UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

SCUOLA DI DOTTORATO DI RICERCA IN BIOSCIENZE E BIOTECNOLOGIE

INDIRIZZO GENETICA E BIOLOGIA MOLECOLARE DELLO SVILUPPO

CICLO XXVIII

Development of computational pipelines for transcriptome and miRNome characterization from RNA-seq data applied to swine adipose tissue

Direttore della Scuola: Ch.mo Prof. Paolo Bernardi

Coordinatore d'indirizzo: Ch.mo Prof. Rodolfo Costa

Supervisore: Ch.ma Prof.ssa Stefania Bortoluzzi

Dottorando: Enrico Gaffo

TABLE OF CONTENTS

RIASSUNTO

Le tecnologie per il sequenziamento massivo del DNA sono spesso usate per studiare il trascrittoma e ottenre profili d'espressione genica su scala genomica (RNA-seq). Rispetto ad altre tecnologie come i microarray, l'RNA-seq ha una maggiore sensibilità nel campionare e quantificare le molecole espresse e permette inoltre l'identificazione di trascritti sconosciuti o non caratterizzati. Il processamento di dati RNA-seq prevede molteplici passaggi di analisi (preprocessamento degli input per la valutazione della qualità e pulizia, allineamento delle read al genoma di riferimento, identificazione, quantificazione e annotazione dei trascritti, stima di espressione differenziale) che devono essere eseguiti in ordine sequenziale, mediante pipeline computazionali. Ogni singolo esperimento di RNA-seq può produrre grandi quantità di dati che richiedono l'impiego di metodi efficienti per ottenere la caratterizzazione qualitativa e quantitativa del trascrittoma. Esistono diversi metodi che implementano ogni passaggio concettuale di analisi e nuovi ne vengono continuamente proposti. Questo e' anche dovuto alla varietà dei quesiti biologici e disegni sperimentali a cui gli esperimenti di RNA-seq possono essere applicati. Di converso, non esiste un'implementazione comunemente adottata dello schema di processamento.

In questa tesi, abbiamo sviluppato una pipeline computazionale per l'analisi di dati RNA-seq focalizzata sul trascrittoma lineare; abbiamo esteso una pipeline esistente che analizza dati di RNA-seq di microRNA (miRNA) e piccoli RNA simili ai miRNA ed abbiamo iniziato a sviluppare una pipeline computazionale per l'identificazione e la quantificazione di RNA circolari. Gli obiettivi principali delle prime due pipeline sono il profiling dell'insieme dei trascritti (trascrittoma) e piccoli RNA (miRNoma) espressi, con l'identificazione di RNA noti e nuovi. Inoltre, è stato possibile studiare le variazioni di sequenza degli RNA (come gli isomiR dei miRNA), dei livelli di espressione di trascritti e piccoli RNA, e confrontare i profili di espressione tra diversi gruppi di campioni biologici.

Il maiale (Sus scrofa) è un organismo modello per numerose malattie o condizioni umane, ma anche molto importante di per sé per l'industria di carne e derivati di alto pregio economicamente importanti. Il tessuto adiposo e il lardo dorsale sono oggetto di attiva ricerca, poichè alcune caratteristiche qualitative e quantitative del grasso e i meccanismi e tassi di deposito e accumulazione del grasso sono in stretta connessione con aspetti tecnologici e risultati qualitativi dei prodotti finali, come il prosciutto crudo. Tuttavia, il quadro complessivo dei processi biologici e molecolari che regolano il deposito del lardo dorsale nei maiali è ancora incompleto.

