	67,986,656	67,991,285	-	5	806	1.009	0.798	1.205	-0.6310	0.004	0.091
MSTRG.5042	15	27,503,347	27,512,980	+	7	3,002	0.843	1.044	0.658	0.6330	0.043	0.287
MSTRG.7472	18	39,037,005	39,043,726	+	7	1,920	11.200	11.016	11.370	-0.1053	0.886	0.966

lncRNA	**FEELnc Analysis**					***cis* Action**		
Locus ID	**Best Potential Partner Gene**	**Direction**	**Type**	**Distance**	**Subtype Location**	**Interaction Partner Gene**	**PCIT (r) [7]**	**Direction**
MSTRG.4390	*ENSBTAG00000046026*	AS [6]	genic	overlapping	exonic	no *cis* interaction with a minimal correlation of r = 0.65		
MSTRG.4802	*ENSBTAG00000001521 (UQCRB)*	AS	genic	nested	exonic	*ENSBTAG00000001521 (UQCRB)* *MSTRG.4780* *ENSBTAG00000032432 MSTRG.4798*	0.69 0.67 0.67 0.67	antisense sense sense sense
MSTRG.5042	*ENSBTAG00000002258 (APOA1)*	AS	genic	containing	exonic	*ENSBTAG00000002258 (APOA1)*	0.98	antisense
MSTRG.7472	*ENSBTAG00000006354 (HP)*	AS	genic	containing	exonic	*ENSBTAG00000006354 (HP)*	0.97	antisense

[1] BTA = bovine chromosome, [2] bp = base pair, [3] FPKM = fragment per kilobase per million, [4] FC = foldchange, [5] BH = Benjamini–Hochberg, [6] AS = anti-sense, [7] PCIT (r) = correlation coefficient r from partial correlation and information theory analysis.

2.9. Pathway Enrichment Analysis

In order to detect generally enriched pathways in the liver transcriptome between the two efficiency groups, the DE genes were submitted to Ingenuity Pathway Analysis (IPA). The pathway PPARα/RXRα activation was significantly enriched ($p \leq 0.01$, equaling $-\log_{10}(p) \geq 2.0$) (downregulated in highly efficient bulls ($-\log_{10}(p) = 6.12$, z-score = −0.707), as was the pathway of VDR/RXR activation ($-\log_{10}(p) = 3.06$, z-score = −1.0). A slight upregulation of the NRF2-mediated oxidative stress response for high efficiency bulls was also detected ($-\log_{10}(p) = 2.83$, z-score = 0.447).

Focusing on transcriptional upstream regulators, we observed the strongest inhibition in the high efficiency group for the peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PPARGC1A, activation = −2.268, $p = 1.3 \times 10^{-3}$) and the strongest activation for the hypoxia-inducible factor 1-alpha (HIF1A, activation = 2.348, $p = 3.21 \times 10^{-3}$). Detailed results for enriched canonical pathways and upstream regulators are given in Supplement 7 and Supplement 8, respectively. The Ingenuity pathway enrichment analysis with genes associated with the four selected top hub lncRNAs (the NATs *MSTRG.4390*, *MSTRG.4802*, *MSTRG.5042*, *MSTRG.7472*) revealed significant enrichments of specific biological pathways. A summary of the top five enriched canonical pathways is provided in Table 3. Transcriptional upstream regulators are listed in Table 4, prioritized for results with an activation score if available.

MSTRG.4802 had by far the strongest z-score (−2.236) for the pathway of oxidative phosphorylation ($-\log_{10}(p) = 7.00$), followed by mitochondrial dysfunction ($-\log_{10}(p) = 6.02$). The enrichment of both pathways was based on the correlated genes *ATP5MF, ATP5PD, COX5A, NDUFB10* and *UQCRB* encoding protein members of mitochondrial respiratory chain complexes. The gene *ubiquinol-cytochrome c reductase binding protein* (*UQCRB*) was predicted by FEELnc to be the positional interaction partner for *MSTRG.4802*, which was confirmed by the finding of a *cis*interaction. The lncRNA *MSTRG.4802* was found in the antisense direction to its interaction partner and both loci displayed a positive correlation of their expression levels with r = 0.69. The abovementioned upstream regulators PPARGC1A ($p = 6.2 \times 10^{-3}$) and HIF1A ($p = 1.44 \times 10^{-3}$) were detected to be significant for genes correlated with *MSTRG.4802* (see Supplement 7), as well as the paralogue transcription regulator PPARGC1B ($p = 3.0 \times 10^{-3}$).

Of all performed enrichment analyses, *MSTRG.7472* had the overall lowest p-value ($-\log_{10}(p) = 11.2$) for the pathway of acute phase response signaling, which was downregulated in the high efficiency group (z-score = −0.378). One of the major genes involved in this pathway is *haptoglobin* (*HP*), which we predicted to be in *cis*-interaction with lncRNA *MSTRG.7472*. In addition, the pathway unfolded protein response was found to be upregulated in highly efficient bulls ($-\log_{10}(p) = 6.82$, zscore = 0.447) for *MSTRG.7472*. One of the correlated genes, *STAT3*, was also found to be a downregulated transcription regulator (activation = −0.877, $p = 6.51 \times 10^{-5}$). Again, an upregulation of HIF1A (activation = 1.932, $p = 3.21 \times 10^{-3}$) could be registered, as well a positive activation of hepatocyte nuclear factor 1 homeobox a (HNF1A, activation = 1.114, $p = 1.77 \times 10^{-6}$) (Supplement 8).

Table 3. Top 5 enriched canonical pathways for key lncRNAs related to feed efficiency.

Lnc RNA	Ingenuity Canonical Pathways	$-\log_{10}(p)$	*p*-Value	Ratio	z-Score	Molecules
MSTRG.4390	Fatty Acid β-oxidation I	5.56	2.75×10^{-6}	8.89×10^{-2}	1.00	*ACADM, ACSL1, ECHS1, HADHB*
	Palmitate Biosynthesis I (Animals)	3.52	3.02×10^{-4}	1.67×10^{-1}	NaN	lauric acid, palmitic acid
	Stearate Biosynthesis I (Animals)	3.52	3.02×10^{-4}	5.00×10^{-2}	NaN	*ACSL1*, palmitic acid, stearic acid
	Ketolysis	3.11	7.76×10^{-4}	1.05×10^{-1}	NaN	*HADHB*, succinic acid
	γ-linolenate Biosynthesis II (Animals)	2.91	1.23×10^{-3}	8.33×10^{-2}	NaN	*ACSL1*, linoleic acid

MSTRG.4802	Oxidative Phosphorylation	7.00	1.00×10^{-7}	4.2×10^{-2}	−2.236	*ATP5MF, ATP5PD, COX5A, NDUFB10, UQCRB*
	Mitochondrial Dysfunction	6.02	9.55×10^{-7}	2.66×10^{-2}	NaN	*ATP5MF, ATP5PD, COX5A, NDUFB10, UQCRB*
	Spermine Biosynthesis	2.16	6.92×10^{-3}	1.43×10^{-1}	NaN	*SMS*
	Sirtuin Signaling Pathway	1.40	3.98×10^{-2}	6.17×10^{-3}	NaN	*ATG3, NDUFB10*
	TNFR1 Signaling	1.32	4.79×10^{-2}	2.00×10^{-2}	NaN	*MADD*
MSTRG.5042	TCA Cycle II (Eukaryotic)	3.48	3.31×10^{-4}	7.14×10^{-2}	NaN	fumaric acid, L-malic acid, succinic acid
	Palmitate Biosynthesis I (Animals)	3.19	6.46×10^{-4}	1.67×10^{-1}	NaN	lauric acid, palmitic acid
	Glycerol Degradation I	3.12	7.59×10^{-4}	1.54×10^{-1}	NaN	*GK*, glycerol
	Stearate Biosynthesis I (Animals)	3.03	9.33×10^{-4}	5.00×10^{-2}	NaN	*ACSL1*, palmitic acid, stearic acid
	γ-linolenate Biosynthesis II (Animals)	2.58	2.63×10^{-3}	8.33×10^{-2}	NaN	*ACSL1*, linoleic acid
MSTRG.7472	Acute Phase Response Signaling	1.12×10^{1}	6.31×10^{-12}	5.52×10^{-2}	−0.378	*C5, FGG, HP, HPX, HRG, LBP, OSMR, SAA2, SOCS3, STAT3*
	Unfolded protein response	6.82	1.51×10^{-7}	8.93×10^{-2}	0.447	*CANX, DNAJC3, P4HB, PDIA6, XBP1*
	Role of JAK family kinases in IL-6-type Cytokine Signaling	4.64	2.29×10^{-5}	1.20×10^{-1}	NaN	*OSMR, SOCS3, STAT3*
	Role of JAK2 in Hormone-like Cytokine Signaling	4.24	5.75×10^{-5}	8.82×10^{-2}	NaN	*GHR, SOCS3, STAT3*
	Role of Tissue Factor in Cancer	3.85	1.41×10^{-4}	3.36×10^{-2}	NaN	*CFL1, FGG, P4HB, PDIA6*

NaN = not a number.

Table 4. Transcriptional upstream regulators for key lncRNAs with an activation score (except for *MSTRG.4802*: here all transcriptional regulators are listed).

Lnc RNA	Upstream Regulator	Activation z-Score	*p*-Value of Overlap	Target Molecules in Dataset
MSTRG.4390	PML	−2.433	1.22×10^{-6}	*ACADM, APOA1, HADHB*, myristic acid, palmitic acid, stearic acid
	TP53	0.113	3.21×10^{-2}	*ACADM, ACSL1, APOA1, HADHB, IDH1, INHBA, NDRG2, PCK1*
	SIRT1	0.317	8.98×10^{-3}	*ACADM*, glycerol, *MAT2A, PCK1*
	MYC	0.577	2.51×10^{-2}	*IDH1, INHBA, MAT2A, NDRG2, PCK1, SHMT2*
	SREBF1	0.652	1.69×10^{-3}	*ACSL1, ARF4, IDH1, PCK1*
	HNF4A	1.181	8.09×10^{-3}	*ACSL1, APOA1, HADHB, HSDL2, INHBA, MAT2A, MPP1, PCK1, RAB30, TRIP11*
	PPARGC1A	1.729	7.33×10^{-5}	*ACADM, INHBA*, myristic acid, palmitic acid, *PCK1*, stearic acid
	PPARGC1B	2.177	4.51×10^{-7}	*ACADM*, myristic acid, palmitic acid, *PCK1*, stearic acid
MSTRG.4802	PPARGC1B		3.03×10^{-3}	*ATP5MF, COX5A*
	ARID5B		4.15×10^{-3}	*UQCRB*
	Esrra		5.68×10^{-3}	*ATP5MF, COX5A*
	PPARGC1A		6.22×10^{-3}	*ATP5MF, ATP5PD, COX5A*
	HNF1A		1.44×10^{-2}	*AP3M1, ATG3, CLTRN*
	KMT2D		1.85×10^{-2}	*FBXL21P, PTGR2*
	SUB1		2.97×10^{-2}	NDUFB10

	HTT		4.36×10^{-2}	*AGRN, ATP5MF, UQCRB*
MSTRG.5042	PML	−2.000	1.89×10^{-3}	*APOA1*, myristic acid, palmitic acid, stearic acid
	SREBF1	0.652	7.56×10^{-3}	*ACSL1, ARF4, IDH1, PCK1*
	TCF7L2	0.728	2.99×10^{-3}	*ACSL1, ADIPOR2, FBP1, IDH1, PCK1*
	HNF4A	1.505	1.03×10^{-2}	*ACSL1, APOA1, ASGR2, FBP1, HSDL2, INHBA, MAT2A, PABPN1, PCK1, RAB30, RTCB, SOAT2, TRIP11*
	PPARGC1A	1.673	7.26×10^{-4}	*GK, INHBA*, myristic acid, palmitic acid, *PCK1*, stearic acid
	SP1	1.934	2.66×10^{-2}	*ACSL1, APOA1, MAT2A, PCK1, THRB*
	PPARGC1B	2.000	8.73×10^{-5}	myristic acid, palmitic acid, *PCK1*, stearic acid
MSTRG.7472	STAT3	−0.877	6.51×10^{-5}	*C5, FGG, HP, LBP, PDIA4, SOCS3, STAT3, XBP1*
	TP53	−0.640	3.11×10^{-2}	*CD44, HDLBP, NARS1, P4HB, PDIA6, STAT3, TMSB10/TMSB4X, UGDH, XBP1*
	ATF4	−0.152	3.90×10^{-5}	*CANX, NARS1, OSMR, SLC39A14, STAT3*
	CEBPB	−0.133	5.64×10^{-5}	*HP, HPX, LBP, SAA2, SOCS3, STAT3, XBP1*
	NFE2L2	0.000	6.00×10^{-8}	*C5, DNAJC3, GHR, NARS1, PDIA4, PDIA6, SOCS3, TMED2, UGDH, XBP1*
	XBP1	0.262	1.16×10^{-6}	*DNAJC3, FKBP2, P4HB, PDIA4, PDIA6, SEC61G, XBP1*
	ATF6	0.762	1.50×10^{-5}	*DNAJC3, PDIA4, SLC39A14, XBP1*
	TCF3	1.000	6.56×10^{-8}	*AZGP1, EPRS1, GPLD1, NUF2, PDIA4, PDIA6, RASSF4, SOCS3, XBP1*
	TCF4	1.000	3.11×10^{-4}	*NUF2, PDIA4, PDIA6, SOCS3, STAT3, XBP1*
	HNF1A	1.114	1.77×10^{-6}	*C5, FGL1, HOPX, HPX, LBP, NUF2, SOCS3, TARS1, XBP1*
	PRDM1	1.176	1.91×10^{-3}	*CD44, FGG, TRIB1, XBP1*
	HIF1A	1.932	3.21×10^{-3}	*CD44, GHR, HP, SOCS3, STAT3*

The lncRNAs *MSTRG.4390* and *MSTRG.5042* were highly correlated with each other (r = 0.80). The analysis showed that, based on their correlation partners, they were enriched for functionally related pathways: fatty acid ß-oxidation ($-\log_{10}(p)$ = 5.56, z-score = 1) was upregulated for *MSTRG.4390* in highly efficient animals and *MSTRG.5042* showed an enrichment for the TCA cycle II ($-\log_{10}(p)$ = 3.48, no z-score) in this experimental group. The analysis of potential upstream regulators revealed the same strongest transcriptional regulators for both lncRNAs: a downregulation of promyelocytic leukemia (PML; *MSTRG.4390*: activation = -2.433, $p = 1.22 \times 10^{-6}$, *MSTRG.5042*: activation = −2.000, $p = 1.89 \times 10^{-3}$) and an upregulation of PPARGC1B (*MSTRG.4390*: activation = 2.177, $p = 4.51 \times 10^{-7}$, *MSTRG.5042*: activation = 2.000, $p = 8.73 \times 10^{-5}$) in the high efficiency group could be observed (Table 4).

The analysis of both lncRNAs and their correlation partners combined showed a significant enrichment for valine degradation ($-\log_{10}(p)$ = 5.18, z-score = 0), followed by the pathways that had been detected on an individual basis as well: fatty acid ß-oxidation ($-\log_{10}(p)$ = 4.74, z-score = 1.00) and the TCA cycle II ($-\log_{10}(p)$ = 3.37, no z-score). Analogously, to the individual analysis of potential upstream regulators, the strongest activation for transcription regulators was observed for PPARGC1B (activation = 2.177, $p = 6.26 \times 10^{-6}$) and PML was significantly inhibited (activation = 2.433, $p = 2.68 \times 10^{-5}$) in animals of high efficiency (see Supplement 7 and Supplement 8).

3. Discussion

We studied crossbred F2-bulls (Charolais x Holstein Friesian) with divergent feed efficiency and fat deposition at a transcriptomic (liver) and metabolomics (blood plasma) level and integrated these data to identify lncRNAs and predict their potential biological function through biological pathway enrichment analyses. Using the bioinformatics lncRNA prediction tool FEELnc [22], which has been applied to determine lncRNAs in different species, including dogs [44], chicken [45], cattle [30,36] and pigs [46], we have identified 3590 lncRNA loci expressed in the liver transcriptome.

In a previous study, our group employed the herein presented pipeline, which applied a systems biology approach combining RIF and PCIT algorithms with biological network prediction to identify potential key regulatory lncRNAs in a tissue- and sex-overarching approach [36]. However, other studies have shown that many lncRNAs are tissue-specific in their expression pattern [47]. To better understand the function of lncRNAs and their interactions in the liver, we have now focused on this single organ due to its relevance in the context of metabolism [38] and the immune system [48].

We therefore adjusted the pipeline, especially regarding the loci set prioritization: the category of tissue-specific loci was excluded and instead potential positional partner loci of lncRNAs, as predicted by FEELnc, were included. Furthermore, we lowered the minimal expression threshold to at least 0.1 FPKM in at least six animals of one group. In contrast to the previous study, we used raw FPKM values for calculations instead of log-transformed values. This step presented itself as necessary to account for the relatively low abundance of lncRNAs compared with mRNAs in the transcriptome [37,49]. Indeed, in our study, 2335 out of the 3590 lncRNA loci (65%) had an average expression level in the liver of less than 1 FPKM.

The prediction of potential biological functions of the identified key lncRNAs was based on the premise that they were involved in the same biological pathway as their correlated partner loci or metabolites. This guilt-by-association heuristic, in which correlating genes or metabolites are used to perform enrichment analyses for biological pathways and then to infer functional involvement for novel, non-coding elements has already been applied to miRNAs and lncRNAs, e.g., [27,32], and [30]. When interpreting the results of such analyses, it should be kept in mind that these predictions heavily depend on the statistical method used to calculate the correlation coefficients [50]. The PCIT algorithm that we applied in our study ensures that the detected pairwise loci correlations are independent of any other third locus in the dataset [35].

Up to now, the combined application of the RIF and PCIT allowed for the discovery of regulatory genomic elements in cattle with regard to a variety of phenotypes: e.g., feed efficiency [51], puberty [52,53], as well as the mineral content [54], intramuscular fat content [55] and fatty acid composition in muscle [56]. Our study showed that the functional prediction of lncRNAs with potential regulatory activity in cattle that differed in their phenotypes in terms of feed efficiency, pointed towards their involvement in immunological pathways, the TCA-cycle, fatty acid β-oxidation, and mitochondrial function.

3.1. LncRNAs Participating in Fatty Acid β-Oxidation and TCA-Cycle

The relevance of mitochondrial function and energy metabolism for feed efficiency was underlined by the key lncRNAs *MSTRG.4390* and *MSTRG.5042* and their respective pathway enrichments for fatty acid β-oxidation and the TCA-cycle. In the mitochondria, the fatty acids are broken down to produce acetyl-CoA that then enters the TCA cycle. The β-oxidation is MTPdependent (mitochondrial trifunctional protein), which is encoded by the genes *HADHA* and *HADHB*. The latter was part of our prioritized loci set because it was predicted as the positional interaction gene of lncRNA *MSTRG.2669*, but it turned out to be significantly correlated ($r = 0.7153$) with *MSTRG.4390*. Though no differential abundance was found for carnitine or acetylcarnitine, which are indicative of a challenged β-oxidation when decreased [57], a number of long-chain fatty acids (e.g., stearoyl carnitine, palmitoyl carnitine, docosapentaenoate) was positively correlated with *MSTRG.4390* expression, along with the related enzyme encoding gene *ACSL1*. Both *MSTRG.4390* and *MSTRG.5042* shared most of their correlation partners, including fatty acids, which suggests a common biological function. However, only *MSTRG.5042* correlated with all three successive TCA cycle products: succinate, fumarate, and malate. Analogous to these findings, Wang and Kadarmideen [58] also found an enrichment for the citrate cycle in an integrative study of metabolomics and transcriptomic data in cattle divided into high and low residual feed intake.

A definitive functional prediction for *MSTRG.5042* remained challenging, because its strongest associations ($r > 0.9$) were with its *cis*-partners *APOA1* (*Apolipoprotein A1*) and *MAT2A* (*Methionine Adenosyltransferase 2A*). The protein encoded by *MAT2A* catalyses the production of S-adenosylmethionine from methionine. While *MAT2A* had higher expression levels in animals of high feed efficiency, methionine itself was of significantly higher abundance in plasma in bulls of low feed efficiency (high RFI). *APOA1* was downregulated in bulls of low efficiency, which is in accordance with findings of Gondret et al. [59] in pigs and Zhuo et al. [60] in chickens.

It is noteworthy that the lncRNA *MSTRG.5042* was exactly in the antisense position to *APOA1*, but displayed a 50-fold lower average expression. We found that PPARGC1B, a key regulator of mitochondrial biogenesis [61], is the most strongly activated upstream regulator (z-score = 2.177, $p = 6.26 \times 10^{-6}$) when comparing animals of high efficiency with low efficiency animals, which is supported by the findings of Vigors et al. [62] in pigs.

3.2. LncRNA Linked to Mitochondrial Function and Energy Metabolism

Exploring the potential regulatory impact of the hub lncRNAs revealed that they might modulate mitochondrial processes and energy metabolism. In our study, the enrichment hits for lncRNA *MSTRG.4802* suggest its involvement in oxidative phosphorylation and mitochondrial dysfunction. *MSTRG.4802* was particularly interesting, because it did not only have a significant RIF1 score—and thereby a predicted high regulatory potential—but was also DE with a significantly lower expression in high efficiency bulls. In addition, its *cis*-interaction partner *UQCRB* (*UbiquinolCytochrome C Reductase Binding Protein*) also displayed a lower expression level in animals of high feed efficiency. *UQCRB*, which is fundamental for the functioning of the mitochondrial respiratory chain complex III [63], is on the opposite strand and in complete overlap with *MSTRG.4802*. Interestingly, this locus falls within a remapped QTL region for RFI as well [4], which supports its putative relevance in the regulation of the related biological processes.

3.3. LncRNA Associated with Immunological Functions

There is a tight relationship between the animal's immune response and its performance in feed efficiency or growth-related traits. Although not DE in our dataset, the correlation of *MSTRG.7472* with *HP*, *LBP, SOCS3* and *SAA2* indicates that this lncRNA is functionally involved in the acute phase signaling. Already in early life stages, inflammation negatively affects growth rates and the average daily gain in feedlot calves [64]. Subsequently, at puberty, gene modules that were associated with feed efficiency in bulls showed enrichments for an immune and an inflammatory response, whereby the authors had reasons to assume that this was due to a bacterial infection of the liver [39]. Mukiibi et al. [41] assessed the liver transcriptome of bulls—similar in age to our cohort—in different breeds and found the acute phase signaling pathway to be among the top enrichment hits in Angus steers of divergent growth performance.

3.4. LncRNAs Putatively Involved in Gluconeogenesis

As Ingenuity Pathway Analysis is deeply rooted in human research, biological processes and pathways that are specific to other species might therefore be overlooked. We considered it noteworthy that *MSTRG.4390* and *MSTRG.5042* both correlated with the gene *PCK1* at expression level and that *MSTRG.5042* expression also correlated with that of *FBP1*. Both *PCK1* and *FBP1* occupy key roles in gluconeogenesis, a biological pathway that is particularly important for the energy balance in cattle [65]. The correlation of *MSTRG.5042* with the metabolite glycerol supports the assumption that these lncRNAs might be involved in the regulation of hepatic gluconeogenesis in cattle [66]. Additionally, we found lactate to be differentially abundant and at significantly higher levels in the plasma of highly efficient animals. The available amount of the glucogenic precursors lactate and glycerol, next to glucogenic amino acids and volatile fatty acids, substantially influences the hepatic glucose production [66]. In this context, we found that the high-connectivity key lncRNA *MSTRG.9118* was co-expressed with *G6PC,* encoding the enzyme that controls the glucose release in hepatocytes and thereby plays a central role in this biological pathway [67]. *MSTRG.9118* is also antisense oriented to *G6PC.*

3.5. LncRNAs as Natural Antisense Transcripts

The above-mentioned four hub lncRNAs (*MSTRG.4390, MSTRG.4802, MSTRG.5042, MSTRG.7472*) lie in antisense orientation to and almost completely overlap with a protein-coding gene on the opposite strand. Furthermore, all four hub lncRNAs were positively co-expressed with their *cis*-partner locus. The observation of nearly complete or perfect antisense overlaps between the paired protein-coding genes and non-coding RNAs has already been described and reviewed for natural antisense transcripts (NATs) by Latgé et al. [68]. Our observation of predominantly positive

correlations between key lncRNAs and the paired locus on the opposite strand confirmed the findings of Wenric et al. [17]. The authors found that strong negative correlations ($r < -0.4$) between

the mirroring pairs were rare and the correlation coefficients ranged between 0.431 and 0.533 [17]. Indeed, we only found two strong negative correlations between key antisense lncRNAs and the overlapping paired partner locus. We could also confirm strong differences in expression levels between the non-coding NATs and their protein-coding partners, although not as strong as described by Wenric et al. (up to a 1000-fold). Indeed, the observed expression ratios of partner gene expression level divided by antisense lncRNA expression level were rather variable and ranged from 0.21 to 392.77. Only in two exceptional cases of *cis*-interactions of NAT lncRNAs (out of 44) did these have higher expression levels than their *cis*-partner gene, and in both cases, the expression ratio was below 0.5. As reported by Napoli et al. [69], NATs have been found to be implicated in multiple regulatory mechanisms, including RNA masking, alternative splicing and chromatin remodelling. A conceivable function of our key lncRNAs, which are positively correlated with their associated antisense locus partner at expression level, could be the stabilization of the corresponding paired transcript. The stabilization might occur by protecting the transcript from degradation, binding to miRNAs or corrosive post-transcriptional processes [70]. Such lncRNAs with potential protective properties would easily have been overlooked in the past before the introduction of stranded RNA sequencing libraries in 2008 [71].

4. Materials and Methods

4.1. Animals

The bulls in our study were part of a F2-population of a Charolais x Holstein Friesian cross (SEGFAM [72]). The animals were bred and raised at the Leibniz Institute for Farm Animal Biology (FBN) in Dummerstorf (Germany) and kept under standardized housing conditions, as previously described by Eberlein et al. [73] and Widmann et al. [74]. The bulls' individual feed intake was measured daily, and body weight was recorded on a monthly basis. Animals were slaughtered at 18 months of age and the carcasses underwent detailed dissection, including measurements for intramuscular (IMF) and carcass (CF) fat percentage. The bulls were split into groups of high or low efficiency depending on their residual feed intake (RFI) in the last month of life, their IMF in *M. longissimus dorsi* and their CF percentage. Bulls were assigned to the high efficiency group if they had a low RFI (at least one standard deviation (SD) below average) and a lower CF as well as a lower IMF than the population mean (CF: mea*n* = 16.5% ± 4.0%; IMF: mea*n* = 3.67% ± 1.76%; *n* = 246). All animals had to have a positive daily weight gain and no less than the population average minus one SD. Accordingly, bulls were grouped to low efficiency if they had a high RFI (at least one SD above average), and a higher CF and IMF than the mean (see Table 5). Archer's formula [75] was used to calculate the individual RFI, which equals the bulls' energy intake while accounting for the average daily weight gain and metabolic mid-weight (average body weight during the last month of life raised to the power of 0.75). For the current study, out of 246 deeply phenotyped F2-bulls, 26 bulls were selected with extremely high (*n* = 13) or low efficiency (*n* = 13).

All experimental procedures were carried out according to the German animal care guidelines and were approved (27 March 2003) and supervised by the relevant authorities of the State Mecklenburg-Vorpommern, Germany (State Office for Agriculture, Food Safety and Fishery; LALLF M-V/TSD/7221.3-2.1-010/03).

Table 5. Phenotypic characteristics of bulls in high and low efficiency group.

Group	Number of Animals.	CF (%)		IMF (%)		RFI in MJ ME/Day	
		Mean	SD	Mean	SD	Mean	SD
high efficiency	13	14.39	2.86	2.77	0.95	−20.91	4.47
low efficiency	13	20.28	4.06	4.59	1.71	20.48	4.40

CF = carcass fat content, IMF = intramuscular fat content in *M. longissimus dorsi*, RFI = residual feed intake, MJ ME = megajoule metabolizable energy, SD = standard deviation.

4.2. Plasma Metabolites

Blood samples were taken on the day of slaughter before transit to the slaughterhouse and holistic metabolite profiles with 640 biochemical compounds and molecules in plasma were established by Metabolon Inc. (Durham, NC, USA, https://www.metabolon.com/). With ultra-highperformance liquid chromatography and tandem accurate mass spectrometry (UHPLC/MS/MS) methods, compounds and derivatives of eight different metabolite classes were determined: amino acids (n = 167), carbohydrates (n = 27), cofactors and vitamins (n = 19), energy (n = 10), lipids (n = 278), nucleotides (n = 36), peptides (n = 35), and xenobiotics (n = 68). As animal B002 (high efficiency group) clustered unexpectedly within the inefficient group in the transcriptomic analysis, this animal was excluded from further metabolomics analysis steps.

For differential abundance analysis of metabolites in the blood plasma, the R-package MetaboDiff [76] was used and the author's instructions were closely followed. As recommended, metabolites with more than 40% missing cases were excluded and for the remaining metabolites, missing values were imputed with the k-nearest neighbor algorithm. A total of 552 metabolites remained in the dataset, which was then normalized using a variance stabilization transformation. For the comparison of the high and low efficiency group, a Student's *t*-Test was applied, and p-values were corrected for multiple testing with the Benjamini–Hochberg procedure [77].

4.3. Sampling, RNA Isolation, Library Preparation, and Sequencing

Immediately after slaughter and dissection, tissue samples were taken from the liver (*Lobus caudatus*), shock frozen in liquid nitrogen and then stored at −80 °C. For RNA extraction, the samples were ground in liquid nitrogen and 30 mg were subjected to an on-column-purification with the NucleoSpin RNA II kit (Macherey and Nagel, Düren, Germany), which included a DNase digestion to remove genomic DNA. RNA was subsequently tested for remaining DNA residues and further cleansed, if necessary, according to Weikard et al. [78]. The RNA concentration and integrity were measured with a Qubit Fluorometer (Invitrogen, Karlsruhe, Germany) and a 2100 Bioanalyzer Instrument (Agilent Technologies, Waldbronn, Germany). From 1 µg of total RNA per sample, stranded, ribodepleted and indexed libraries were prepared with the TruSeq Stranded RNA-RiboZero H/M/R Gold Kit (Illumina, San Diego, CA, USA). Paired-end reads were sequenced (2 × 100 bp) in a multiplexed design on a HiSeq 2500 Sequencing System (Illumina).

4.4. Alignment and Assembly

Raw reads were subjected to quality control with FastQC [79], adapter trimming with Cutadapt v.1.6.1 [80] and thereafter quality trimming with Quality Trim v. 1.6.0 [81]. For quality trimming, the sequence start was also processed (option -s), the maximum number of missing bases (N) was set to 3, and the minimum base quality was set to 15. In a guided alignment, the reads were then mapped with HISAT2 v.2.1.0 [82] to the latest bovine reference genome ARS-UCD1.2 [83] with Ensembl annotation release 97 [84]. The sorting and indexing of BAM files were performed with samtools v.1.6 [85] and Stringtie v.1.3.4d [86] was used for the individual assembly while using the reference genome and annotation in a guided approach. For this study, we created a project-specific annotation with Stringtie merge (default settings for Stringtie merge and a minimal read alignment per exonic base (-c) of 15). To this end, we made use of the bovine reference genome, the 26 bull liver samples, as well as 178 other samples available from a previous study [36]. These samples included 26 liver samples from cows of the same resource population, as well as muscle (n = 52), jejunum (n = 48) and rumen (n = 52) samples of these cows and the bulls used in the present study.

The merged annotation was checked for plausibility, i.e., the number of exons for each transcript and the number of transcripts for each locus. We excluded loci that had over 20 transcripts, unless one of these transcripts was already annotated, in which case only that particular transcript was kept for the locus. In the reference annotation (Ensemble release 97), the maximum number of exons per transcript was 173 and therefore we set a cut-off threshold of 200 exons per transcript. Transcripts with more than 200 exons were excluded from the merged annotation, except for two transcripts overlapping with the gene *titin,* which is highly expressed in muscle tissue and has been annotated with 335 exons in NCBI (National Center for Biotechnology Information, annotation release 106).

The transcriptome dataset examined in this study was already used in a previous study ([36], aligned to UMD.3.1, Ensembl annotation release 92) and is stored in the Functional Annotation of Animal Genomes (FAANG) database (https://data.faang.org/dataset) under project number PRJEB34570.

4.5. Long Non-Coding RNA Prediction and Fragment Counting

The computational identification of lncRNAs was carried out with FEELnc [22], while making use of the merged annotation and the bovine reference genome and annotation ARS-UCD1.2. (Ensembl 97). Annotated loci of the protein coding biotype were excluded, and the minimal transcript length was kept at the default of 200 nt. To reduce the number of false positives, monoexonic transcripts were discarded, unless they were in antisense localization. The coding potential for all remaining transcripts was determined in shuffling mode.

Except for the differential expression analysis, fragments per kilobase per million mapped reads (FPKM) were used in all further analysis steps. These were calculated based on fragment counts derived with featureCounts [87]. All loci needed to have a minimal expression of at least 0.1 FPKM in at least six animals of one experimental group. The expression threshold was deliberately set this low in order to keep as many predicted lncRNAs in the dataset as possible. Loci that were annotated as ribosomal, spliceosomal, metazoan or Y-RNA genes were generally discarded.

4.6. Loci Set Prioritization

To enable the construction of meaningful co-expression networks, we compiled a list of prioritized loci, which included loci that belonged to at least one of the following four categories: predicted lncRNA (lncRNA), potential interaction partner of the lncRNA (partner locus), overlapping or in close proximity of up to 3 Mb of a QTL (QTL locus), and differentially expressed between the groups of high and low efficiency (DE locus).

Loci were included in the 'lncRNA' category if one of the locus' transcripts was predicted as lncRNA using FEELnc and the minimal expression threshold was exceeded. Loci were included in the category 'partner locus' of the prioritized loci set if FEELnc predicted them to be positional interaction partners and rated them 'best choice' with a score of 1. FEELnc determines the most likely positional interaction partner for a lncRNA based on its physical genomic position relative to the nearest locus. The best choice thereby is a locus that overlaps with the lncRNA, preferentially at an exon, and if no overlapping locus can be found, the closest neighbor is chosen instead.

Loci were included in the category 'QTL locus' if they were minimally expressed and overlapped with or were no farther away than 3 Mb from a QTL for residual feed intake (RFI) in cattle. QTLs were downloaded from the Animalgenome QTL database (https://www.animalgenome.org/cgibin/QTLdb/BT/index, accessed 10 October 2019) and only QTL based on SNP array studies were kept. The QTL positions were then remapped to the new reference genome ARS-UCD1.2 with the NCBI Genome Remapping Service and default options (https://www.ncbi.nlm.nih.gov/genome/tools/remap, accessed on 22 November 2019).

The differential expression analysis was performed with the R-package DESeq2 [88]. Cluster analysis revealed unexpected clustering of animal B002 in a PCA-plot based on read counts. Due to pathological findings in the liver, this animal was excluded from all further analyses. The model for differential expression analysis included the efficiency group; an effect of year of slaughter or birth could not be included because all animals of the high efficiency group were born between 2002 and 2007 and all animals of the low efficiency group were born between 2008 and 2011. Loci were considered significantly differentially expressed (DE) if they were minimally expressed and withstood a correction for multiple testing with the Benjamini–Hochberg [77] procedure (adjusted *p*value (q) ≤ 0.1).

4.7. Regulatory Impact Factor Analysis

The regulatory impact factor (RIF) algorithm of Reverter et al. [34] is designed to detect loci with high regulatory potential in a prioritized loci set while contrasting two biological conditions or groups. The analysis makes use of two metrics: RIF1 and RIF2. A high RIF1 score was attributed to lncRNAs that were co-expressed with abundant target loci (DE, QTL, partner) in both efficiency groups. A high RIF2 score was assigned to lncRNAs if they were strongly correlated with a target

locus in one group but displayed no or a reversed correlation to the same target locus in the other group. Since some lncRNAs were also categorized as DE, QTL or partner loci, they could also be targets in the RIF analysis. RIF scores were standardized with a z-transformation and lncRNAs with either a RIF1 or RIF2 score of ≥ 1.96 were deemed significant, which corresponds to a significance threshold of $p \leq 0.05$ in a t-test. Subsequently, lncRNAs with a significant RIF score (key lncRNAs) were closely scrutinized in the co-expression networks.

4.8. Partial Correlation and Information Theory

The partial correlation and information theory (PCIT [35]) calculates pairwise correlations between loci while accounting for the influence of a third locus. Unlike likelihood-based approaches, which invoke a parametric distribution (e.g., normal) assumed to hold under the null hypothesis and then a nominal p-value (e.g., 5%) used to ascertain significance, PCIT is an information theoretic approach. Its threshold is an informative metric, in this case the partial correlation after exploring all trios in judging the significance of a given correlation, which might then become a connection when inferring a network. It thereby tests all possible three-way combinations in a dataset and only keeps correlations between loci if they are significant and independent of the expression of another locus, whereas no hard threshold is set for the correlation strength. The significance threshold for each combination of loci depends on the average ratio of partial and direct correlation [35]. The set of prioritized loci that was subjected to the RIF analysis was also used for the PCIT.

4.9. Correlation of Plasma Metabolites with Key LncRNAs

A Pearson correlation coefficient was calculated with the function rcorr of the Hmisc R-package [89] for all key lncRNAs (significant RIF score) and plasma metabolites. The data curation was independent from the differential abundance analysis of metabolites and a lower number of missing cases was accepted for the correlation analysis. The raw metabolite values were filtered for metabolites with less than five missing cases and missing values imputed with the minimum observation, assuming that the missing value was due to a value below the detection limit and not a technical error. The values were then scaled with the scale-function in R (without centering). Correlations were considered significant if they had a $p \leq 0.01$.

4.10. Natural Antisense Transcripts

The results from FEELnc were filtered for key lncRNAs (significant RIF score) that overlapped with a predicted positional interaction partner locus on the opposite strand (antisense direction and a distance of 0 bp to the partner locus). LncRNAs that are in antisense position to another gene have been described as natural antisense transcripts (NATs) in the literature before [17] and fall into the category of *cis*-interaction partners. We wanted to screen for valid *cis*-interactions, meaning a correlation in expression and not a mere positional neighborhood. To this end, we checked for significant PCIT correlations between the antisense lncRNAs and the respective partner loci, regardless of correlation strength or direction (positive or negative).

4.11. Selection of Hub Key lncRNAs in Co-Expression Network

The visualization of the co-expression network was realized in Cytoscape 3.6.1 [90]. All significant PCIT correlations with a minimum strength of $|r| \geq 0.65$ between lncRNAs with a significant RIF score and any other locus from the prioritized loci set were included. Additionally, significant correlations between the above-mentioned lncRNAs and plasma metabolites were also included if they had a minimal correlation strength of $|r| \geq 0.65$. We filtered for lncRNAs with a

significant RIF score that were correlated with at least 10 annotated genes, having an official gene symbol available and not predicted to be a lncRNA. To further narrow down the selection to impactful lncRNAs, we filtered for lncRNAs that fulfilled either of the following three criteria: I) categorization as a DE or QTL locus, II) additional correlation with at least 10 metabolites, or III) exceptionally high connectivity with >50 annotated genes with an official gene symbol in the bovine genome annotation. LncRNAs that fulfilled these criteria were labelled key lncRNAs.

4.12. Cis-Action of Hub LncRNAs

In addition to the screening for NATs, we searched for *cis*-interaction partners for hub lncRNAs in a larger radius. All loci within a physical distance of up to 1 Mb and with a correlation significant according to PCIT and substantial in magnitude such that $|r| \geq 0.65$ were considered for each individual hub lncRNA. Since the lncRNA prediction in FEELnc works in a transcript-based manner, only the transcript of a locus that was actually predicted to be non-coding was considered.

4.13. Pathway Enrichment Analysis

In order to discern the probable biological functions of hub lncRNAs, we conducted pathway enrichment analyses with significantly and substantially correlated loci and metabolites ($|r| \geq 0.65$) for each of them. Additionally, to investigate which biological pathways are generally to be addressed for our animal material and phenotype, an enrichment analysis was done for all DE between the high and low efficiency group. The list of metabolites and genes and their logged fold changes were submitted to the Ingenuity Pathway Analysis (IPA: QIAGEN, Inc., http://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) [91]. Pathways were considered significantly enriched at a *p*-value of $p \leq 0.05$ equalling a $-\log_{10}(p\text{-value})$ of 1.3. The same significance threshold was applied to upstream regulators in the pathway enrichment analyses.

5. Conclusions

With this study, we enlarged the catalogue of lncRNAs from bovine liver, identified hub lncRNAs that are potentially involved in biological processes and pathways modulating feed efficiency in bulls and made first predictions contributing to their functional annotation. Our results underline the importance of immunological pathways and metabolic pathways associated with mitochondrial processes of the metabolic phenotype related to feed efficiency in bulls and suggest a possible regulatory function of key lncRNAs with regard to their modulating and fine-tuning role within these biological pathways.

A substantial proportion of the identified lncRNAs fall into the category of natural antisense transcripts, which most likely perform a stabilizing function with respect to mRNAs transcribed from the opposite strand. This function needs to be validated by further studies. To what extent these lncRNAs and the associated biological processes and pathways are also relevant in cows or bulls at other life stages requires further investigations.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/21/9/3292/s1.