In questa tesi, abbiamo applicato i metodi di analisi sviluppati a dati RNA-seq di RNA poliadenilati e piccoli RNA da campioni di tessuto adiposo sottocutaneo di 20 soggetti di razza Italian Large White (ILW). Gli animali selezionati sono stati allevati in condizioni molto standardizzate, ma presentano, riguardo i tratti del grasso, fenotipi e corrispondenti meriti genetici estremi e divergenti (maiali FAT e LEAN). L'analisi del profilo trascrizionale del lardo dorsale ha identificato l'espressione di 23.483 geni, dei quali solo il 54,1% rappresentato da geni noti. Dei 63.418 trascritti espressi, circa l'80% erano isoforme non precedentemente annotate. Confrontando i livelli di espressione dei maiali FAT contro i maiali LEAN, abbiamo poi identificato, con criteri molto stringenti, 86 trascritti differenzialmente espressi: 72 espressi a livelli più alti nei maiali obesi (tra cui ACP5, BCL2A1, CCR1, CD163, CD1A, EGR2, ENPP1, GPNMB, INHBB, LYZ, MSR1, OLR1, PIK3AP1, PLIN2, SPP1, SLC11A1, STC1) e 14 meno espressi (inclusi ADSSL1, CDO1, DNAJB1, HSPA1A, HSPA1B, HSPA2, HSPB8, IGFBP5, OLFML3). I geni sovraespressi sono implicati in processi del sistema immunitario, di risposta allo stimolo, attivazione cellulare e sviluppo dell'apparato scheletrico. I geni sottoespressi includono cinque proteine heat shock e sono associati a categorie funzionali quali il legame di proteine mal ripiegate, e la risposta allo stress. Nel tessuto adiposo un'eccessiva adiposità combinata a carenze nei meccanismi di risposta allo stress sono collegate ad uno stato infiammatorio del tessuto e, di conseguenza, ad alterazioni dell'attività secretoria del tessuto adiposo, similmente a quanto è stato osservato nell'obesità umana.

I miRNA sono importanti regolatori dell'espressione genica nel differenziamento, nell sviluppo e nella fisiologia cellulare dei diversi tessuti. Essi agiscono come regolatori post-trascrizionali dell'espressione genica, silenziando i trascritti bersaglio. Lo studio del miRNoma del lardo dorsale di maiale ha identificato l'espressione di centinaia di piccoli RNA, includendo potenziali nuovi miRNA, nuove isoforme di miRNA (isomiR) e nuovi microRNA-offeset RNA (moRNA), probabilmente prodotti dalle regioni terminali di precursori a forcina processate in modo non canonico. Da uno studio preliminare condotto su due campioni abbiamo rilevato 222 miRNA noti, 68 nuovi miRNA e 17 moRNA espressi da forcine note, e 312 nuovi miRNA espressi da 253 nuove forcine. L'espressione di cinque piccoli RNA, inclusi il moRNA ssc-moR-21-5p e un miRNA prodotto da un precursore da noi predetto, è stata validata mediante qRT-PCR, confermando l'affidabilità dei nostri risultati. In accodo con questi dati, un secondo studio condotto su 18 campioni ha identificato un miRNoma molto simile in termini di elementi espressi e varianti. Questo ha inoltre permesso di identificare miRNA e moRNA differenzialmente espressi tra soggetti FAT e LEAN, potenziali regolatori di trascritti la cui modulazione dell'espressione potrebbe essere implicata nelle variazioni fenotipiche dei soggetti considerati. Abbiamo predetto i potenziali bersagli dei miRNA e dei moRNA (nell ipotesi che i moRNA possano funzionare come miRNA) modulati prendendo in considerazione, per analisi ad hoc le sequenze dei trascritti ricostruite in precedenza e gli isomiR dei miRNA risultati maggiormente espressi e quindi rilevanti. Abbiamo integrato i risultati di queste predizioni con l'analisi combinata dei profili d'espressione di miRNA e trascritti, per selezionare le relazioni miRNA-trascritto maggiormente supportate dai dati d'espressione. La rete di interazioni miRNA-trascritti ottenuta in questo modo è stata arricchita dall'informazione su espressione differenziale, annotazione funzionale e predizioni del potenziale codificante e sovrapposizione dei trascritti con regioni genomiche di QTL di maiale. In questo modo siamo stati in grado di identificare un numero ristretto di interazioni potenzialmente molto significative che necessitano di essere investigate sperimentalmente. Ulteriori considerazioni stanno emergendo dallo studio del potenziale impatto di specifici miRNA differenzialmente espressi su geni appartenenti a pathway molto attinenti alla biologia del tessuto adiposo.