Author Contributions: Conceptualization, C.K. and R.W.; Methodology, C.K., A.R., R.W. and W.N.; Software, A.R.; Formal Analysis, W.N.; Investigation, W.N., C.K. and R.W.; Resources, C.K., H.M.H., R.M.B., E.A.; Data Curation, W.N., C.K.; Writing – Original Draft Preparation, W.N., R.W. and C.K.; Writing – Review & Editing, all authors; Visualization, W.N.; Supervision, C.K. and R.W.; Project Administration, C.K. and R.W.; Funding Acquisition, C.K. and R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the German Research Foundation (DFG–grant numbers: KU 771/8-1 and WE 1786/5-1).

Acknowledgments: The authors thank Frieder Hadlich for his support in bioinformatics issues, as well as Simone Wöhl and Bärbel Pletz for the excellent work accomplished in the lab.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

lncRNA	long non-coding RNA
RFI	residual feed intake
QTL	quantitative trait locus
NAT	natural antisense transcript
FAANG	Functional Annotation of Animal Genomes
RIF	regulatory impact factor
PCIT	partial correlation and information theory
nt	nucleotide
FC	foldchange
PCA	principal component analysis
DE	differentially expressed
FPKM	fragments per kilobase million
IPA	Ingenuity Pathway Analysis
SD	standard deviation
CF	carcass fat
IMF	intramuscular fat content
Mb	megabase

References

1. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74, doi:10.1038/nature11247.
2. Salzman, J.; Gawad, C.; Wang, P.L.; Lacayo, N.; Brown, P.O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **2012**, *7*, e30733, doi:10.1371/journal.pone.0030733.
3. Kenny, D.A.; Fitzsimons, C.; Waters, S.M.; McGee, M. Invited review: Improving feed efficiency of beef cattle – the current state of the art and future challenges. *Animal* **2018**, *12*, 1815–1826, doi:10.1017/S1751731118000976.
4. Saatchi, M.; Beever, J.E.; Decker, J.E.; Faulkner, D.B.; Freetly, H.C.; Hansen, S.L.; Yampara-Iquise, H.; Johnson, K.A.; Kachman, S.D.; Kerley, M.S., et al. QTLs associated with dry matter intake, metabolic midtest weight, growth and feed efficiency have little overlap across 4 beef cattle studies. *BMC Genomics* **2014**, *15*, 1004, doi:10.1186/1471-2164-15-1004.
5. Seabury, C.M.; Oldeschulte, D.L.; Saatchi, M.; Beever, J.E.; Decker, J.E.; Halley, Y.A.; Bhattarai, E.K.; Molaei, M.; Freetly, H.C.; Hansen, S.L., et al. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* **2017**, *18*, 386–386, doi:10.1186/s12864-017-3754-y.
6. de Oliveira, P.S.; Cesar, A.S.; do Nascimento, M.L.; Chaves, A.S.; Tizioto, P.C.; Tullio, R.R.; Lanna, D.P.; Rosa, A.N.; Sonstegard, T.S.; Mourao, G.B., et al. Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet* **2014**, *15*, doi:10.1186/s12863-014-0100-0.

7. Higgins, M.G.; Fitzsimons, C.; McClure, M.C.; McKenna, C.; Conroy, S.; Kenny, D.A.; McGee, M.; Waters, S.M.; Morris, D.W. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Sci Rep* **2018**, *8*, 14301–14301, doi:10.1038/s41598-018-32374-6.
8. Long, Y.; Wang, X.; Youmans, D.T.; Cech, T.R. How do lncRNAs regulate transcription? *Sci Adv* **2017**, *3*, doi:10.1126/sciadv.aao2110.
9. Marchese, F.P.; Raimondi, I.; Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol* **2017**, *18*, 206, doi:10.1186/s13059-017-1348-2.
10. Lu, W.; Cao, F.; Wang, S.; Sheng, X.; Ma, J. LncRNAs: The Regulator of Glucose and Lipid Metabolism in Tumor Cells. *Front Oncol* **2019**, *9*, 1099, doi:10.3389/fonc.2019.01099.
11. Muret, K.; Désert, C.; Lagoutte, L.; Boutin, M.; Gondret, F.; Zerjal, T.; Lagarrigue, S. Long noncoding RNAs in lipid metabolism: literature review and conservation analysis across species. *BMC Genomics* **2019**, *20*, 882, doi:10.1186/s12864-019-6093-3.
12. Yang, L.; Li, P.; Yang, W.; Ruan, X.; Kiesewetter, K.; Zhu, J.; Cao, H. Integrative Transcriptome Analyses of Metabolic Responses in Mice Define Pivotal LncRNA Metabolic Regulators. *Cell Metab* **2016**, *24*, 627–639, doi:10.1016/j.cmet.2016.08.019.
13. Pradas-Juni, M.; Hansmeier, N.R.; Link, J.C.; Schmidt, E.; Larsen, B.D.; Klemm, P.; Meola, N.; Topel, H.; Loureiro, R.; Dhaouadi, I., et al. A MAFG-lncRNA axis links systemic nutrient abundance to hepatic glucose metabolism. *Nat Commun* **2020**, *11*, 644, doi:10.1038/s41467-020-14323-y.
14. Zhang, Y.; Liu, X.S.; Liu, Q.R.; Wei, L. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res* **2006**, *34*, 3465–3475, doi:10.1093/nar/gkl473.
15. Katayama, S.; Tomaru, Y.; Kasukawa, T.; Waki, K.; Nakanishi, M.; Nakamura, M.; Nishida, H.; Yap, C.C.; Suzuki, M.; Kawai, J., et al. Antisense transcription in the mammalian transcriptome. *Science* **2005**, *309*, 1564–1566, doi:10.1126/science.1112009.
16. Pelechano, V.; Steinmetz, L.M. Gene regulation by antisense transcription. *Nat Rev Genet* **2013**, *14*, 880–893, doi:10.1038/nrg3594.
17. Wenric, S.; ElGuendi, S.; Caberg, J.-H.; Bezzaou, W.; Fasquelle, C.; Charloteaux, B.; Karim, L.; Hennuy, B.; Frères, P.; Collignon, J., et al. Transcriptome-wide analysis of natural antisense transcripts shows their potential role in breast cancer. *Sci Rep* **2017**, *7*, 17452, doi:10.1038/s41598-017-17811-2.
18. Li, B.; Hu, Y.; Li, X.; Jin, G.; Chen, X.; Chen, G.; Chen, Y.; Huang, S.; Liao, W.; Liao, Y., et al. Sirt1 Antisense Long Noncoding RNA Promotes Cardiomyocyte Proliferation by Enhancing the Stability of Sirt1. *J Am Heart Assoc* **2018**, *7*, e009700, doi:10.1161/jaha.118.009700.
19. Ma, X.-Y.; Wang, J.-H.; Wang, J.-L.; Ma, C.X.; Wang, X.-C.; Liu, F.-S. Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of earlystage hematopoietic cells. *BMC Genomics* **2015**, *16*, 676, doi:10.1186/s12864-015-1881-x.
20. Washietl, S.; Kellis, M.; Garber, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res* **2014**, *24*, 616–628, doi:10.1101/gr.165035.113.
21. Li, A.; Zhang, J.; Zhou, Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinf* **2014**, *15*, 311, doi:10.1186/1471-2105-15-311.
22. Wucher, V.; Legeai, F.; Hedan, B.; Rizk, G.; Lagoutte, L.; Leeb, T.; Jagannathan, V.; Cadieu, E.; David, A.; Lohi, H., et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res* **2017**, *45*, e57, doi:10.1093/nar/gkw1306.
23. Hezroni, H.; Koppstein, D.; Schwartz, M.G.; Avrutin, A.; Bartel, D.P.; Ulitsky, I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* **2015**, *11*, 1110–1122, doi:10.1016/j.celrep.2015.04.023.
24. Kong, L.; Zhang, Y.; Ye, Z.Q.; Liu, X.Q.; Zhao, S.Q.; Wei, L.; Gao, G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* **2007**, *35*, 345–349, doi:10.1093/nar/gkm391.
25. Wang, L.; Park, H.J.; Dasari, S.; Wang, S.; Kocher, J.-P.; Li, W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **2013**, *41*, e74, doi:10.1093/nar/gkt006.
26. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* **2013**, *41*, e166, doi:10.1093/nar/gkt646.
27. Oliveira, G.B.; Regitano, L.C.A.; Cesar, A.S.M.; Reecy, J.M.; Degaki, K.Y.; Poleti, M.D.; Felicio, A.M.; Koltes, J.E.; Coutinho, L.L. Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics* **2018**, *19*, 16, doi:10.1186/s12864-018-4514-3.
28. Deng, L.; Wang, J.; Zhang, J. Predicting Gene Ontology Function of Human MicroRNAs by Integrating Multiple Networks. *Front Genet* **2019**, *10*, doi:10.3389/fgene.2019.00003.
29. Bansal, A.; Singh, T.R.; Chauhan, R.S. A novel miRNA analysis framework to analyze differential biological networks. *Sci Rep* **2017**, *7*, 14604, doi:10.1038/s41598-017-14973-x.

30. Weikard, R.; Hadlich, F.; Hammon, H.M.; Frieten, D.; Gerbert, C.; Koch, C.; Dusel, G.; Kuehn, C. Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of preweaning calves in response to different diets. *Oncotarget* **2018**, *9*, 21052–21069, doi:10.18632/oncotarget.24898.
31. Lin, X.C.; Zhu, Y.; Chen, W.B.; Lin, L.W.; Chen, D.H.; Huang, J.R.; Pan, K.; Lin, Y.; Wu, B.T.; Dai, Y., et al. Integrated analysis of long non-coding RNAs and mRNA expression profiles reveals the potential role of lncRNAs in gastric cancer pathogenesis. *Int J Oncol* **2014**, *45*, 619–628, doi:10.3892/ijo.2014.2431.
32. Bakhtiarizadeh, M.R.; Salami, S.A. Identification and Expression Analysis of Long Noncoding RNAs in FatTail of Sheep Breeds. *G3-Genes Genom Genet* **2019**, *9*, 1263–1276, doi:10.1534/g3.118.201014.
33. Yue, B.; Li, H.; Liu, M.; Wu, J.; Li, M.; Lei, C.; Huang, B.; Chen, H. Characterization of lncRNA–miRNA–mRNA Network to Reveal Potential Functional ceRNAs in Bovine Skeletal Muscle. *Front Genet* **2019**, *10*, doi:10.3389/fgene.2019.00091.
34. Reverter, A.; Hudson, N.J.; Nagaraj, S.H.; Perez-Enciso, M.; Dalrymple, B.P. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* **2010**, *26*, 896–904, doi:10.1093/bioinformatics/btq051.
35. Reverter, A.; Chan, E.K. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* **2008**, *24*, 2491–2497, doi:10.1093/bioinformatics/btn482.
36. Nolte, W.; Weikard, R.; Brunner, R.M.; Albrecht, E.; Hammon, H.M.; Reverter, A.; Kühn, C. Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle. *Front Genet* **2019**, *10*, doi:10.3389/fgene.2019.01130.
37. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, doi:10.1101/gad.17446611.
38. Rui, L. Energy metabolism in the liver. *Compr Physiol* **2014**, *4*, 177–197, doi:10.1002/cphy.c130024.
39. Alexandre, P.A.; Kogelman, L.J.A.; Santana, M.H.A.; Passarelli, D.; Pulz, L.H.; Fantinato-Neto, P.; Silva, P.L.; Leme, P.R.; Strefezzi, R.F.; Coutinho, L.L., et al. Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC Genomics* **2015**, *16*, 13, doi:10.1186/s12864-0152292-8.
40. Fonseca, L.D.; Eler, J.P.; Pereira, M.A.; Rosa, A.F.; Alexandre, P.A.; Moncau, C.T.; Salvato, F.; RosaFernandes, L.; Palmisano, G.; Ferraz, J.B.S., et al. Liver proteomics unravel the metabolic pathways related to Feed Efficiency in beef cattle. *Sci Rep* **2019**, *9*, 5364, doi:10.1038/s41598-019-41813-x.
41. Mukiibi, R.; Vinsky, M.; Keogh, K.; Fitzsimmons, C.; Stothard, P.; Waters, S.M.; Li, C. Liver transcriptome profiling of beef steers with divergent growth rate, feed intake, or metabolic body weight phenotypes1. *J Anim Sci* **2019**, *97*, 4386–4404, doi:10.1093/jas/skz315.
42. Salleh, S.M.; Mazzoni, G.; Lovendahl, P.; Kadarmideen, H.N. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. *BMC Bioinf* **2018**, *19*, 15, doi:10.1186/s12859-018-2553-z.
43. Tizioto, P.C.; Coutinho, L.L.; Decker, J.E.; Schnabel, R.D.; Rosa, K.O.; Oliveira, P.S.; Souza, M.M.; Mourão, G.B.; Tullio, R.R.; Chaves, A.S., et al. Global liver gene expression differences in Nelore steers with divergent residual feed intake phenotypes. *BMC Genomics* **2015**, *16*, 242, doi:10.1186/s12864-015-1464-x.
44. Le Beguec, C.; Wucher, V.; Lagoutte, L.; Cadieul, E.; Botherel, N.; Hedan, B.; De Brito, C.; Guillory, A.S.; Andre, C.; Derrien, T., et al. Characterisation and functional predictions of canine long non-coding RNAs. *Sci Rep* **2018**, *8*, 12, doi:10.1038/s41598-018-31770-2.
45. Muret, K.; Klopp, C.; Wucher, V.; Esquerré, D.; Legeai, F.; Lecerf, F.; Désert, C.; Boutin, M.; Jehl, F.; Acloque, H., et al. Long noncoding RNA repertoire in chicken liver and adipose tissue. *Genet Sel Evol* **2017**, *49*, 6, doi:10.1186/s12711-016-0275-0.
46. Jin, L.; Hu, S.; Tu, T.; Huang, Z.; Tang, Q.; Ma, J.; Wang, X.; Li, X.; Zhou, X.; Shuai, S., et al. Global Long Noncoding RNA and mRNA Expression Changes between Prenatal and Neonatal Lung Tissue in Pigs. *Genes* **2018**, *9*, 443, doi:10.3390/genes9090443.
47. Ulitsky, I.; Shkumatava, A.; Jan, C.H.; Sive, H.; Bartel, D.P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell.* **2011**, *147*, doi:10.1016/j.cell.2011.11.055.
48. Bogdanos, D.P.; Gao, B.; Gershwin, M.E. Liver immunology. *Compr Physiol* **2013**, *3*, 567–598, doi:10.1002/cphy.c120011.
49. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G., et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **2012**, *22*, 1775–1789, doi:10.1101/gr.132159.111.
50. Ehsani, R.; Drabløs, F. Measures of co-expression for improved function prediction of long non-coding RNAs. *BMC Bioinf* **2018**, *19*, 533, doi:10.1186/s12859-018-2546-y.
51. Alexandre, P.A.; Naval-Sanchez, M.; Porto-Neto, L.R.; Ferraz, J.B.S.; Reverter, A.; Fukumasu, H. Systems Biology Reveals NR2F6 and TGFB1 as Key Regulators of Feed Efficiency in Beef Cattle. *Front Genet* **2019**, *10*, doi:10.3389/fgene.2019.00230.

52. Canovas, A.; Reverter, A.; DeAtley, K.L.; Ashley, R.L.; Colgrave, M.L.; Fortes, M.R.S.; Islas-Trejo, A.; Lehnert, S.; Porto-Neto, L.; Rincon, G., et al. Multi-Tissue Omics Analyses Reveal Molecular Regulatory Networks for Puberty in Composite Beef Cattle. *PLoS One* **2014**, *9*, 17, doi:10.1371/journal.pone.0102551.
53. Nguyen, L.T.; Reverter, A.; Cánovas, A.; Venus, B.; Anderson, S.T.; Islas-Trejo, A.; Dias, M.M.; Crawford, N.F.; Lehnert, S.A.; Medrano, J.F., et al. STAT6, PBX2, and PBRM1 Emerge as Predicted Regulators of 452 Differentially Expressed Genes Associated With Puberty in Brahman Heifers. *Front Genet* **2018**, *9*, doi:10.3389/fgene.2018.00087.
54. Afonso, J.; Fortes, M.R.S.; Reverter, A.; da Silva Diniz, W.J.; Cesar, A.S.M.; de Lima, A.O.; Petrini, J.; de Souza, M.M.; Coutinho, L.L.; Mourão, G.B., et al. Genetic regulators of mineral amount in Nelore cattle muscle predicted by a new co-expression and regulatory impact factor approach. *bioRxiv* **2019**, 10.1101/804419, 804419, doi:10.1101/804419.
55. Cesar, A.S.M.; Regitano, L.C.A.; Koltes, J.E.; Fritz-Waters, E.R.; Lanna, D.P.D.; Gasparin, G.; Mourao, G.B.; Oliveira, P.S.N.; Reecy, J.M.; Coutinho, L.L. Putative Regulatory Factors Associated with Intramuscular Fat Content. *PLoS One* **2015**, *10*, 21, doi:10.1371/journal.pone.0128350.
56. de Oliveira, P.S.N.; Coutinho, L.L.; Cesar, A.S.M.; Diniz, W.J.d.S.; de Souza, M.M.; Andrade, B.G.; Koltes, J.E.; Mourão, G.B.; Zerlotini, A.; Reecy, J.M., et al. Co-Expression Networks Reveal Potential Regulatory Roles of miRNAs in Fatty Acid Composition of Nelore Cattle. *Front Genet* **2019**, *10*, doi:10.3389/fgene.2019.00651.
57. Adam, A.C.; Lie, K.K.; Moren, M.; Skjaerven, K.H. High dietary arachidonic acid levels induce changes in complex lipids and immune-related eicosanoids and increase levels of oxidised metabolites in zebrafish (Danio rerio). *Br J Nutr* **2017**, *117*, 1075–1085, doi:10.1017/s0007114517000903.
58. Wang, X.; Kadarmideen, H.N. Metabolomics Analyses in High-Low Feed Efficient Dairy Cows Reveal Novel Biochemical Mechanisms and Predictive Biomarkers. *Metabolites* **2019**, *9*, doi:10.3390/metabo9070151.
59. Gondret, F.; Vincent, A.; Houée-Bigot, M.; Siegel, A.; Lagarrigue, S.; Causeur, D.; Gilbert, H.; Louveau, I. A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. *BMC Genomics* **2017**, *18*, 244, doi:10.1186/s12864-017-3639-0.
60. Zhuo, Z.; Lamont, S.J.; Lee, W.R.; Abasht, B. RNA-Seq Analysis of Abdominal Fat Reveals Differences between Modern Commercial Broiler Chickens with High and Low Feed Efficiencies. *PLoS One* **2015**, *10*, e0135810, doi:10.1371/journal.pone.0135810.
61. Lehman, J.J.; Barger, P.M.; Kovacs, A.; Saffitz, J.E.; Medeiros, D.M.; Kelly, D.P. Peroxisome proliferator–activated receptor γ coactivator-1 promotes cardiac mitochondrial biogenesis. *J Clin Invest* **2000**, *106*, 847– 856, doi:10.1172/JCI10268.
62. Vigors, S.; O'Doherty, J.V.; Bryan, K.; Sweeney, T. A comparative analysis of the transcriptome profiles of liver and muscle tissue in pigs divergent for feed efficiency. *BMC Genomics* **2019**, *20*, 461, doi:10.1186/s12864-019-5740-z.
63. Haut, S.; Brivet, M.; Touati, G.; Rustin, P.; Lebon, S.; Garcia-Cazorla, A.; Saudubray, J.M.; Boutron, A.; Legrand, A.; Slama, A. A deletion in the human QP-C gene causes a complex III deficiency resulting in hypoglycaemia and lactic acidosis. *Hum Genet* **2003**, *113*, 118–122, doi:10.1007/s00439-003-0946-0.
64. Gifford, C.A.; Holland, B.P.; Mills, R.L.; Maxwell, C.L.; Farney, J.K.; Terrill, S.J.; Step, D.L.; Richards, C.J.; Burciaga Robles, L.O.; Krehbiel, C.R. Growth and Development Symposium: Impacts of inflammation on cattle growth and carcass merit. *J Anim Sci* **2012**, *90*, 1438–1451, doi:10.2527/jas.2011-4846.
65. Fassah, D.M.; Jeong, J.Y.; Baik, M. Hepatic transcriptional changes in critical genes for gluconeogenesis following castration of bulls. *Asian Austral J Anim Sci* **2018**, *31*, 537–547, doi:10.5713/ajas.17.0875.
66. Larsen, M.; Kristensen, N.B. Effect of abomasal glucose infusion on splanchnic and whole-body glucose metabolism in periparturient dairy cows. *J Dairy Sci* **2009**, *92*, 1071–1083, doi:10.3168/jds.2008-1453.
67. Tanaka, A.; Urabe, S.; Takeguchi, A.; Mizutani, H.; Sako, T.; Imai, S.; Yoshimura, I.; Kimura, N.; Arai, T. Comparison of activities of enzymes related to energy metabolism in peripheral leukocytes and livers between Holstein dairy cows and ICR mice. *Vet Res Commun* **2006**, *30*, 29–38, doi:10.1007/s11259-005-3223y.
68. Latgé, G.; Poulet, C.; Bours, V.; Josse, C.; Jerusalem, G. Natural Antisense Transcripts: Molecular Mechanisms and Implications in Breast Cancers. *Int J Mol Sci* **2018**, *19*, 123, doi:10.3390/ijms19010123.
69. Napoli, S.; Piccinelli, V.; Mapelli, S.N.; Pisignano, G.; Catapano, C.V. Natural antisense transcripts drive a regulatory cascade controlling c-MYC transcription. *RNA Biol* **2017**, *14*, 1742–1755, doi:10.1080/15476286.2017.1356564.
70. Rosikiewicz, W.; Makalowska, I. Biological functions of natural antisense transcripts. *Acta biochimica Polonica* **2016**, *63*, 665–673, doi:10.18388/abp.2016_1350.
71. Cloonan, N.; Forrest, A.R.; Kolle, G.; Gardiner, B.B.; Faulkner, G.J.; Brown, M.K.; Taylor, D.F.; Steptoe, A.L.; Wani, S.; Bethel, G., et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **2008**, *5*, 613–619, doi:10.1038/nmeth.1223.
72. Kühn, C.; Bellmann, O.; Voigt, J.; Wegner, J.; Guiard, V.; Ender, K. An experimental approach for studying the genetic and physiological background of nutrient transformation in cattle with respect to nutrient secretion and accretion type. *Arch Anim Breed* **2002**, *45*, 317–330, doi:10.5194/aab-45-317-2002.
73. Eberlein, A.; Takasuga, A.; Setoguchi, K.; Pfuhl, R.; Flisikowski, K.; Fries, R.; Klopp, N.; Furbass, R.; Weikard, R.; Kuhn, 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*, 951–964, doi:10.1534/genetics.109.106476.