I risultati applicativi di questi studi hanno allargato la conoscenza dei trascritti e dei piccoli RNA espressi nel tessuto adiposo di maiale, e anche delle interazioni regolative tra piccoli RNA e trascritti, fornendo utili informazioni per una miglior comprensione del lardo dorsale di maiali ILW e nuove ipotesi per studi futuri sulla regolazione dell'espressione genica in questo tessuto. In aggiunta, stiamo attualmente sviluppando ed estendendo ulteriormente i metodi qui presentati, con applicazioni e obiettivi ulteriori rispetto a quelli descritti in questa tesi.

ABSTRACT

High throughput technologies for DNA sequencing are used more and more frequently for gene expression profiling studies (RNA-seq). With respect to other techniques such as microarrays, RNA-seq has higher sensitivity in retrieving the expressed molecules and presents the advantageous feature of allowing the detection of unknown or uncharacterized transcripts. RNA-seq data processing involves several computational steps (input preprocessing for quality evaluation and cleaning; read alignment to reference genome; transcript identification, quantification, and annotation; differential expression assessment) that have to be performed in sequential order, thus resulting in a computational pipeline. Each single RNA-seq experiment can produce large amounts of data that require the use of efficient computational methods to obtain transcriptome qualitative and quantitative characterization. There are different methods that implement each conceptual pipeline step, and new ones are continuously proposed. However, because of the variety of biological questions and study designs to which RNA-seq experiments can be applied to, there is not a commonly adopted implementation of the processing workflow.

In this thesis, we developed a computational pipeline for the analysis of RNA-seq data focused on the linear transcriptome, extended an existing pipeline that analyzes RNA-seq data of microRNAs (miRNAs) and miRNA-like small RNAs, and started to develop a computational pipeline for the detection and quantification of circular RNAs. The main objectives of the first two pipelines were the profiling of the set of the transcripts (transcriptome) and small RNAs (miRNome) expressed in the considered samples, by the identification of known and new RNAs. They allowed as well to investigate RNA sequence variations (such as miRNA isomiRs), transcripts and small RNAs expression levels, and to compare expression profiles between different sample groups.

The pig (*Sus scrofa*) is a model organism for human diseases, and very important per se for the meat industry. Fat and backfat tissues are subject of very active research since fat attributes and deposition traits are in strong connection with technological aspects and quality of pig products. However, the global framework of the biological and molecular processes regulating backfat deposition in pig is still incomplete. We applied our pipelines to RNA-seq data of polyadenylated and of small RNAs from pig subcutaneous adipose tissue samples from 20 Italian Large White (ILW) individuals. Selected animals were reared under very standard conditions but presented, for fat traits, extreme and divergent phenotypes (FAT and LEAN pigs) and genetic merits.

The backfat transcription profile was characterized by the expression of 23,483 genes, of which only 54.1% were represented by known genes. Of 63,418 expressed transcripts, about 80% were non-previously annotated isoforms. By comparing the expression level of FAT vs. LEAN pigs, we detected 86 robust differentially expressed transcripts, 72 more expressed in fat pigs (including *ACP5, BCL2A1, CCR1, CD163, CD1A, EGR2, ENPP1, GPNMB, INHBB, LYZ, MSR1, OLR1, PIK3AP1, PLIN2, SPP1, SLC11A1, STC1*) and 14 less expressed (including *ADSSL1, CDO1, DNAJB1, HSPA1A, HSPA1B, HSPA2, HSPB8, IGFBP5, OLFML3*). Overexpressed genes were implied particularly in immune system processes, response to stimulus, cell activation and skeletal system development. Underexpressed genes included five heat shock proteins and were involved in unfolded protein binding and stress response functional categories. Adipose tissue alterations and impaired stress response are linked to inflammation and, in turn, to adipose tissue secretory activity, similar to what is observed in human obesity.