74. Widmann, P.; Nuernberg, K.; Kuehn, C.; Weikard, R. Association of an ACSL1 gene variant with polyunsaturated fatty acids in bovine skeletal muscle. *BMC Genet* **2011**, *12*, 96, doi:10.1186/1471-2156-12-96.
75. Archer, J.A.; Arthur, P.F.; Herd, R.M.; Parnell, P.F.; Pitchford, W.S. Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. *J Anim Sci* **1997**, *75*, 2024–2032, doi:10.2527/1997.7582024x.
76. Mock, A.; Warta, R.; Dettling, S.; Brors, B.; Jäger, D.; Herold-Mende, C. MetaboDiff: an R package for differential metabolomic analysis. *Bioinformatics* **2018**, *34*, 3417–3418, doi:10.1093/bioinformatics/bty344.
77. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Stat Soc Ser B* **1995**, *57*, 289–300, doi:10.2307/2346101.
78. Weikard, R.; Goldammer, T.; Brunner, R.M.; Kuehn, C. Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. *Physiol Genomics* **2012**, *44*, 728–739, doi:10.1152/physiolgenomics.00007.2012.
79. Andrew, S. FastQC: a quality control tool for high throughput sequence data. Availabe online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 28 March 2018)
80. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **2011**, *17*, 10–12, doi:10.14806/ej.17.1.200.
81. Robinson, A. Quality Trim version 1.6.0. Availabe online: https://bitbucket.org/arobinson/qualitytrim/downloads/ (accessed on 29 March 2018).
82. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **2015**, *12*, 357, doi:10.1038/nmeth.3317.
83. Rosen, B.D.; Bickhart, D.M.; Schnabel, R.D.; Koren, S.; Elsik, C.G.; Tseng, E.; Rowan, T.N.; Low, W.Y.; Zimin, A.; Couldrey, C., et al. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **2020**, *9*, doi:10.1093/gigascience/giaa021.
84. Frankish, A.; Vullo, A.; Zadissa, A.; Yates, A.; Thormann, A.; Parker, A.; Gall, A.; Moore, B.; Walts, B.; Aken, B.L., et al. Ensembl 2018. *Nucleic Acids Res* **2017**, *46*, 754–761, doi:10.1093/nar/gkx1098.
85. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
86. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.-C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **2015**, *33*, 290–295, doi:10.1038/nbt.3122.
87. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930, doi:10.1093/bioinformatics/btt656.
88. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**, *15*, 550, doi:10.1186/s13059-014-0550-8.
89. Harrell Jr, F.E. Hmisc: Harrell Miscellaneous. **2019**, *R-package version 4.2-0*.
90. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.
91. Kramer, A.; Green, J.; Pollard, J., Jr.; Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30*, 523–530, doi:10.1093/bioinformatics/btt703.

9.3. Metabogenomic analysis to functionally annotate the regulatory role of long non-coding RNAs in the liver of cows with different nutrient partitioning phenotype.

Nolte W, Weikard R, Albrecht E, Hammon HM, Kühn C.

Eingereicht am 7. November 2020 bei *Genomics*,
unter Begutachtung seit dem 28 November 2020.

Metabogenomic analysis to functionally annotate the regulatory role of long non-coding RNAs in the liver of cows with different nutrient partitioning phenotype

Wietje Nolte[1], Rosemarie Weikard[1], Elke Albrecht[2], Harald M. Hammon[3], Christa Kühn*[1,4]

[1]Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany

[2]Institute of Muscle Biology and Growth, Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany

[3]Institute of Nutritional Physiology "Oskar Kellner", Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany

[4]Faculty of Agricultural and Environmental Sciences, University Rostock, 18059 Rostock, Germany

*** Correspondence:**
Prof. Dr. Christa Kühn
kuehn@fbn-dummerstorf.de

Keywords: Long non-coding RNA; metabogenomics; nutrient partitioning; Bos taurus; functional genome annotation; Functional Annotation of Animal Genomes (FAANG)

Abstract

Long non-coding RNAs (lncRNAs) hold gene regulatory potential, but require substantial further functional annotation in livestock. Applying two metabogenomic approaches by combining transcriptomic and metabolomic analyses, we aimed to identify lncRNAs with potential regulatory function for divergent nutrient partitioning of lactating crossbred cows and to establish metabogenomic interaction networks comprising metabolites, genes and lncRNAs. Through correlation analysis of lncRNA expression with transcriptomic and metabolomic data, we unraveled lncRNAs that have a putative regulatory role in energy and lipid metabolism, the urea and TCA cycles, and gluconeogenesis. Especially *FGF21,* which correlated with a plentitude of differentially expressed genes, differentially abundant metabolites, as well as lncRNAs, suggested itself as a key metabolic regulator. Notably, lncRNAs in close physical proximity to coding-genes as well as lncRNAs with natural antisense transcripts appear to perform a fine-tuning function in gene expression involved in metabolic pathways associated with different nutrient partitioning phenotypes.

1. Introduction

The rise of long non-coding RNA (lncRNA) research began with the discovery of H19 in 1989 [1] and Xist in the early 1990s [2-4]. Over the past three decades, knowledge and understanding of the regulatory power of lncRNA molecules have steadily increased. Mostly in human disease and especially cancer research, lncRNAs have established themselves as regulators and potential biomarkers and have received much attention compared to the non-coding genome of farm animals. Despite the fact that lncRNAs have been at the center of an increasing number of studies, few have been thoroughly characterized with regard to their biological function and mode of action [5].

LncRNAs have a wide range of functions in the cytoplasm and the nucleus. As summarized by Rinn and Chang [6] and Marchese and colleagues [7], lncRNAs act as repressors and enhancers of gene activation and expression and do so both in *cis* and in *trans*. Modes of action hereby include a decoy-function where lncRNAs bind to DNA-regulatory proteins such as transcription factors and thereby influence transcription; they perform scaffolding tasks and serve as adapters for other proteins; they guide protein complexes to target sites on the DNA with temporal and spatial specificity; and they can act in *cis* in enhancer-like function on neighbouring protein-coding genes [8]. Furthermore, lncRNAs can modify chromatin structure and thereby allow for as well as inhibit transcription, e. g. through the interaction with chromatin-modifying complexes [6,7]. Especially natural antisense transcripts (NATs), which are located on the opposite strand of another (usually protein-coding) gene, are assumed to impact the chromatin state and thus have epigenetic regulatory potential (reviewed by Magistri et al. [9]).

Despite recognized species specificity [10], the non-coding genome of farm animals remains relatively unexplored with a limited number of studies on this topic, as reviewed by Weikard and colleagues in 2017 [11] and by Kosinska-Selbi in 2020 [12]. In cattle, phenotypic traits of particular economic importance are feed efficiency, milk yield, disease susceptibility or resistance, and meat quality including fat deposition. Baumgard et al. have stressed the importance of resource efficiency and nutrient partitioning in the lactating dairy cow and have highlighted the opportunity that genomics and omics-research represent for the field [13]. In the critical period around the onset of lactation, cows must adapt to considerable challenges associated with an enormous increase in energy demands. Optimal metabolic adaptation and balancing of complex metabolic and immune processes are necessary to cope with the metabolic priority of the mammary gland. Simultaneously, nutrient and energy homeostasis need to be maintained to prevent the development of metabolic and infectious diseases (as summarized in numerous reviews, e.g. [13-17]). As the liver is the key organ that controls and modulates the metabolic and immunological adaptation, particularly in early lactation, studies of the hepatic transcriptome have been performed to analyze the patterns, changes and adaptations at the molecular level. The major focus has so far been on the protein-coding part of the transcriptome. Tthe potential influence of lncRNAs on biological processes related to nutrient partitioning in lactating cows has only sparsely been addressed. LncRNAs have been shown to be potentially involved in fat metabolism of bovine liver [18], lactation [19,20], and energy metabolism of growing calves in response to different diets [21] As mentioned above, lncRNAs are assumed to facilitate fine-tuning of gene expression, e.g. by *cis*-acting [8], and indeed, we have recently identified a number of NATs associated with regulatory potential in feed efficiency in bulls [22]. With the goal to further elucidate the molecular background of nutrient partitioning, we analyzed crossbred cows in their second lactation (F_2-population of a beef x dairy cross), which strongly differed in their disposition to secrete milk and accrete body fat [23]. Cows of the same population have been examined in previous studies for the expression of candidate genes in liver and selected plasma metabolites related to insulin-dependent glucose metabolism [24,25] as well as the expression of genes playing a regulatory role in liver, mammary gland and skeletal muscle [26]. In the present study, we wanted to investigate the liver transcriptome and plasma metabolome of the cows on a holistic scale with a special focus on lncRNAs that are co-regulated with gene expression levels and metabolite abundances. Moreover, we investigated whether the crossbred cows, which phenotypically differed in terms of nutrient partitioning, showed differences at transcriptional level to dairy cows despite their substantially lower performance level. In order to identify putative key

regulatory lncRNAs, we opted for an integrative metabogenomic approach by using knowledge based joint network and pathway analyses as well as nonbiased co-expression analysis.

2. Materials and Methods

2.1 Animals

The animals in the study were selected from 243 cows of an F_2-population (Charolais x Holstein) bred and kept at the Leibniz Institute for Farm Animal Biology (FBN) in Dummerstorf (Germany). Animals' housing conditions and feeding regime have been described previously [24]. For this study, a cohort of 25 second lactation cows were selected for milk production and nutrient partitioning based on the energy corrected milk (ECM) yield during seven days prior to slaughter at 30 days in lactation, the fat content in carcass (CFC), and the intramuscular fat content (IMF) in *M. longissimus dorsi*. Phenotyping and sampling strategy have been explained in our previous study [27]. The cows were grouped into animals of nutrient partitioning predominantly directed to milk secretion (SEC, n = 13) or to body fat accretion (ACC, n = 12) type (see Supplementary Table 1). The ACC cows displayed higher accretion of body fat and were characterized by higher CFC (25.32 ± 3.27% vs. 17.09 ± 2.59%), higher IMF content (6.18 ± 2.30% vs. 4.16 ± 1.12%), and a lower milk yield (19.28 ± 7.92 kg ECM vs. 190.87 ± 22.02 kg ECM) compared to SEC cows.

2.2 Plasma Metabolites

At slaughter, blood plasma samples (n = 25) were collected and forwarded to Metabolon Inc. (Durham, NC, USA) to establish holistic metabolite profiles with 640 measured biochemical compounds. For the analysis of differential metabolite abundance (DA) in the blood, metabolites were first filtered for compounds that had less than 50% missing cases. Missing values in the remaining 613 metabolites were imputed with half of the measured minimum of the respective metabolite. Group differences were assessed in R with analysis of variance for a linear model including fixed effects for year of birth and group. A correction for multiple testing was done with the Benjamini-Hochberg procedure (www.jstor.org/stable/2346101) and group differences were considered significant with $q \leq 0.05$. Metabolite abundance for each metabolite and group was then calculated with least square means based on normalized values (z-transformation). Based on group means from raw values, fold changes between groups were calculated.

2.3 Sampling, RNA Isolation, Library Preparation, and Sequencing

Liver samples (*Lobus caudatus*) were taken immediately after slaughter and dissection, shock frozen in liquid nitrogen and stored at -80°C. To extract total RNA, samples were ground in liquid nitrogen and 30 mg were used in an on-column-purification with the NucleoSpin RNA II kit (Macherey & Nagel, Düren, Germany), including a DNase digestion step to remove genomic DNA. In case of persisting contamination with DNA residues, the RNA was further cleansed according to Weikard et al. [26]. Indexed, ribodepleted and stranded libraries were prepared from 1 µg of total RNA with the TruSeq Stranded RNA-Ribo-Zero H/M/R Gold Kit (Illumina, San Diego, USA). In a multiplexed design, the libraries were sequenced in paired-end mode (2 x 100 bp) on a HiSeq 2500 platform (Illumina).

2.4 Downstream analysis pipeline

The pipeline for the alignment and assembly of reads, the creation of a project specific annotation as well as the identification of lncRNAs has been described previously in detail [22]. Identical program versions and parameter settings have been used in this study. The pipeline has been applied to the current *Bos taurus* reference genome ARS-UCD.1.2, Ensembl annotation release 97 (https://doi.org/10.1093/nar/gkx1098) and the project specific annotation was subjected, as previously described, to additional sanity checks and quality filters [22]. The transcriptomic data is stored in the Functional Annotation of Animal Genomes (FAANG) database (https://data.faang.org/dataset) under project number PRJEB34570. The project specific annotation is available as Supplementary Data 1.

Gene expression levels (FPKM, fragments per kilobase per million mapped reads) were calculated based on the fragment counts obtained from featureCounts v.1.6.1 [28]. A minimal expression filter was applied to all loci: minimum 0.1 FPKM in six or more animals of either experimental group. We opted for an unconventionally low expression threshold in order to capture as many lncRNAs as possible. Loci that were annotated as spliceosomal, metazoan, ribosomal or Y-RNA genes were excluded from the dataset.

The *in-silico* prediction of lncRNAs was performed with the FEELnc program [29] based on the project-specific merged annotation and the bovine reference genome and annotation ARS-UCD1.2 (Ensembl release 97). We discarded loci, which were annotated as the biotype protein coding and we assumed a minimal transcript length of 200 nt (default). In order to minimize the number of false positives, we discarded monoexonic transcripts unless they were in antisense localization to another locus. For lncRNA assignment, the coding potential of transcripts under scrutiny was evaluated as well as their k-mer composition via shuffling mode.

2.5. Differential Expression Analysis of Loci

The differential gene expression was calculated and analyzed with the R-package DESeq2 [30]. The model for the differential expression analysis (DEA) included the phenotype and the year of slaughter. Loci were considered to be differentially expressed (DE) at a significance level of q (Benjamini Hochberg) ≤ 0.05. A principal component analysis (PCA) was performed in R for the transcriptome and metabolome based on the minimally expressed loci (variance stabilizing transformation, DESeq2 package) and based on the filtered and imputed 613 metabolite values (z-transformed) respectively.

2.6. Joint Pathway Enrichment Analysis and Network Exploration

For an integrative omics-data analysis, DE genes and DA metabolites (FDR ≤ 0.05) were submitted to the web-based application tool MetaboAnalyst 4.0 (https://www.metaboanalyst.ca/, [31]). Genes that had an official gene symbol were considered, excluding *Bos taurus* specific miRNAs, and in addition, metabolites that had an unambiguous identifier in the Human Metabolome Database (HMDB, https://hmdb.ca/, [32]). Gene names and HMDB identifiers were submitted alongside their $\log_2$ fold changes between animals of the two nutrient partitioning phenotypes. For the Joint Pathway Enrichment Analysis, the Gene-Metabolite-Interaction network was selected based on the organism *Bos taurus* and the KEGG (Kyoto Encyclopedia of Genes and Genomes) reference pathway database, version October 2019 (https://www.genome.jp/kegg/, [33]). Default settings were chosen with a hypergeometric test and the topology measure with the degree centrality. The applied integration method was the default option "combine queries". The analysis was conducted for metabolic integrated pathways as well as all integrated pathways. Subsequently, we performed a knowledge-based Network Exploration with the same multi-omics data that were used for the Joint Pathway Enrichment Analysis.

2.7. Correlation of Plasma Metabolites and Loci

To explore locus-locus, locus-metabolite and metabolite-metabolite abundance relationships, we performed a correlation analysis in R. Pearson correlation coefficients were calculated between expression levels (FPKM) of DE genes (FDR ≤ 0.05) and predicted lncRNAs as well as the plasma levels of selected DA metabolites (FDR ≤ 0.05). All loci had to pass a minimal expression threshold of 0.1 FPKM in at least six animals of either group. For the metabolites, a subset ($n = 40$) was selected based on significant differential abundance (FDR ≤ 0.05), highest absolute fold change and with regard to their relevance in metabolic processes, respectively. Furthermore, we only included single representatives for groups of highly correlated metabolites. After the correlation analysis, a correction for multiple testing was applied and correlations were deemed significant with FDR ≤ 0.05.

3. Results and Discussion

In our study, we took alternative approaches to obtain indication on potential regulatory roles of lncRNAs for the phenotypic differentiation between crossbred cows characterized by different nutrient partitioning. (i) We applied a knowledge-based pipeline taking advantage of the joint gene expression – metabolic profile analysis exploiting KEEG pathway information for network and pathway enrichment analyses (via MetaboAnalyst). (ii) We performed agnostic correlation analyses between DA metabolites, DE genes and identified lncRNAs to set up metabogenomic networks. Both approaches were used to highlight key metabolites, for which subsequently highly correlated lncRNAs and DE genes were identified. These alternative data exploration strategies were applied to functionally annotate the potential regulatory role of lncRNAs in the liver of cows with different nutrient partitioning phenotype.

3.1 RNA Sequencing, Transcriptome Assembly and Prediction of Long Non- Coding RNAs

The 25 liver RNA samples, which were used for library preparation, had an average RNA integrity (RIN) of 8.3 ± 0.20 (Supplementary Table 2). The average sequencing depth per sample was above 50.0 million read pairs after read quality trimming. The average read alignment rate to the reference genome ARS-UCD.1.2 was 98.72 ± 0.16%. The project specific annotation, which was used for mapping, contained 30,806 loci and 82,628 transcripts after quality filtering. The averaged mapping rate for fragment counts against this annotation was 85.83 ± 1.13%.

Within the project-specific annotation that incorporated information from the current bovine reference genome, FEELnc predicted 6,161 transcripts as lncRNAs (3,590 loci). For 202 lncRNA transcripts, no potential adjacent interaction partners based on physical neighbourhood in the genome were found within the default window size. For all other lncRNA transcripts, a total of 19,184 interactions were predicted by FEELnc. We performed the lncRNA (structural) characterization in a locus-based manner because the expression level is also measured per locus and because a very different number of transcripts per locus would hamper comparisons (Supplementary Table 3). Therefore, the transcript with the highest number of exons was selected for each lncRNA locus. We observed an almost equal distribution of strandedness for the 3,590 lncRNA loci with 50.8% being on the plus strand and 49.2% on the minus strand. The average number of exons was at 4.5 ± 7.1 and the geometric mean of the total exonic length was 1,724 nucleotides.

3.2 Differential Metabolite Abundance and Gene Expression

The cow groups contrasted in this study showed a divergent phenotype resulting in different ways or priority for utilizing metabolic energy supplied from nutrients, i.e. in secretion of milk or accretion of body fat. It has to be considered that even the SEC cows of this experiment were not at the level of high milk performance of a dairy cow and that both cow groups had a substantially higher body fat deposition than high yielding dairy cows at the same stage of lactation (e.g., IMF for high lactating dairy cows at day 30: 0.9% - 1.42%, [34]). The average daily ECM in the SEC group amounted to 26.7 kg ECM, while at this lactation stage truly high-lactating multiparous dairy cows yield up to 45 kg ECM and beyond [34].

The different plasma metabolomic and the hepatic transcriptomic patterns of both cow groups characterize these two different metabolic phenotypes differing with regard to nutrient partitioning (Figure 1).

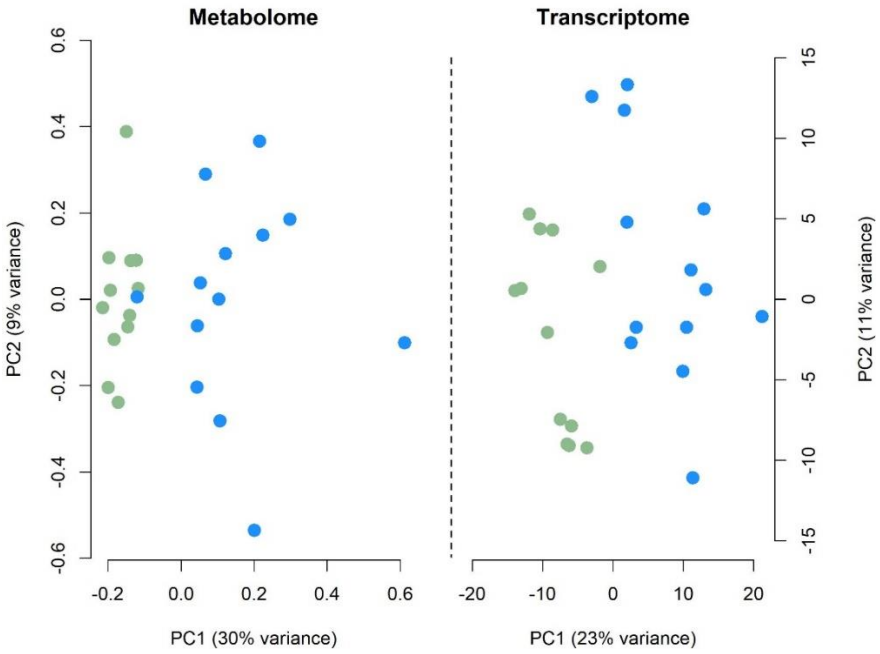

Figure 1: Principal Component Analyses (PCA) based on the metabolome (613 metabolites) and on the transcriptome (18,363 loci) for 25 cows of divergent milk yield and fat deposition (grouped into animals of milk secretion type in blue and animals of body fat accretion type in green).

Out of 613 plasma metabolites that were assessed for differential abundance, 185 were significantly ($q \leq 0.05$) different between the cow groups. While 154 metabolites were found to be of significantly higher abundance in the SEC group, 31 had a significantly lower abundance (see Supplementary Figure 1 and Supplementary Table 4). Plasma metabolomic differences between the animals of the two groups were clearly reflected in a principal component analysis (PCA) by the first component (see Figure 1). The strongest positive differences between the SEC and the ACC group, based on the log-transformed fold change (log_2(FC)), were observed for N-octanoylglycine (log_2(FC) = 3.44, q = 2.96E-02), 3-dehydrocholate (log_2(FC) = 3.08, q = 3.49E-03) and 7-ketodeoxycholate (log_2(FC) = 2.84, q = 2.36E-02). A significantly higher abundance in the SEC group was also observed for a plenitude of lipid metabolites. The most reduced plasma levels in this group were detected for 3-phosphoglycerate (log_2(FC) = -2.33, q = 3.94E-02) and homoarginine (log_2(FC) = -2.12, q = 5.19E-05). Plasma levels of a number of amino acids were also lower in SEC animals (see Supplementary Figure 1 and Supplementary Table 4).

In the hepatic transcriptomic patterns of both cow groups, a total of 18,363 loci were found to exceed the minimum expression limit (≥ 0.1 FPKM in at least six animals of one phenotypic group). A total of 2,114 loci were differentially expressed (DE) at a significance threshold of an adjusted p-value (FDR) ≤ 0.05 between the different phenotypic groups. Thereof, 247 loci, which encompassed a total of 469 transcripts, were characterized as lncRNAs. Among the genes with most prominent DE status, i.e. the lowest adjusted p-value (FDR) and the highest log_2 fold change, were *FGF21* (log_2(FC) = 5.80, FDR = 8.51E-16), *MFSD2A* (log_2(FC) = 3.43, FDR = 1.34E-16), *ANGPTL4* (log_2(FC) = 2.83, FDR = 4.85E-11), *APOA4* (log_2(FC) = 2.56, FDR = 4.06E-12), *LIPG* (log_2(FC) = 4.12, FDR = 1.23E-07), *ADIPOR2* (log_2(FC) = 1.39, FDR = 6.77E-14), and *SLC25A47* (log_2(FC) = 1.68, FDR = 1.19E-13), all indicating higher gene expression levels in the SEC compared to the ACC group. The annotated genes with particular low abundance in cows of a SEC versus ACC type were *REC8* (log_2(FC) = -4.83, FDR = 3.99E-04), *HBB* (log_2(FC) = -4.23, FDR = 3.07E-09), and *ALAS2* (log_2(FC) = -4.17, FDR = 1.86E-05).

3.3 Joint Pathway Enrichment Analysis and network exploration

For functional interpretation of metabolomics and transcriptomic data with special emphasis on lncRNA function associated with different cow phenotypes, we used the joint gene expression – metabolic profile data for network and pathway enrichment exploration. A total of 175 DA HMDB identifiers of metabolites and 1,644 DE gene symbols were submitted alongside their respective $\log_2$ fold changes between cow groups to the MetaboAnalyst tool. The Network Exploration Analysis yielded to two gene-metabolite interaction subnetworks.

Subsequent enrichment analysis of the complex subnetwork 1 (Figure 2) using MetaboAnalyst resulted in 16 significantly ($p \leq 0.05$) enriched KEGG pathways (Supplementary Table 5) with the lowest p-values for arginine biosynthesis (p = 1.86E-04), linoleic acid metabolism (p = 3.3E-04), and glycine, serine and threonine metabolism (p = 4.18E-04). In this network, the metabolites palmitic acid, arachidonic acid, linoleic acid, ornithine, arginine and sphingomyelin (d18:1/18:0) occupy central hub positions. At gene level, *AGTR1, P2RY1, CYP2E1, APOA1, CYP3A4, CYP4A22, GNA14* and *GAA* display the highest number of connections to other nodes in the network (Figure 2). Subnetwork 2 (not shown) contained merely three metabolites and two genes and a KEGG-based enrichment analysis of these five components indicated a significant enrichment for the pyrimidine metabolism (p = 4.76E-03).

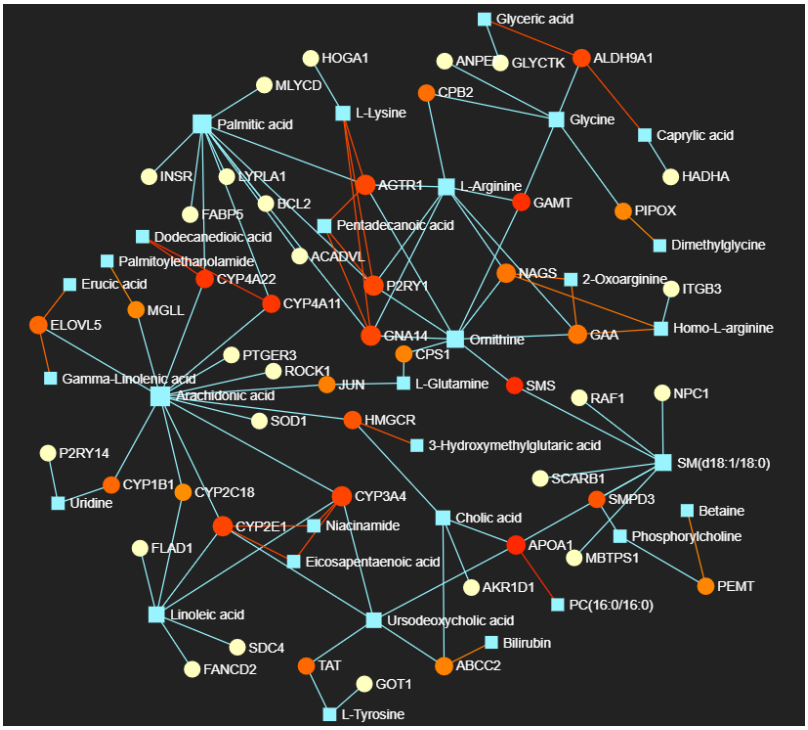

Figure 2: Gene-Metabolite Subnetwork 1 from Metaboanalyst Network Exploration analysis based on DE genes and DA metabolites between cows differing with regard to nutrient partitioning. Circles represent genes, their colour is proportional to the betweenness centrality value of genes (0 = light yellow, medium = orange, and high = red), and metabolites are marked by squares in light blue. Degree reflects the number of connections to other nodes in a network, and betweenness centrality measure reflects the number of shortest paths going through the node (nodes between or connecting networks have high between centrality measures although they might be low in degree themselves).

The Joint Pathway Enrichment Analysis for metabolic pathways (Figure 3, Supplementary Table 6) yielded significant ($p \leq 0.05$) pathway enrichments with highest pathway impact (PI) for phenylalanine, tyrosine and tryptophan biosynthesis (-log(p) = 1.61, PI = 2.4). Enriched pathways with the lowest p-values were glycerolipid metabolism (-log(p) = 6.81, PI = 1.06) and arginine biosynthesis (-log(p) = 6.88, PI = 1.31). When all KEGG pathways were included in the Joint Pathway Enrichment Analysis, the pathways with lowest p-values were protein processing in endoplasmatic reticulum (-log(p) = 11.93, PI = 0.24), PPAR signaling pathway (-log(p) = 8.02, PI = 1.68), and again arginine biosynthesis (-log(p) = 6.88, PI = 1.8). The retinol metabolism scored highest with regard to PI in this analysis (-log(p) = 1.64, PI = 2.44).

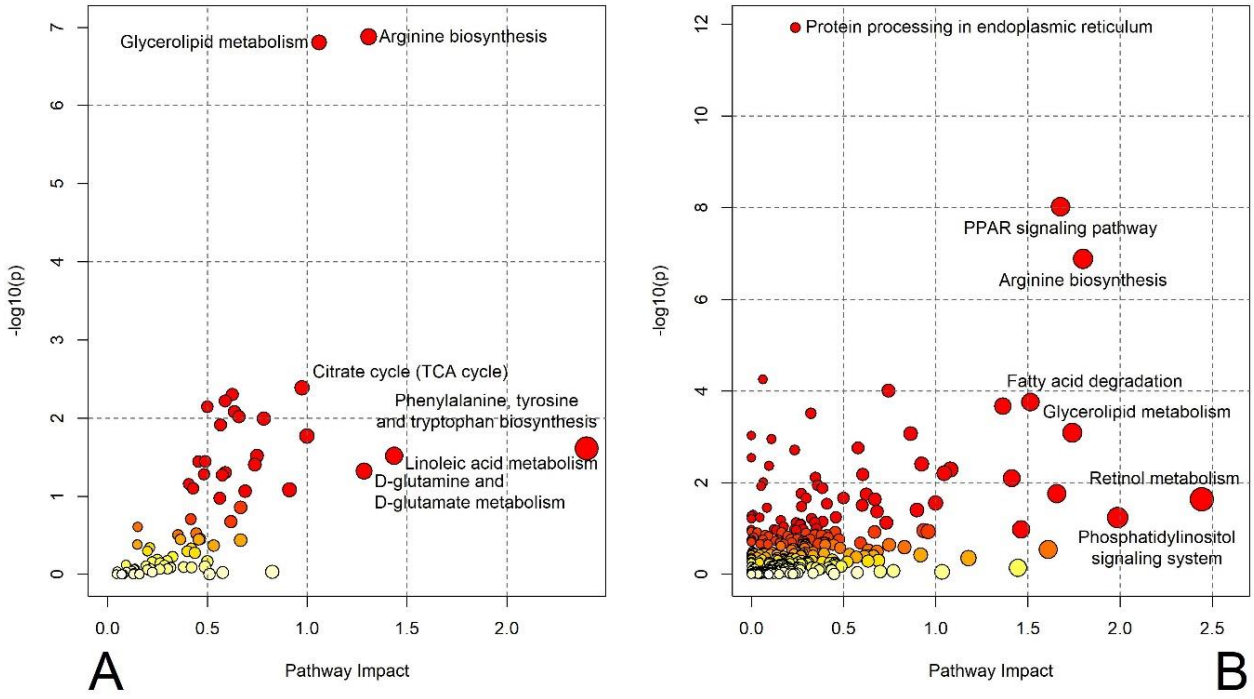

Figure 3: Bubble plots showing enriched metabolic (A) and all (B) pathways from a Joint Pathway Enrichment Analysis in MetaboAnalyst 4.0 based on DA plasma metabolites and DE genes in the hepatic transcriptome from cows with different nutrient partitioning phenotype. Bubble colour shifts from white to red with increasing -$\log_{10}$(p-value) and bubble size increases with pathway impact.

3.4 Correlations of Plasma Metabolites and Loci

Due to the known species-specificity of lncRNAs and the scarce knowledge on lncRNAs in non-model organisms represented in public data bases, we also followed an agnostic, purely data guided pipeline complementary to the (KEGG) pathway based, knowledge-based data analysis to set up metabogenomic networks of DA metabolites, DE genes and identified lncRNAs. For this, correlation coefficients were calculated for the transcriptome and metabolome as follows: locus – locus correlations, locus – metabolite correlations, and metabolite – metabolite correlations. Out of the 185 DA metabolites, 40 were chosen for the correlation analysis (see Supplementary Table 4). Overall, the Pearson correlation analyses between 3,492 loci (1,866 DE genes and all 2,076 minimally expressed lncRNAs) and 40 selected DA metabolites yielded a total of 1,100,147 significant correlations (FDR $\leq$ 0.05) and of them 63,541 with a correlation coefficient $|r| \geq 0.8$ (Supplementary Table 7). Of all significant correlations, more than half of the connections (55.15%) were between loci that were classified as non-lncRNAs (Supplementary Table 8). Between the DA metabolites and lncRNAs a total of 8,300 significant correlations were calculated, of which 222 showed a higher strength (correlation coefficient $|r| \geq 0.8$), (Figure 4).

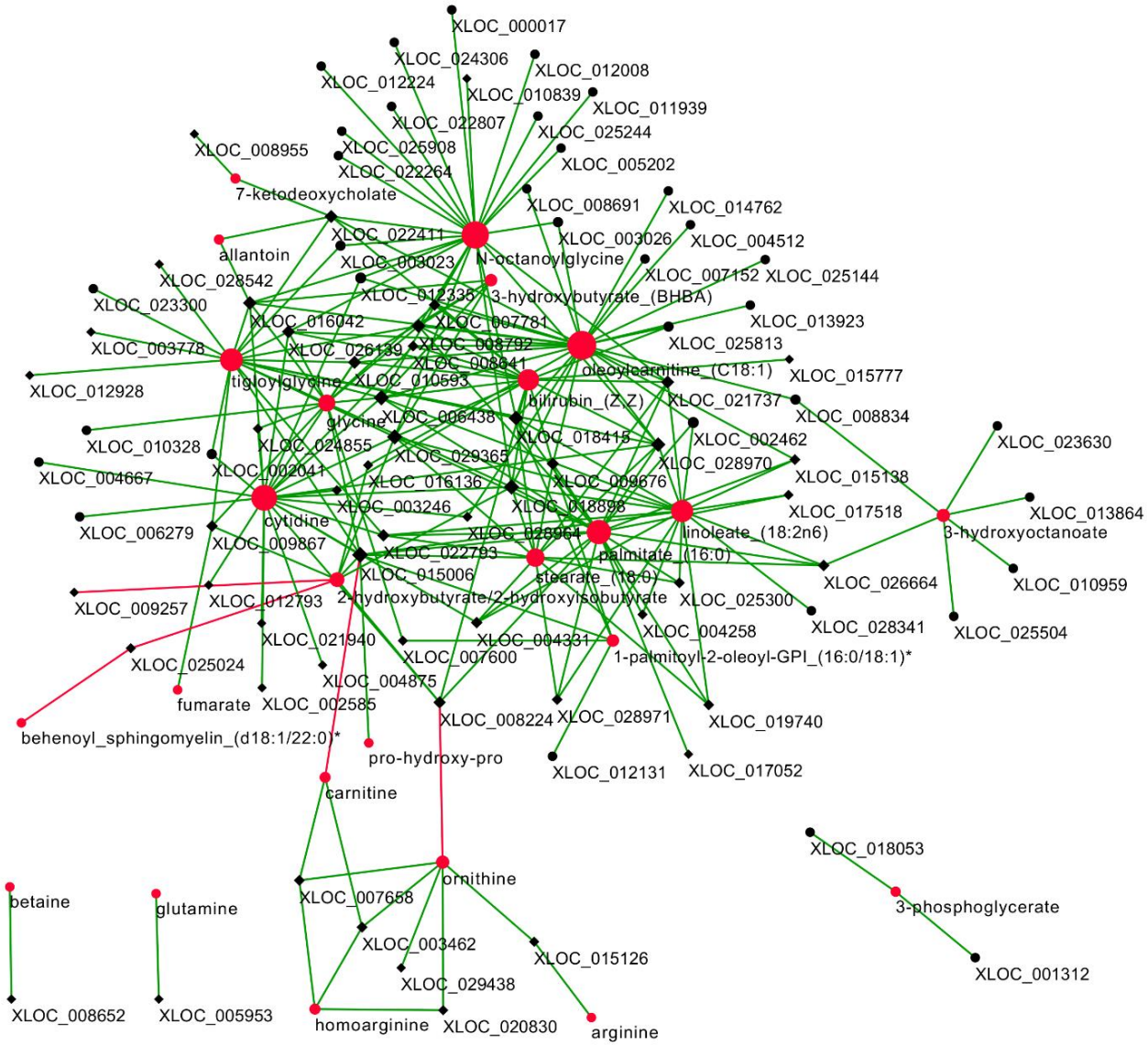

Figure 4: Interaction network between differentially abundant metabolites (red nodes) that were significantly and strongly (FDR ≤ 0.05, |r|≥ 0.8) correlated with lncRNAs (black nodes). LncRNAs that are significantly DE (FDR ≤ 0.05) between cows of different nutrient partitioning phenotype are highlighted in diamond shape. Node size reflects degree (number of connections to other nodes) and edge colour indicates correlation direction (positive = green, negative = red).

These strong correlations involved 25 metabolites and 86 lncRNAs. Among them were metabolites that had connections to over 10 lncRNAs: glycine, stearate (18:0), bilirubin (Z,Z), linoleate (18:2n6), tigloylglycine, palmitate (16:0), cytidine, N-octanoylglycine, and oleoylcarnitine (C18:1). The co-expression network visualization (Figure 4) highlights that lncRNAs in the center with higher connectivity, i.e. connections to multiple nodes, were also DE between the two phenotypic groups. The higher proportion of significant correlations between genes and metabolites compared to lncRNAs and metabolites might be due to the fact that genes had to be DE while lncRNAs only needed to pass the minimal expression threshold (with 247 lncRNAs being DE).

3.5 LncRNAs co-expressed with central players in lipid, glucose and energy metabolism may regulate hepatic glucose and lipid homeostasis and energy balance

Results from the Joint Pathway Enrichment and Network analysis indicated that the biological pathways associated with lipid metabolism belong to the major biological processes that differentiate the cow groups characterized by divergent nutrient partitioning (Figure 3). In the PPAR-signalling pathway, which is clearly enriched in the SEC cows (log(p) = 8.02, PI = 1.68), we found a number of DE genes encoding proteins with impact on fatty acid uptake and activation, intracellular fatty acid binding, mitochondrial and peroxisomal fatty acid oxidation, ketogenesis, triglyceride turnover, lipid droplet biology, gluconeogenesis and bile synthesis (Supplementary Table 3, Supplementary Figure 3). Among them are for example, *SLC27A1*, *FABP4, FABP5*, *ACSL1*, *CPT1A*, *ANGPTL4*, *CYP4A1, CYP4A22, PLIN4, PLIN2*, *APOA1, APOA5*, GK, *PCK1* and *RXRG*.