MiRNAs play important roles in cell differentiation and physiology acting as post-transcriptional regulators of gene expression by silencing targeted transcripts. The pig backfat miRNome showed the expression of hundreds of small RNAs, including putative new miRNAs, new miRNA isoforms (isomiRs), and new moRNAs, likely produced from the terminal regions of non-canonically processed hairpin precursors. From a first study on two samples, we detected 222 known miRNAs, 68 new miRNAs and 17 moRNAs expressed from known hairpins, and 312 new miRNAs expressed from 253 new hairpins. The expression of five small RNAs, including moRNA *ssc-moR-21-5p* and a miRNA from a new hairpin, was validated by a qRT-PCR assay, thus confirming the robustness of our results. A second study on 18 samples identified a largely overlapping miRNome in terms of expressed elements and variations, and was important to identify differentially expressed miRNAs and moRNAs in FAT and LEAN subjects. We predicted putative regulatory interactions between small RNAs and transcripts by sequence analysis, using custom target predictions on reconstructed transcript sequences and miRNA isomiRs. We then integrated target prediction results with combined analysis of miRNA and transcript expression data, to eventually select miRNA-transcript relations most supported by negative correlation of expression profiles. Further, the predicted network of miRNA-transcript interactions was enriched by information on transcript differential expression, functional annotations and coding potential predictions, and transcript overlap with pig QTL genomic regions. In this way we were able to focus on a restricted and possibly most significant number of interactions that need to be experimentally investigated. Additional considerations are coming from the study of the possible impact of specific differentially expressed miRNAs to genes belonging to the pathways most germane to adipose tissue features.

The applicative results of these studies enlarged the knowledge of transcripts and small RNAs expressed in the pig adipose tissue, as well as small RNA-transcripts regulatory interactions, providing information helpful for a better understanding of ILW pig backfat and future studies on gene expression regulation in this tissue. Moreover, the methods presented here are currently undergoing further development and extension, and have applications well over and above those presented in this thesis.

1 BACKGROUND

1.1 GENE EXPRESSION

In living cells, genetic information flows from the genes contained in deoxyribonucleic acid (DNA) linear sequences, to ribonucleic acid (RNA) transcripts, to amino acid chains (proteins), consisting of two basic steps, transcription and translation. This description, referred to as the "central dogma of molecular biology", oversimplifies the numerous and complex biological mechanisms coming into play in the process of gene expression. The variety of cellular organisms is determined by the differences in their DNA sequences and thus their genetic makeup. Besides, in multicellular organism different cell types share the same DNA sequence and their diversity is defined, in a first instance, by the set of activated genes and the genome-wide expressed RNAs, namely their transcriptome, which in turn determines the protein products. Moreover, cells with the same set of activated genes can finely modulate their expression patterns by regulating gene expression at different levels, including transcriptional control, RNA processing, RNA transport and localization, translational control, RNA degradation, and protein activity organization.

The number of gene products found in a cell is much larger than the number of expressed genes, both in terms of transcripts and protein products. Various mechanisms determine transcriptome complexity. A single gene can generate several transcripts; for instance by the use of alternative promoters, or by post-transcriptionally modifications to the transcribed RNA, such as alternative polyadenylation sites and alternative splicing. In eukaryotes, transcription occurs in the nucleus from the activity of RNA polymerases, forming single stranded RNAs (primary transcripts). Primary transcripts bearing information for proteins (pre-mRNA) are organized in sequence modules, exons and introns, which are defined in the frame of the RNA splicing process. Splicing is a two-step biochemical process co-transcriptionally regulated accomplished by two complex macromolecular machineries (spliceosomes) that process pre-mRNA by removing the introns and ligating the exons to form the mRNA transcript. Exons can be chained in the same order of their transcription, or can present multiple and developmentally regulated alternative patterns of splicing producing multiple mRNA variants (transcript isoforms), in which the exon chain excludes some exons and/or is rearranged with a different exon order (Figure 1). Alternative splicing occurs for many genes, as much as in > 90% of human multi-exon protein-coding genes (Pan et al., 2008; Wang et al., 2008a).