Due to its interactive function in PPAR signaling, particularly *RXRG* encoding the retinoic protein receptor G could play a critical upstream regulatory role in modulating the biological processes of lipid and glucose metabolism in the liver of SEC and ACC cows. Functionally, RXRG acts as a master regulator by producing various physiological effects by activating multiple nuclear receptor complexes. RXRG can regulate gene expression in a ligand-dependent manner and in cooperative interaction with nuclear transcription factors such as PPARA and PPARG. Their activation leads to a reduction of freely circulating triglycerides in the plasma and increases the uptake of free fatty acids and further impacts the glucose metabolism as well as the insulin sensitivity [35]. Park et al. [36] reported that RXRG may play an important role in tight control of glucose metabolism in the fasting/feeding cycle. In our study, *RXRG* displayed highly different hepatic transcript levels between both cow groups and showed in agreement with the enriched pathways exceptionally high wiring to levels of lncRNAs, DA metabolites, and DE genes involved in lipid metabolism. The 17 correlated DA metabolites (Supplementary Table 7), out of all 40 DA metabolites included in the correlation analysis, were, as expected, predominantly fatty acids (palmitate (16:0), stearate (18:0), linoleate (18:2n6), oleoylcarnitine (C18:1)), but also hydroxybutyrate and bilirubin (Z,Z). *RXRG* expression was strongly correlated with that of 796 protein coding genes (Supplementary Table 3), of which the highest correlated genes were those involved in PPAR signaling-mediated lipid, glucose and energy metabolism (see above). Further correlations were found with *MLYCD*, *SLC25A20* and *ACADVL* (fatty acid oxidation), *MPC1* and *PC* (Pyruvate and energy metabolism, TCA cycle/gluconeogenesis), *ELOVL5* (fatty acid (PUFA) biosynthesis), *SLC22A5* (carnitine transport), *LIPG* and *CREB3L3* (lipid metabolism), *ABCD3* (fatty acid transport and bile acid synthesis) and *FGF21* (metabolic regulation, see below).

Surprisingly, the highest correlation of *RXRG* expression level was found with *LDHA* expression together with a yet unannotated *LDHA* isoform transcript (ENSBTAG00000016688) and the unannotated lncRNA XLOC_019740, an antisense transcript to *LDHA.* All loci were differentially expressed between both cow groups, which indicates a link of *RXRG* to anaerobic glycolysis. In total, *RXRG* transcript level was correlated with that of 202 lncRNAs (FDR $\leq$ 0.05), 89 of which were DE between the cow groups. Inspection of their genomic position revealed that several lncRNAs are located adjacently and mostly in antisense direction to protein coding genes, which are known to be associated with PPAR signaling and to play important roles in glucose, lipid or energy metabolism. One example is the lncRNA XLOC_016136 with an expression highly correlated with that of *RXRG*. It is located antisense to *ELOVL5*, both potentially interacting loci were DE between cow groups. Another example is the natural antisense lncRNA XLOC_018898 to *ACSL1*. Both loci strongly correlated with *RXRG* at expression level and were DE between the two phenotypic groups. Of the remaining ones, the following lncRNAs also need to be mentioned: lncRNA XLOC_015138 with its potential antisense interaction partner gene *SLC25A47* and lncRNA XLOC_028970 that is located antisense to *MPC1*. As shown in Figure 4, lncRNA XLOC_028970 displays a central hub position in the interaction network between DA metabolites and lncRNAs. Based on the observed correlations to *RXRG* at expression level, the similarities in the differential hepatic expression of RXRG, lncRNAs and interacting partner genes and the differences in metabolite abundance between phenotypically different cow groups, it can be assumed

that the addressed lncRNAs could play a modulating, possibly fine-tuning role in the expression of their potential partner genes in conjunction with *RXRG* in the PPAR signaling pathway.

In our transcriptome study, *FGF21* was one of the most DE genes in the liver of cows differing in their nutrient partitioning phenotype (FDR = 8.51E-16; $\log_2$(FC) = 5.8, see Supplementary Figure 2). As mentioned above, *FGF21* expression level showed a significant correlation to *RXRG* expression (|r| 0.76, FDR = 6.7E-04). In our metabogenomics analysis, *FGF21* transcript level was significantly correlated with the abundance of 24 out of 40 of the DA metabolites included in the correlation analysis. Among them are those metabolites with the highest connectivity in the lncRNA-metabolite co-expression network (e.g., oleoylcarnitine (C18:1), palmitate (16:0), bilirubin (Z,Z), linoleate (18:2n6), stearate (18:0)). Except for ornithine, arginine and carnitine, all metabolites displayed a positive correlation with *FGF21* transcript abundance (see Supplementary Table 9). In addition, *FGF21* transcript level showed significant correlations with multiple DE transcript loci of the liver transcriptome (n = 1.370, FDR ≤ 0.05) including numerous genes involved in fatty acid and lipid metabolism (e.g., *APOA1, LIPG, ELOVL2, GPAT3*) and also with the expression level of *CREB3L3,* which encodes a transcription factor controlling *FGF21* expression and its plasma level in cooperation with PPARA (e.g., [37]). Of the DE loci correlating to *FGF21* expression level are 410 lncRNAs, comprising more than the average number of significant correlations observed in our study. Strong transcript correlations (|r| ≥ 0.8) were observed for 29 lncRNAs out of a list with 131 loci in total (Supplementary Table 6).

The lncRNA XLOC_010660, which is overlapping with and antisense to *FGF21*, displayed the strongest correlation to *FGF21* transcript abundance (Figure 5, |r| = 0.98), which indicates its regulatory potential as natural antisense transcript to mediate the expression of its parental gene. Positionally related lncRNAs are also deposited for the orthologous human *FGF21* gene in the database for human lncRNAs, LNCipedia v. 5.2 (https://lncipedia.org/db/), (e.g. lnc-FGF21-2). The bovine antisense *FGF21*-lncRNA XLOC_010660 showed a higher hepatic transcript level in SEC type cows and represented, similar to the *FGF21* gene, one of the loci with the strongest differential expression levels in the hepatic transcriptome between the cow groups.

Apart from XLOC_010660, the lncRNA XLOC_007781 was also highly correlated with *FGF21* (Figure 5), although it is located distantly from *FGF21*: intergenic, antisense to and 5' upstream of the transcription start site of *SLC25A33*. The expression level of *SLC25A33* was not only strongly correlated with those of XLOC_007781 and *FGF21* but also highly different between the SEC and ACC cows indicating a metabolic link between these loci. The *SLC25A33* gene encodes a mitochondrial carrier protein that promotes cell growth and survival by controlling the mitochondrial genome and by preventing mitochondrial dysfunction [38]. Upon insulin or IGF1 stimulation, SLC25A33 as pyrimidine (deoxy-) nucleotide transporter, is known to be involved in the regulation of cell growth and proliferation by controlling mitochondrial DNA replication and transcription [39,40]. In our study, the lncRNA XLOC_007781, putatively *cis*-interacting with *SLC25A33* or possibly *trans*-interacting with *FGF21,* is significantly higher expressed in the SEC cows compared to their counterparts, which displayed almost no lncRNA XLOC_007781 expression. Remarkably, this DE lncRNA holds one of the central hub positions in the metabolite-lncRNA interaction network underlining its potential critical role associated with the metabolic challenge of the liver of cows in early lactation. This hypothesis is also supported by the observation that this lncRNA is only very marginally expressed in the liver of bulls from the same experimental crossbred population [22].

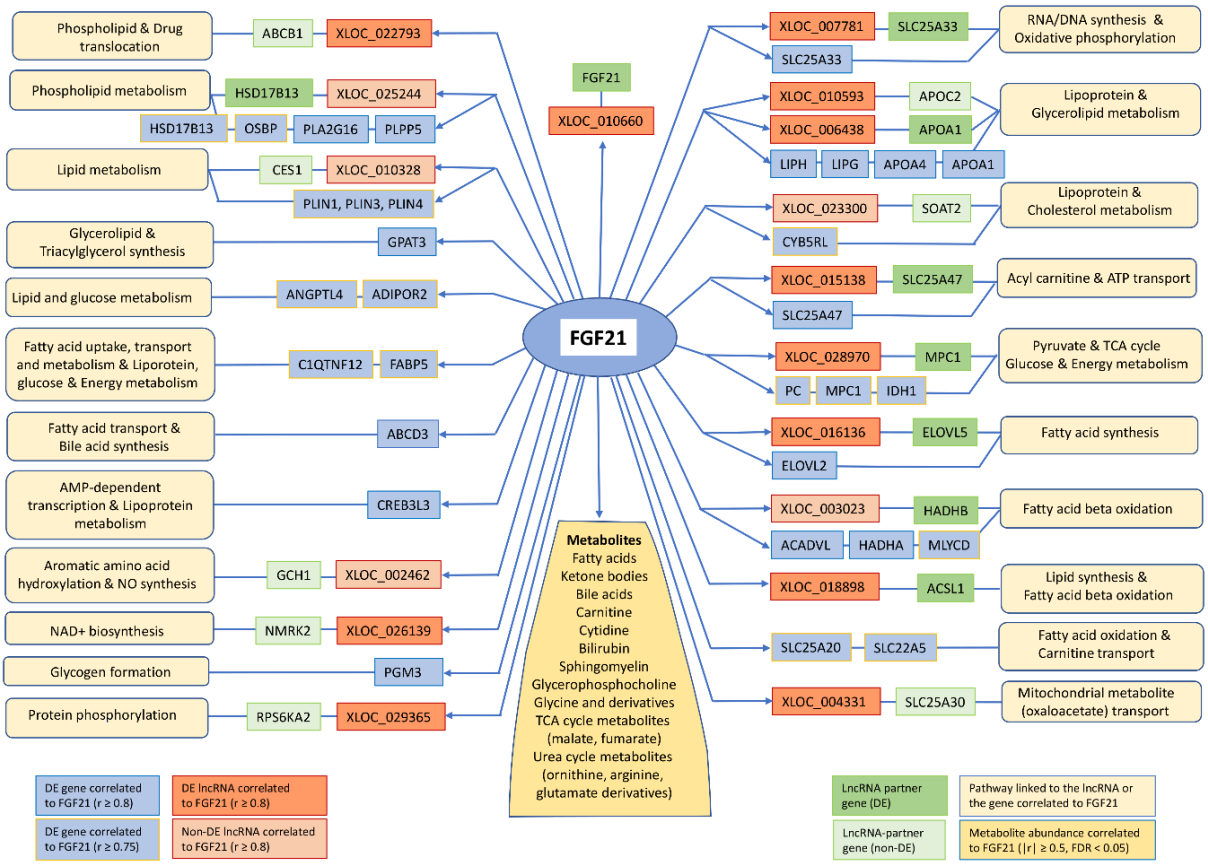

Figure 5: Correlation of the *FGF21* expression level with hepatic transcript levels of lncRNAs and protein coding genes involved in lipid, glucose and energy metabolism pathways and with the abundance of plasma metabolites. DE gene: gene differentially expressed between SEC and ACC cow groups at FDR < 0.05; DE lncRNA or non-DE lncRNA: lncRNA differentially or not differentially expressed between SEC and ACC cow groups at FDR < 0.05.

Central hub positions in the metabolite-lncRNA interaction network (Figure 4) are represented by further DE lncRNAs, whose transcript levels are strongly correlated with those of *FGF21*, for example XLOC_015138, XLOC_023300, XLOC_006438, XLOC_010593 and XLOC_028970 (Figure 5). Particularly XLOC_028970 that is antisense to the *MPC1* gene and also correlated with *RXRG* expression level seems to play a central role in this network. The MPC1 protein is responsible for transporting pyruvate into mitochondria for fatty acid oxidation in the TCA cycle and is required for an efficient gluconeogenesis regulation (e.g., [41]). Also, the other four lncRNAs have potentially *cis*-interacting partner genes in genomic antisense orientation to *SLC25A47, OAT2, APOA1* and *APOC2,* respectively, with important roles in energy and lipid metabolism.

The FGF21 protein, known as a hepatokine with pleiotropic metabolic effects, which regulates glucose and lipid metabolism in the liver and energy balance and metabolism in mammals and modulates many pathways in multiple target tissues in metabolically compromised animals and humans (e.g., [42]). Dietary imbalances such as nutrient deprivation (fasting or starvation), ketogenic or high carbohydrate diets, overfeeding, protein restriction and conditions associated with metabolic stress such as physical exercise or metabolic diseases (e.g., type 2 diabetes, nonalcoholic fatty liver disease and obesity) trigger FGF21 expression and signaling. It has been reported that FGF21 is a physiological regulator essential both for energy balance in the baseline state and for the adaptation to changes in dietary imbalances in human and rodent species [43,44]. It seems to be the key signal that communicates and coordinates the metabolic response to reverse different nutritional stresses and restores metabolic homeostasis [45]. In early lactating dairy cows, FGF21 has been found to be associated with lactation performance and was identified as a sensitive biomarker for detecting and monitoring ketosis [46-49]. Furthermore, it has

been observed that administration of FGF21 to energy-deficient, early lactating dairy cows resulted in lower triglyceride levels in liver biopsies, and plasma free fatty acid concentrations tended to be lower [50]. In our study, a higher hepatic *FGF21* transcript level in the SEC cows was accompanied by a higher plasma level of fatty acids that are known to induce FGF21 expression in the liver in order to restore the metabolic homeostasis.

Based on the results from our study and from literature, it is conceivable that the *FGF21* gene and associated lncRNAs (Figure 5) together with their interacting partner genes can modulate and fine-tune energy balance as well as glucose and lipid homeostasis in the liver of cows challenged by changes in energy demands and nutrient conditions or which differ with regard to their physiological priority for nutrient partitioning.

Another gene most differentially expressed in the liver between SEC and ACC cow groups (Supplementary Figure 2, Supplementary Table 3) was *MFSD2A.* It had a higher transcript level in SEC compared to ACC cows that correlated with a substantial number of transcript levels of DE genes functionally involved in fatty acid and lipid metabolism, such as *MPC1, ANGPTL4, ADIPOR2, SLC25A34, PLIN1, SLC25A47* and *ACSL1* as well as with those of *RXRG* and *FGF21*. The *MFSD2A* gene is known to be fasting-induced and regulated by both PPARA and glucagon signaling in the liver. The encoded protein, a sodium-dependent lysophosphatidylcholine and plasma membrane transporter for omega-3 fatty acids, has a regulatory function in growth, development and lipid metabolism [51]. In our study, *MFSD2A* expression levels correlated with several plasma metabolite levels (palmitate (16:0), linoleate (18:2n6), glycine, tigloylglycine, stearate (18:0), bilirubin (Z,Z) and carnitine (Supplementary Table 7).

At the top of the lncRNAs with transcript levels highly correlated with that of *MFSD2A was* the intergenic lncRNA XLOC_018415 that possesses a hub position in the metabolite-lncRNA interaction network and is functionlessly annotated in the bovine genome as ENSBTAG00000052009. Additional strong correlations were found for a number of lncRNAs, e.g., XLOC_022793, XLOC_004331, XLOC_028970, which are all antisense located to putative *cis*-interaction genes (*ABCB1*, *SLC25A30, MPC1*) associated with pathways included in lipid and energy metabolism. A pair of lncRNAs (XLOC_029365 and XLOC_028971) with strong correlation to *MFSD2A* at transcript level is positionally co-located with the *RPS6KA2* gene, which encodes a kinase that has been implicated in controlling of cell growth and differentiation. Remarkably, XLOC_028971 is the most differentially expressed lncRNA between the cow groups in our study and the transcript levels of both lncRNAs were also highly correlated with that of the *RXRG*, *FGF21* and *ANGPTL4* genes playing crucial roles in metabolic regulation.

The expression of *ANGPTL4* showed a higher level in the SEC group and was strongly positively correlated with circulating fatty acids and derivates (stearate 18:0, palmitate 16:1, linoleate 18:2n6), oleoylcarnitine (C18:1), 1-palmitoyl-2-oleoyl-GPI (16:0/18:1)), and 3-hydroxybutyrate (BHBA), 2-hydroxybutyrate/2-hydroxyisobutyrate as well as with expression levels of genes associated with fatty acid oxidation, energy and lipid metabolism, such as *SLC25A47, PLIN2, MPC1, SLC22A5, SLC25A34, SLC45A3, SLC25A33* and *LDHA*. The ANGPTL4 protein is known to mediate the inactivation of LPL [52], and thereby plays a role in regulating triglyceride clearance in blood serum and lipid metabolism [53]. Functionally known as adipokine and hepatokine, it also regulates glucose homeostasis and insulin sensitivity [54]. A decreased *ANGPTL4* expression has been found to be associated with type 2 diabetes [55] and thus, it might be a branch point in nutrient partitioning. Wang et al. [56] found that cows with clinical ketosis and fatty liver had significantly higher serum ANGPTL4 concentrations than healthy cows and suggested that it could play an important role in adjusting energy metabolism deregulated in dairy cows during the peripartum period. In our study, we detected several lncRNAs that were highly correlated with the *ANGPTL4* gene expression level suggesting a potential regulatory fine-tuning function for them.

Most highly correlated with *ANGTL4* and with a higher transcript level in the SEC cows was lncRNA XLOC_022793, which is antisense to the *ABCB1* gene. The ABCB1 protein is a multidrug and lipid translocase with broad specificity [57], e.g., glycosphingolipids and membrane phospholipids [58].

ABCB1 expression is predominantly regulated at the transcriptional level. In addition, lncRNAs XLOC_019740, XLOC_004331, XLOC_015138 and XLOC_028970, which all have already been mentioned in our study (see above), have antisense partner genes (*LDHA, SLC25A30, SLC25A47* and *MPC1*) involved in pathways of energy metabolism, showed strong correlations to *ANGPTL4* at transcription level. The role of *ANGPTL4* in mediating the cross talk between metabolic syndromes, such as diabetes and obesity and cancer by modulating its expression by PPARs, has been discussed [59]. However, a regulatory fine-tuning role of lncRNAs should be also considered in this context and when elucidating its role in phenotype-driven interpretation of the metabogenomic data characterizing the different cow groups in our study.

The expression levels of the genes *ANGTL4, MFSD2A, RXRG* and *FGF21* highlighted in this study correlated with plasma levels of linoleate (18:2n6), palmitate (16:0) and cytidine (Supplementary Table 7). These metabolites also held hub positions with high connectivity to other DA metabolites as well as a number of lncRNAs in the metabolite-lncRNA interaction network (Figure 4). It is noteworthy that lncRNAs with central network positions that are connected to these metabolites were also DE: e.g. XLOC_028970 (antisense lncRNA of *MPC1*), XLOC_018898 (antisense lncRNA of *ACSL1*) and XLOC_006438 (antisense lncRNA of *APOA1*). The strong and numerous correlations of predominantly DE lncRNAs to DA metabolites as well as DE genes with relation to lipid metabolism strongly suggest their functional involvement in this biological pathway.

3.6 LncRNAs involved in linking urea cycle, TCA cycle and gluconeogenesis

Besides long-chain fatty acids, main metabolite nodes in the MetaboAnalyst subnetwork 1 comprise arginine and ornithine (including also other members of the arginine metabolism, such as homo-L-arginine, 2-oxoarginine and glycine). The high impact of these metabolites on the metabolic differentiation between cow groups is also confirmed by the results of the top enriched biological pathways when analyzing all DA metabolites and the joint list of all DA metabolites and DE genes, which together highlighted the pathway of arginine biosynthesis. Closer inspection of metabolites and genes specified that a large number of them is particularly involved in the urea cycle and linked to it, the TCA cycle and the gluconeogenesis pathways (Figure 6). The ACC cows displayed an increased abundance of a large number of DA amino acids in plasma compared to SEC cows. This was also true for ketogenic amino acids, but a few glucogenic amino acids showed the same pattern. Except for glycine, their abundance was consistently higher in the ACC group. It can be hypothesized that this is the effect of a limited amino acid flux into the mammary gland for milk protein synthesis in these cows. The increased levels of ketogenic amino acids (Tyr, Phe, Lys, Leu) together with the increased expression of genes encoding enzymes from the urea cycle, suggest the hypothesis that amino acids were used for oxidation within the TCA cycle. ACC cows can be assumed to be in a positive energy balance and would have no need for substrate supply for additional energy production and do not mobilize fat from their body fat depots. In consequence, the surplus nitrogen (N) from the diet [60], indicating a positive N-balance in beef cows fed a dairy diet at the beginning of lactation had to be detoxified resulting in elevated activity of the urea cycle. This is demonstrated e.g. by highly significantly elevated arginine and ornithine abundance and higher *CPS1, OTC* and *ASS1* expression in ACC cows. Interestingly, N-acetylglutamate is significantly decreased in ACC cows, although it is regarded as an essential cofactor for CPS1 for the synthesis of carbamoyl phosphate from ammonium and bicarbonate generated by protein catabolic processes.

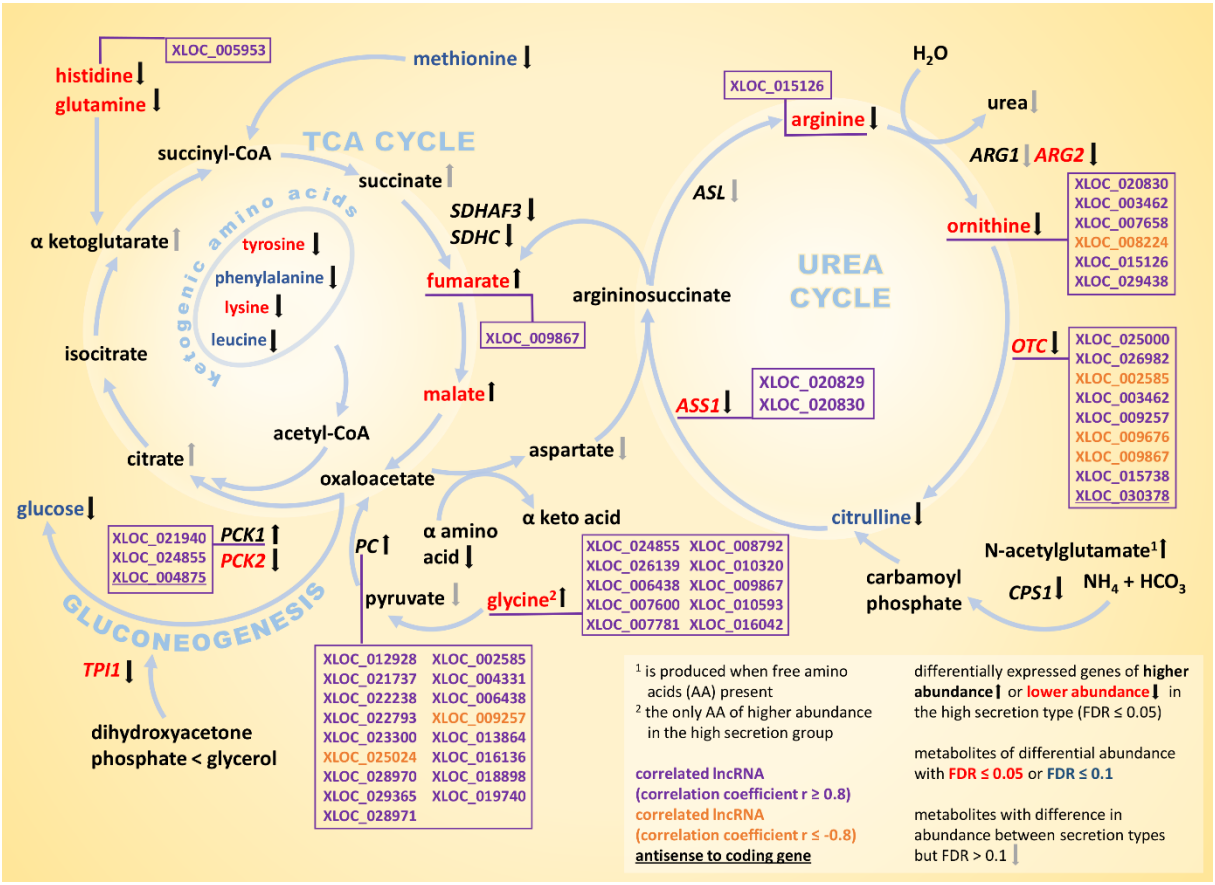

Figure 6: Urea Cycle, TCA cycle and gluconeogenesis with differentially expressed genes (liver) and differentially abundant metabolites (plasma) of cows with different nutrient partitioning phenotype. LncRNAs with significant (FDR ≤ 0.05) and strong correlations ($|r| \geq 0.8$) of their transcript levels with those of genes and or metabolite abundance involved in the cycles are depicted.

Bobe et al. [61] had suggested a regulation of ureagenic and gluconeogenic genes in dairy cows via glucagon due to their increased expression after external glucagon application. Glucagon is known to stimulate lipolysis, which fits the increased levels of long-chain fatty acids and hepatic ketogenesis as indicated by elevated BHBA as observed in plasma of SEC cows. In addition, glucagon has been shown to increase hepatic *FGF21* expression in cows [50], which is in line with its higher expression in the SEC group and would indeed suggest a modulated action of glucagon in those cows. Also elevated expression of genes coding for key enzymes of gluconeogenesis (*PC*, *PCK1*) in the SEC group supported the postulated glucagon action. But, this would be in contrast to the observed lower expression of ureagenic genes (*OTC, ASS1, CPS1*) in the SEC compared to ACC cows, which in turn appeared to be in contrast to the significantly higher N-acetylglutamate plasma concentration in this group. However, recently Galsgaard et al. [62] confirmed that glucagon receptor-mediated activation of ureagenesis is not required when N-acetylglutamate levels are sufficient to activate the first step of the urea cycle. Instead of mirroring the level of free amino acids, the elevated N-acetylglutamate concentration could be due to the energetic challenge in the SEC group. In calorie-restricted fed mice, Yanckello et al. [63] observed higher levels of N-acetylglutamate.

Thus, other regulators than the glucagon effects in early lactating dairy cows described above seem to be involved in fine-tuning the expression of ureagenic genes in our crossbred cattle population. In this context, we observed a large number of lncRNAs, whose expression levels correlated with those of members of the urea cycle, in particular with that of *OTC*. For this gene, which is central in the urea cycle, a large number of correlated and DE lncRNAs was observed (Figure 6).

For example, the lncRNA XLOC_009676 expression was negatively correlated with *OTC* expression, and this lncRNA is antisense and negatively correlated with *PEPD*, which functionally contributes to dipeptide degradation and was significantly higher expressed in the ACC group. The expression of lncRNA XLOC_009867 was also negatively correlated with *OTC* and is located intergenic within the *APOE/C4/C1/-C2* gene cluster. This lncRNA is already annotated in the Ensembl database as ENSBTAG00000053946, although it has not yet been assigned a functional role. The protein coding *APOC1* gene, which is missing in the bovine genome, is located at the corresponding syntenic position in the human genome. The expression of lncRNA XLOC_003462, also already annotated as ENSBTAG000000000051368 albeit without an assigned function, was significantly and positively correlated with the *OTC* expression. This lncRNA is interesting because genomically, no protein coding gene is localized within 200 kb, which could possibly indicate a potential target for *cis*-regulation. The strongest positive correlation of lncRNAs with *OTC* at the transcriptional level was observed for lncRNA XLOC_026982, located 5' to *SLC25A23*, which encodes a mitochondrial solute carrier responsible for shuttling of metabolites, nucleotides and cofactors through the mitochondrial inner membrane [64] and is nominally ($p < 0.05$) lower expressed in the SEC group.

The energy metabolism, namely TCA cycle and gluconeogenesis pathways (Figure 6), are tightly linked to the urea cycle via fumarate and also displayed a large number significantly DA metabolites and DE genes in our study. It has to be remembered that the overall performance level of the cows under investigation, even in the SEC group, was substantially below the average-performance of dairy cows and that cows from both groups displayed highly elevated levels of body fat deposition compared to early lactating dairy cows. Increased levels of fumarate, malate and increased expression of the *PCK1* gene indicate that the SEC cows promote gluconeogenesis compared to the ACC group, which is also documented e.g., by increased *PC* expression. These findings are in line with previous studies of trait-differentiated cows of the same population, which demonstrated significant differences in glucose metabolism [24].

In our study, the transcript levels of genes encoding key enzymes of gluconeogenesis (*PCK1, PC*) and energy metabolism via TCA cycle (*PC*) displayed a large number of significant and strong correlations to the expression of several lncRNAs suggesting their functional role in these pathways. The intergenic and significantly DE lncRNA XLOC_025024 has no annotation in the bovine genome up to now, but the corresponding syntenic position in the human genome carries the annotated lncRNA LINC02473 (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000001405.39, GRch38.p13). Besides *PC* expression, the XLOC_025024 transcript level was significantly correlated with 1,137 DE loci and with the abundance of 33 of the 40 plasma metabolites included in the correlation analysis. This suggests that this lncRNA may be deeply involved in the modulation of biological processes linked to trait-differentiated phenotypes of the two cow groups. Analogously, lncRNA XLOC_012928, besides being positively correlated with *PC,* displayed also a significant correlation to large number of DE genes (1,042) and DA metabolites (31). Of particular interest is the second highest correlation to *SLC25A20.* The gene encodes a carrier protein mediating the transport of acylcarnitines for mitochondrial fatty acid oxidation [65] and displayed a significantly higher expression in the SEC compared to ACC cows. Another lncRNA positively correlated with *PC* transcript abundance was XLOC_019740. This lncRNA is antisense to *LDHA* that displayed a significantly higher expression in the SEC group, was highly correlated to RXRG expression level (see above), and the corresponding protein converts lactate to pyruvate. Finally, the *PC* transcript level is also positively correlated with lncRNA XLOC_028970, which was significantly higher expressed in the SEC group analogous to its antisense protein coding gene *MPC1*. The protein encoded by *MPC1* has a key function in the transport of pyruvate into the mitochondrion [66], which highlights the potential functional link between the lncRNA XLOC_028970 and gluconeogenesis and TCA cycle via *PC*. Due to the correlation with many genes in the PPAR signaling pathway including *FGF21* (see above), this lncRNA represents a link between the TCA cycle, gluconeogenesis and lipid metabolism.

4. Conclusions

In our study, we used enrichment, network and correlation analyses of our metabolomics and transcriptomics data to reduce the complexity of the huge data and to improve understanding and interpretation of biological systems driving phenotypic differentiation of metabolic phenotypes of nutrient partitioning in lactating cows. To generate hypotheses for the putative involvement of lncRNA modulation in these systems and in the expression of the phenotypic divergency of the cow groups analysed, we focused on the most outstanding observed interconnections. Metabogenomic analyses (merging metabolomics and transcriptomic data) indicated an intricate role of lncRNAs in the regulatory network, modulating nutrient partitioning in lactating cows. Given our data, intergenic as well as natural antisense lncRNAs can be assumed to exert *cis*-regulatory effects on genes involved in lipid and amino acid metabolism, energy metabolism and detoxication of nitrogen via the urea cycle. LncRNAs with hub positions in metabogenomic networks identified in our study should be investigated in more detail in the future and should be considered more closely in the analysis of the physiological adaptation and changes in the metabolic state of lactating cows.

5. Data Availability Statement

The transcriptomic data set examined in this study was already used in a previous study ([27], aligned to UMD.3.1, Ensembl annotation release 92) and is stored in the Functional Annotation of Animal Genomes (FAANG) database (https://data.faang.org/dataset) under project number PRJEB34570. The project specific annotation of the transcriptomic data (gtf file) is available as Supplementary Data 1.

6. Ethics Statement

All experimental procedures were carried out according to the German animal care guidelines and were approved and supervised by the relevant authorities of the State Mecklenburg-Vorpommern, Germany (State Office for Agriculture, Food Safety and Fishery; LALLF M-V/TSD/7221.3-2.1-010/03).

7. Supplementary Materials

Supplementary Figure 1: Volcano plot of differentially abundant plasma metabolites for cows characterized by predominant nutrient partitioning into milk secretion (SEC, n = 13) and body fat accretion (ACC, n = 12) with upregulation (higher abundance) in the SEC type cows labelled blue and downregulation (lower abundance) labelled green. Significance threshold (horizontal dotted line) at q (Benjamini-Hochberg) ≤ 0.05.

Supplementary Figure 2: Volcano plot for differentially expressed loci between cows characterized by predominant nutrient partitioning into milk secretion (SEC, n = 13) and body fat accretion (ACC, n = 12) with upregulation (higher abundance) in SEC type cows on the right side and downregulation (lower abundance) on the left side. Significance threshold (horizontal dotted line) at q (Benjamini-Hochberg) ≤ 0.05. Labels are displayed for loci with an absolute $\log_2$ fold change ≥ 1 and q ≤ 0.05.

Supplementary Figure 3: Genes differentially expressed between SEC and ACC cow groups at FDR < 0.05 (DE) and associated with a biological process associated with the PPAR signal and their correlation with *RXRG* expression at the transcriptional level in the liver (FDR < 0.05)

Supplementary Table 1: Phenotypic characteristics of cows with different nutrient partitioning phenotype

Supplementary Table 2: RNA sequencing, alignment, and mapping statistics

Supplementary Table 3: Results of the differential hepatic expression analysis for minimally expressed loci (≥ 0.1 FPKM for at least six cows of one phenotypic group) in liver with information on position and structure of loci and transcripts, including predicted long non-coding RNAs and their positional partner genes.

Supplementary Table 4: Metabolites that were differentially abundant (FDR ≤ 0.05) between cows of with different nutrient partitioning phenotype. Those metabolites selected for a Pearson correlation analysis with differentially expressed genes and lncRNAs in the liver transcriptome are marked with 'x'.

Supplementary Table 5: Significant KEGG pathway enrichment analysis results of Metaboanalyst subnetworks based on genes differentially expressed in the liver and differentially abundant plasma metabolites between cows with different nutrient partitioning phenotype.

Supplementary Table 6: Joint Pathway Analysis from differentially expressed genes and differentially abundant metabolites (FDR ≤ 0.05) between cows of different nutrient partitioning phenotype using joint omics-data analysis with MetaboAnalyst 4.0 for metabolic and all pathways (based on KEGG version October 2019).

Supplementary Table 7: Correlations between differentially expressed genes (n = 1,866), lncRNAs (n = 2,076) and 40 selected differentially abundant metabolites (FDR ≤ 0.05) with overall number of significant Pearson correlations (FDR ≤ 0.05) in brackets and number of significant correlations with correlation coefficient |r|≥ 0.8 above.

Supplementary Table 8: Correlations between differentially expressed genes (n = 1,866), lncRNAs (n = 2,076) and 40 selected differentially abundant metabolites with overall number of significant Pearson correlations (FDR ≤ 0.05) in brackets and number of significant correlations with correlation coefficient |r|≥ 0.8 above.

Supplementary Table 9: Blood plasma metabolite abundance correlated with gene expression of *FGF21* in liver (FDR ≤ 0.05).

Supplementary Data 1: Project specific annotation of the transcriptomic data (gtf file). For details on establishing the annotation see [22].

8. Acknowledgments

This study was funded by the German Research Foundation (DFG grant numbers: KU 771/8-1 and WE 1786/5-1). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668. This work was supported by European COST action FAANG-Europe CA15112. The authors thank Frieder Hadlich for his support in bioinformatics issues, as well as Simone Wöhl and Bärbel Pletz for the excellent lab work.

9. Author Contributions

Conceptualization, C.K. and R.W.; Methodology, C.K., R.W. and W.N.; Software, W.N.; Investigation, W.N., C.K. and R.W.; Resources, C.K., H.M.H., R.M.B., E.A.; Data Curation, W.N., C.K.; Writing – Original Draft Preparation, W.N., R.W. and C.K.; Writing – Review & Editing, all authors; Visualization, W.N.; Supervision, C.K. and R.W.; Project Administration, C.K. and R.W.; Funding Acquisition, C.K. and R.W.

10. Conflict of Interest

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

11. References

[1] C.I. Brannan, E.C. Dees, R.S. Ingram, S.M. Tilghman, The product of the H19 gene may function as an RNA, Mol. Cell. Biol., 10 (1990) 28-36.
[2] G. Borsani, R. Tonlorenzi, M.C. Simmler, L. Dandolo, D. Arnaud, V. Capra, M. Grompe, A. Pizzuti, D. Muzny, C. Lawrence, H.F. Willard, P. Avner, A. Ballabio, Characterization of a murine gene expressed from the inactive X chromosome, Nature, 351 (1991) 325-329.
[3] N. Brockdorff, A. Ashworth, G.F. Kay, P. Cooper, S. Smith, V.M. McCabe, D.P. Norris, G.D. Penny, D. Patel, S. Rastan, Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome, Nature, 351 (1991) 329-331.
[4] C.J. Brown, B.D. Hendrich, J.L. Rupert, R.G. Lafrenière, Y. Xing, J. Lawrence, H.F. Willard, The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus, Cell, 71 (1992) 527-542.
[5] M. Sun, W.L. Kraus, From discovery to function: The expanding roles of long noncoding RNAs in physiology and disease, Endocrine Rev., 36 (2015) 25-64.
[6] J.L. Rinn, H.Y. Chang, Genome regulation by long noncoding RNAs, Annu. Rev. Biochem., 81 (2012) 145-166.
[7] F.P. Marchese, I. Raimondi, M. Huarte, The multidimensional mechanisms of long noncoding RNA function, Genome Biol., 18 (2017) 13-13.

[8] N. Gil, I. Ulitsky, Regulation of gene expression by cis-acting long non-coding RNAs, Nat. Rev. Genet., 21 (2020) 102-117.
[9] M. Magistri, M.A. Faghihi, G. St Laurent, C. Wahlestedt, Regulation of chromatin structure by long noncoding RNAs: Focus on natural antisense transcripts, Trends Genet., 28 (2012) 389-396.
[10] I. Sarropoulos, R. Marin, M. Cardoso-Moreira, H. Kaessmann, Developmental dynamics of lncRNAs across mammalian organs and species, Nature, 571 (2019) 510-514.

[11] R. Weikard, W. Demasius, C. Kuehn, Mining long noncoding RNA in livestock, Anim. Genet., 48 (2017) 3-18.
[12] B. Kosinska-Selbi, M. Mielczarek, J. Szyda, Review: Long non-coding RNA in livestock, Animal, 14 (2020) 1-11.
[13] L.H. Baumgard, R.J. Collier, D.E. Bauman, A 100-Year Review: Regulation of nutrient partitioning to support lactation, J. Dairy Sci., 100 (2017) 10353-10366.
[14] J.J. Gross, R.M. Bruckmaier, Review: Metabolic challenges in lactating dairy cows and their assessment via established and novel indicators in milk, Animal, 13 (2019) S75-S81.
[15] J.J. Loor, M. Bionaz, J.K. Drackley, Systems physiology in dairy cattle: nutritional genomics and beyond, Annu. Rev. Anim. Biosci., 1 (2013) 365-392.
[16] J.W. McFadden, Review: Lipid biology in the periparturient dairy cow: contemporary perspectives, Animal, 14 (2020) s165-s175.
[17] J.J. Loor, Genomics of metabolic adaptations in the peripartal cow, Animal, 4 (2010) 1110-1139.
[18] R. Liang, B. Han, Q. Li, Y. Yuan, J. Li, D. Sun, Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver, Sci. Rep., 7 (2017) 6396-6396.
[19] X. Zheng, C. Ning, P. Zhao, W. Feng, Y. Jin, L. Zhou, Y. Yu, J. Liu, Integrated analysis of long noncoding RNA and mRNA expression profiles reveals the potential role of long noncoding RNA in different bovine lactation stages, J. Dairy Sci., 101 (2018) 11061-11073.
[20] B. Yang, B. Jiao, W. Ge, X. Zhang, S. Wang, H. Zhao, X. Wang, Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period, BMC Genomics, 19 (2018) 605-605.
[21] R. Weikard, F. Hadlich, H.M. Hammon, D. Frieten, C. Gerbert, C. Koch, G. Dusel, C. Kuehn, Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets, Oncotarget, 9 (2018) 21052-21069.
[22] W. Nolte, R. Weikard, R.M. Brunner, E. Albrecht, H.M. Hammon, A. Reverter, C. Kühn, Identification and annotation of potential function of regulatory antisense long non-coding RNAs related to feed efficiency in bos taurus bulls, Int. J. Mol. Sci., 21 (2020) 3292.
[23] C. Kühn, O. Bellmann, J. Voigt, J. Wegner, V. Guiard, K. Ender, An experimental approach for studying the genetic and physiological background of nutrient transformation in cattle with respect to nutrient secretion and accretion type, Arch. Anim. Breed., 45 (2002) 317-330.
[24] H.M. Hammon, C.C. Metges, A. Schulz, P. Junghans, J. Steinhoff, F. Schneider, R. Pfuhl, R.M. Bruckmaier, R. Weikard, C. Kühn, Differences in milk production, glucose metabolism, and carcass composition of 2 Charolais x Holstein F2 families derived from reciprocal paternal and maternal grandsire crosses, J. Dairy Sci., 93 (2010) 3007-3018.
[25] M. Mielenz, B. Kuhla, H.M. Hammon, Abundance of adiponectin system and G-protein coupled receptor GPR109A mRNA in adipose tissue and liver of F2 offspring cows of Charolais × German Holstein crosses that differ in body fat accumulation, J. Dairy Sci., 96 (2013) 278-289.
[26] R. Weikard, T. Goldammer, R.M. Brunner, C. Kuehn, Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows, Physiol. Genomics, 44 (2012) 728-739.
[27] W. Nolte, R. Weikard, R.M. Brunner, E. Albrecht, H.M. Hammon, A. Reverter, C. Kühn, Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle, Front. Genet., 10 (2019).
[28] Y. Liao, G.K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics, 30 (2014) 923-930.
[29] V. Wucher, F. Legeai, B. Hédan, G. Rizk, L. Lagoutte, T. Leeb, V. Jagannathan, E. Cadieu, A. David, H. Lohi, S. Cirera, M. Fredholm, N. Botherel, P.A.J. Leegwater, C. Le Béguec, H. Fieten, J. Johnson, J. Alföldi, C. André, K. Lindblad-Toh, C. Hitte, T. Derrien, FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome, Nucleic Acids Res., 45 (2017) e57-e57.
[30] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol., 15 (2014) 550-550.
[31] J. Chong, D.S. Wishart, J. Xia, Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis, Curr. Protoc. Bioinformatics, 68 (2019) e86-e86.
[32] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A.

Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: The human metabolome database for 2018, Nucleic Acids Res., 46 (2018) D608-D617.
[33] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, M. Tanabe, New approach for understanding genome variations in KEGG, Nucleic Acids Res., 47 (2019) D590-D595.
[34] C. Schäff, S. Börner, S. Hacke, U. Kautzsch, H. Sauerwein, S.K. Spachmann, M. Schweigel-Röntgen, H.M. Hammon, B. Kuhla, Increased muscle fatty acid oxidation in dairy cows with intensive body fat mobilization during early lactation, J. Dairy Sci., 96 (2013) 6449-6460.
[35] S. Kersten, Peroxisome proliferator activated receptors and lipoprotein metabolism, PPAR Res., 2008 (2008) 132960-132960.
[36] S. Park, Y.J. Lee, E.H. Ko, J.W. Kim, Regulation of retinoid X receptor gamma expression by fed state in mouse liver, Biochem. Biophys. Res. Comm., 458 (2015) 134-139.

[37] Y. Nakagawa, A. Satoh, S. Yabe, M. Furusawa, N. Tokushige, H. Tezuka, M. Mikami, W. Iwata, A. Shingyouchi, T. Matsuzaka, S. Kiwata, Y. Fujimoto, H. Shimizu, H. Danno, T. Yamamoto, K. Ishii, T. Karasawa, Y. Takeuchi, H. Iwasaki, M. Shimada, Y. Kawakami, O. Urayama, H. Sone, K. Takekoshi, K. Kobayashi, S. Yatoh, A. Takahashi, N. Yahagi, H. Suzuki, N. Yamada, H. Shimano, Hepatic CREB3L3 Controls Whole-Body Energy Homeostasis and Improves Obesity and Diabetes, Endocrinology, 155 (2014) 4706-4719.
[38] S. Floyd, C. Favre, F.M. Lasorsa, M. Leahy, G. Trigiante, P. Stroebel, A. Marx, G. Loughran, K. O'Callaghan, C.M.T. Marobbio, D.J. Slotboom, E.R.S. Kunji, F. Palmieri, R. O'Connor, The insulin-like growth factor-I-mTOR signaling pathway induces the mitochondrial pyrimidine nucleotide carrier to promote cell growth, Mol. Biol. Cell, 18 (2007) 3545-3555.
[39] C. Favre, A. Zhdanov, M. Leahy, D. Papkovsky, R. O'Connor, Mitochondrial pyrimidine nucleotide carrier (PNC1) regulates mitochondrial biogenesis and the invasive phenotype of cancer cells, Oncogene, 29 (2010) 3964-3976.
[40] M.A. Di Noia, S. Todisco, A. Cirigliano, T. Rinaldi, G. Agrimi, V. Iacobazzi, F. Palmieri, The human SLC25A33 and SLC25A36 genes of solute carrier family 25 encode two mitochondrial pyrimidine nucleotide transporters, J. Biol. Chem., 289 (2014) 33137-33148.
[41] L.R. Gray, M.R. Sultana, A.J. Rauckhorst, L. Oonthonpan, S.C. Tompkins, A. Sharma, X. Fu, R. Miao, A.D. Pewa, K.S. Brown, E.E. Lane, A. Dohlman, D. Zepeda-Orozco, J. Xie, J. Rutter, A.W. Norris, J.E. Cox, S.C. Burgess, M.J. Potthoff, E.B. Taylor, Hepatic mitochondrial pyruvate carrier 1 is required for efficient regulation of gluconeogenesis and whole-body glucose homeostasis, Cell Metab., 22 (2015) 669-681.
[42] A. Kharitonenkov, R. DiMarchi, Fibroblast growth factor 21 night watch: advances and uncertainties in the field, J. Intern. Med., 281 (2017) 233-246.
[43] M.K. Badman, A. Koester, J.S. Flier, A. Kharitonenkov, E. Maratos-Flier, Fibroblast growth factor 21-deficient mice demonstrate impaired adaptation to ketosis, Endocrinology, 150 (2009) 4931-4940.
[44] M. Watanabe, G. Singhal, F.M. Fisher, T.C. Beck, D.A. Morgan, F. Socciarelli, M.L. Mather, R. Risi, J. Bourke, K. Rahmouni, O.P. McGuinness, J.S. Flier, E. Maratos-Flier, Liver-derived FGF21 is essential for full adaptation to ketogenic diet but does not regulate glucose homeostasis, Endocrine, 67 (2020) 95-108.
[45] U. Martinez-Garza, D. Torres-Oteros, A. Yarritu-Gallego, P.F. Marrero, D. Haro, J. Relat, Fibroblast Growth Factor 21 and the Adaptive Response to Nutritional Challenges, Int. J. Mol. Sci., 20 (2019).
[46] H. Akbar, F. Batistel, J.K. Drackley, J.J. Loor, Alterations in hepatic FGF21, co-regulated genes, and upstream metabolic genes in response to nutrition, ketosis and inflammation in peripartal holstein cows, PLoS ONE, 10 (2015) e0139963-e0139963.
[47] K.M. Schoenberg, S.L. Giesy, K.J. Harvatine, M.R. Waldron, C. Cheng, A. Kharitonenkov, Y.R. Boisclair, Plasma FGF21 is elevated by the intense lipid mobilization of lactation, Endocrinology, 152 (2011) 4652-4661.
[48] Y. Shen, L. Chen, W. Yang, Z. Wang, Exploration of serum sensitive biomarkers of fatty liver in dairy cows, Sci. Rep., 8 (2018) 13574-13574.
[49] C. Xu, Q. Xu, Y. Chen, W. Yang, C. Xia, H. Yu, K. Zhu, T. Shen, Z. Zhang, FGF-21: promising biomarker for detecting ketosis in dairy cows, Veterinary Research Communications, 40 (2016) 49-54.
[50] L.S. Caixeta, S.L. Giesy, C.S. Krumm, J.W. Perfield, A. Butterfield, K.M. Schoenberg, D.C. Beitz, Y.R. Boisclair, Effect of circulating glucagon and free fatty acids on hepatic FGF21 production in dairy cows, Am. J. Physiol. Regul. Integr. Comp. Physiol., 313 (2017) R526-R534.
[51] J.H. Berger, M.J. Charron, D.L. Silver, Major Facilitator Superfamily Domain-Containing Protein 2a (MFSD2A) Has Roles in Body Growth, Motor Function, and Lipid Metabolism, PLoS ONE, 7 (2012) e50629-e50629.
[52] V. Sukonina, A. Lookene, T. Olivecrona, G. Olivecrona, Angiopoietin-like protein 4 converts lipoprotein lipase to inactive monomers and modulates lipase activity in adipose tissue, Proc. Natl. Acad. Sci. USA, 103 (2006) 17450-17455.
[53] A. Köster, Y.B. Chao, M. Mosior, A. Ford, P.A. Gonzalez-DeWhitt, J.E. Hale, D. Li, Y. Qiu, C.C. Fraser, D.D. Yang, J.G. Heuer, S.R. Jaskunas, P. Eacho, Transgenic angiopoietin-like (Angptl)4 overexpression and targeted disruption of Angptl4 and Angptl3: Regulation of triglyceride metabolism, Endocrinology, 146 (2005) 4943-4950.
[54] A. Xu, M.C. Lam, K.W. Chan, Y. Wang, J. Zhang, R.L.C. Boo, J.Y. Xu, B. Chen, W.S. Chow, A.W.K. Tso, K.S.L. Lam, Angiopoietin-like protein 4 decreases blood glucose and improves glucose tolerance but induces hyperlipidemia and hepatic steatosis in mice, Proc. Natl. Acad. Sci. USA, 102 (2005) 6086-6091.

[55] V. Gusarova, C. O'Dushlaine, T.M. Teslovich, P.N. Benotti, T. Mirshahi, O. Gottesman, C.V. Van Hout, M.F. Murray, A. Mahajan, J.B. Nielsen, L. Fritsche, A.B. Wulff, D.F. Gudbjartsson, M. Sjögren, C.A. Emdin, R.A. Scott, W.J. Lee, A. Small, L.C. Kwee, O.P. Dwivedi, R.B. Prasad, S. Bruse, A.E. Lopez, J. Penn, A. Marcketta, J.B. Leader, C.D. Still, H.L. Kirchner, U.L. Mirshahi, A.H. Wardeh, C.M. Hartle, L. Habegger, S.N. Fetterolf, T. Tusie-Luna, A.P. Morris, H. Holm, V. Steinthorsdottir, P. Sulem, U. Thorsteinsdottir, J.I. Rotter, L.M. Chuang, S. Damrauer, D. Birtwell, C.M. Brummett, A.V. Khera, P. Natarajan, M. Orho-Melander, J. Flannick, L.A. Lotta, C.J. Willer, O.L. Holmen, M.D. Ritchie, D.H. Ledbetter, A.J. Murphy, I.B. Borecki, J.G. Reid, J.D. Overton, O. Hansson, L. Groop, S.H. Shah, W.E. Kraus, D.J. Rader, Y.D.I. Chen, K. Hveem, N.J. Wareham, S. Kathiresan, O. Melander, K. Stefansson, B.G. Nordestgaard, A. Tybjærg-Hansen, G.R. Abecasis, D. Altshuler, J.C. Florez, M. Boehnke, M.I. McCarthy, G.D. Yancopoulos, D.J. Carey, A.R. Shuldiner, A. Baras, F.E. Dewey, J. Gromada, Genetic inactivation of ANGPTL4 improves glucose homeostasis and is associated with reduced risk of diabetes, Nat. Commun., 9 (2018) 2252-2252.

[56] J. Wang, X. Zhu, G. She, Y. Kong, Y. Guo, Z. Wang, G. Liu, B. Zhao, Serum hepatokines in dairy cows: Periparturient variation and changes in energy-related metabolic disorders, BMC Vet. Res., 14 (2018) 236-236.

[57] A. Van Helvoort, A.J. Smith, H. Sprong, I. Fritzsche, A.H. Schinkel, P. Borst, G. Van Meer, MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine, Cell, 87 (1996) 507-517.

[58] P.D.W. Eckford, F.J. Sharom, The reconstituted P-glycoprotein multidrug transporter is a flippase for glucosylceramide and other simple glycosphingolipids, Biochemical Journal, 389 (2005) 517-526.

[59] L. La Paglia, A. Listì, S. Caruso, V. Amodeo, F. Passiglia, V. Bazan, D. Fanale, Potential Role of ANGPTL4 in the Cross Talk between Metabolism and Cancer through PPAR Signaling Pathway, PPAR Res., 2017 (2017) 8187235-8187235.

[60] N. Pareek, J. Voigt, O. Bellmann, F. Schneider, H.M. Hammon, Energy and nitrogen metabolism and insulin response to glucose challenge in lactating German Holstein and Charolais heifers, Livestock Sci., 112 (2007) 115-122.

[61] G. Bobe, J.C. Velez, D.C. Beitz, S.S. Donkin, Glucagon increases hepatic mRNA concentrations of ureagenic and gluconeogenic enzymes in early-lactation dairy cows, J. Dairy Sci., 92 (2009) 5092-5099.

[62] K.D. Galsgaard, J. Pedersen, S.A.S. Kjeldsen, M. Winther-Sørensen, E. Stojanovska, H. Vilstrup, C. Ørskov, N.J. Wewer Albrechtsen, J.J. Holst, Glucagon receptor signaling is not required for N-carbamoyl glutamate- And L-citrulline-induced ureagenesis in mice, Am. J. Physiol. Gastrointest. Liver. Physiol., 318 (2020) G912-G927.

[63] L.M. Yanckello, L.E.A. Young, J.D. Hoffman, R.P. Mohney, M.A. Keaton, E. Abner, A.L. Lin, Caloric restriction alters postprandial responses of essential brain metabolites in young adult mice, Front. Nutr., 6 (2019) 90-90.

[64] G. Fiermonte, F. De Leonardis, S. Todisco, L. Palmieri, F.M. Lasorsa, F. Palmieri, Identification of the mitochondrial ATP-Mg/Pi transporter: Bacterial expression, reconstitution, functional characterization, and tissue distribution, J. Biol. Chem., 279 (2004) 30722-30730.

[65] V. Iacobazzi, F. Invernizzi, S. Baratta, R. Pons, W. Chung, B. Garavaglia, C. Dionisi-Vici, A. Ribes, R. Parini, M.D. Huertas, S. Roldan, G. Lauria, F. Palmieri, F. Taroni, Molecular and functional analysis of SLC25A20 mutations causing carnitine-acylcarnitine translocase deficiency, Hum. Mutat., 24 (2004) 312-320.

[66] D.K. Bricker, E.B. Taylor, J.C. Schell, T. Orsak, A. Boutron, Y.C. Chen, J.E. Cox, C.M. Cardon, J.G. Van Vranken, N. Dephoure, C. Redin, S. Boudina, S.P. Gygi, M. Brivet, C.S. Thummel, J. Rutter, A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans, Science, 336 (2012) 96-100.

10. Anhang

10.1. Danksagung

Diese Disserationsarbeit wäre ohne die fachliche und persönliche Unterstützung von Kollegen, Familie und Freunden nicht möglich gewesen. Mein besonderer Dank geht daher an folgende Personen:

Mein ausdrücklicher Dank gilt Frau Prof. Dr. Christa Kühn, der Leiterin des Instituts für Genombiologie am Leibniz-Insitut für Nutztierbiologie (FBN) für ihre fortwährende fachliche Unterstützung und die Expertise, von der ich in den vergangenen Jahren profitieren durfte. Ihr andauernder Ansporn, Ergebnisse und Methoden zu hinterfragen und differenziert zu betrachten sowie ihr Bestreben, mich in der Wissenschaftskommune zu vernetzen, wird mich auch zukünftig begleiten. Weiterhin bedanke ich mich bei Frau Dr. Rosemarie Weikard dafür, dass sie u.a. ihre Expertise in langer nichtkodierender RNA mit mir geteilt hat und fortwährend Unterstützung bei der Bewältigung vielfältiger Probleme gewährt hat.

Diese Arbeit wäre weiterhin ohne die exzellente Laborarbeit des Teams der Genombiologie des FBNs nicht vorstellbar gewesen. Allen voran danke ich Simone Wöhl für ihre helfenden Hände und die Aufbereitung der RNA-Bibliotheken für die Sequenzierung.

Im Rahmen dieser Dissertation habe ich viel über Datenaufbereitung und -auswertung lernen können und spreche Frieder Hadlich meinen besonderen Dank für seine Hilfe in allen bioinformatischen Fragestellungen aus.

Mein Dank gilt weiterhin dem Deutschen Akademischen Austauschdienst (DAAD), der mir über ein Stipendium einen viermonatigen Aufenthalt in Brisbane (Australien) in der Commonwealth Scientific and Industrial Research Organisation (CSIRO) ermöglicht hat. Die fachliche Begleitung, die Bereitstellung von Algorithmen und der kreative Forschergeist von Dr. Antonio Reverter hat die Bearbeitung und Auswertung der mir zur Verfügung stehenden Daten enorm bereichert.

Im Laufe der letzten vier Jahre habe ich viele wunderbare Wissenschaftler kennen lernen dürfen und bin für den Zusammenhalt und die persönliche Unterstützung der anderen Doktoranden der Genombiologie unschätzbar dankbar.

Großer Dank gilt meiner Familie und meinen Freunden, die mich über die Jahre gestützt haben und immer ein offenes Ohr hatten, wie fern ihnen das Thema und die Personen auch gewesen sein mögen. Allen anderen voran danke ich meiner Mutter: ohne dich hätte ich diesen Weg nicht gehen können.

10.2. Eidesstattliche Erklärung

Hiermit erkläre ich durch eigenhändige Unterschrift, die vorliegende Dissertation selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben. Die aus den Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Dissertation ist in dieser Form noch keiner anderen Prüfungsbehörde vorgelegt worden.

Wietje Nolte

10.3. Liste wissenschaftlicher Veröffentlichungen

Veröffentlichte wissenschaftliche Artikel mit Peer-Review [Teil der Dissertation (*)]

Nolte W, Thaller G, Kühn C (2019). Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. *PLOS ONE* 14(4): e0215913. doi: 10.1371/journal.pone.0215913.

Koch F, Thom U, Albrecht E, Weikard R, Nolte W, Kuhla B, Kühn C (2019). Heat stress directly impairs gut integrity and recruits distinct immune cell populations into the bovine intestine. *Proceedings of the National Academy of Sciences* 116(21) 10333-10338. doi: 10.1073/pnas.1820130116.

(*) Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2019). Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated with Metabolic Efficiency in Cattle. *Frontiers in Genetics* 10:1130. doi: 10.3389/fgene.2019.01130.

(*) Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2020). Identification and Annotation of Potential Function of Regulatory Antisense Long Non-Coding RNAs Related to Feed Efficiency in *Bos taurus* Bulls. *International Journal of Molecular Sciences* 21(9). doi: 10.3390/ijms21093292.

Fachvorträge und Konferenzbeiträge

Nolte W, Kühn C (2017). Nutzung von genomischen Informationen in der Zuchtplanung, *8. Pferde-Workshop* in Uelzen. Schriftlicher Beitrag in *DGfZ-Schriftenreihe* Heft 71, Bonn (ISSN 0949-8842): 110-118.

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2019). Aufdeckung und funktionelle Charakterisierung langer nichtkodierender RNA mit Schlüsselfunktion für Nährstoffansatz beim Rind, bei der *Jahrestagung der Deutschen Gesellschaft für Züchtungskunde (DGfZ)* in Gießen.

Wobbe M, Lehner S, Stock K F, Vosgerau S, Krattenmacher N, von Depka Prondzinski, M,Kalm E, Reents R, Nolte W, Kühn C, Tetens, J, Thaller G (2019). Analyse equiner SNP Arrays mittlerer Dichte im Hinblick auf ihre Einsatzmöglichkeiten beim Reitpferd, bei der *Jahrestagung der Deutschen Gesellschaft für Züchtungskunde (DGfZ)* in Gießen.

Thom U, Koch F, Albrecht E, Weikard R, **Nolte W**, Kuhla B, Kühn C (2019). Auswirkungen von Hitzestress auf die Darmintegrität beim Rind, bei der *Jahrestagung der Deutschen Gesellschaft für Züchtungskunde (DGfZ)* in Gießen.

Nolte W, Kalm E, Krattenmacher N, Lehner S, Reents R, Stock KF, Tetens J, Thaller G, von Depka Prondzinski M, Vosgerau S, Wobbe M, Kühn C (2020). Nutzung der Imputation für den Übergang von der Mikrosatelliten- basierten Abstammungsüberprüfung zur SNP-Genotypisierung, *9. Pferdeworkshop* in Uelzen. Schriftlicher Beitrag in *DGfZ-Schriftenreihe* Heft 80, Bonn.

Vosgerau S, Krattenmacher N, Falker-Gieske C, Blaj I, Wobbe M, Stock KF, von Depka Prondzinski M, Reents R, Kalm E, **Nolte W**, Kühn C, Tetens J, Thaller G (2020). Das Stockmaß als Referenzmerkmal für die genomische Selektion, *9. Pferdeworkshop* in Uelzen. Schriftlicher Beitrag in *DGfZ-Schriftenreihe* Heft 80, Bonn.

Wobbe M, Lehner S, Stock KF, Vosgerau S, Krattenmacher N, von Depka Prondzinski M, Kalm E, Reents R, **Nolte W**, Kühn K, Tetens J, Thaller G (2020). Spektrum und Potenzial genomischer Anwendungen beim Pferd, *9. Pferdeworkshop* in Uelzen. Schriftlicher Beitrag in *DGfZ-Schriftenreihe* Heft 80, Bonn.

Poster

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2018). The role of long non-coding RNA for the gene expression in metabolic processes in cattle, *7. International Symposium on Animal Functional Genomics (ISAFG)* in Adelaide (Australien).

Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2019). Long non-coding RNAs modulate metabolic efficiency in cattle and are linked to arginine biosynthesis (Poster P 411), *37. Konferenz der International Society for Animal Genetics (ISAG)* in Lleida (Spanien).

Weikard R, **Nolte W**, Hammon HM, Brunner RM, Albrecht E, Reverter A, Kühn C (2019). A transcriptional landscape of long noncoding RNAs in tissues from cattle differing in metabolic efficiency (Poster P 409), *37. Konferenz der International Society for Animal Genetics (ISAG)* in Lleida (Spanien).