In addition to the role of intermediate products toward the production of proteins, the importance of cells' transcripts is underlined by the fact that the largest part of eukaryotic

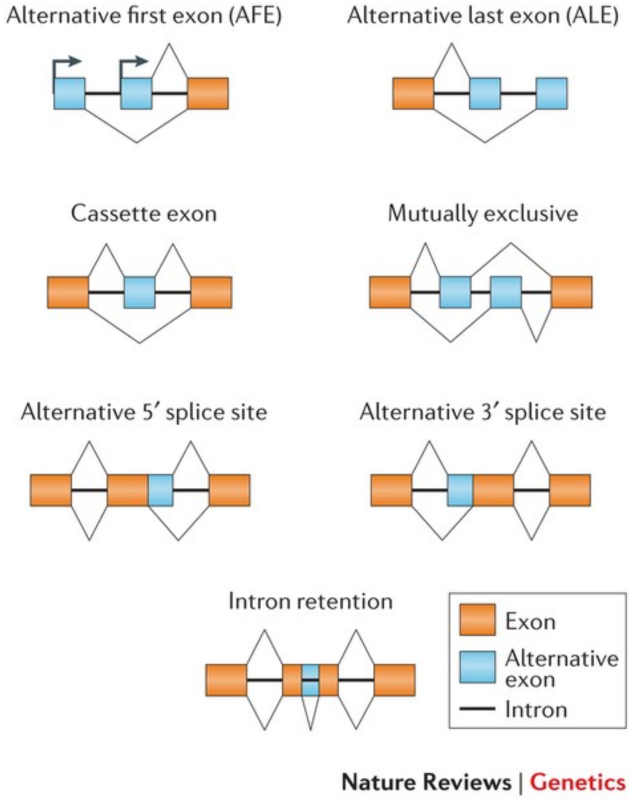

Alternative first exon (AFE)
Alternative last exon (ALE)
Cassette exon
Mutually exclusive
Alternative 5′ splice site
Alternative 3′ splice site
Intron retention
Exon
Alternative exon
Intron
Nature Reviews | Genetics

Figure 1. Alternative splicing patterns. Alternative splicing patterns including, from top left to right, the inclusion of alternative first and last exons (AFE and ALE, respectively), cassette exon, mutually exclusive cassettes, alternative 5' and 3' splice sites, and retained intron. Cassette exon skipping is the most common alternative splicing event in humans. From (Scotti and Swanson, 2016).

DNA does not encode proteins (Fox, 2014): almost 98% of the human genome is non-coding, the majority of which is anyway transcribed into RNA as functional products (Djebali et al., 2012). The definition of "gene", rather than being "a DNA locus encoding a protein", has become more and more as the concept of a transcriptional unit generating a set of sequences that after transcription produce one or more functional transcripts and might encode one or more protein isoforms (Djebali et al., 2012; Sharp, 2009; Wang et al., 2008a). Several types of non-coding RNAs (ncRNAs) have been identified, whose function is not always known. Noncoding RNAs (ncRNAs) are usually categorized in three main groups according to their size: long non coding RNAs (lncRNAs; > 200 nt), medium size ncRNAs (30 to 200 nt), and small RNAs (< 30 nt). ncRNAs play a number of different roles. Housekeeping ncRNAs, such as ribosomal RNA (rRNA) and transfer RNA (tRNA), are abundant in cells and are directly involved in protein synthesis. Others (small nuclear RNAs, snRNAs) are involved in pre-mRNA processing, or in guiding chemical modifications of RNAs, in the biosynthesis of rRNA and tRNAs (small nucleolar RNAs, snoRNAs; small cajal body-specific RNAs, scaRNAs). Small ncRNAs, like microRNAs (miRNA), small interfering RNA (siRNA) and Piwi-interacting RNAs (piRNAs), play critical roles in gene expression regulation at epigenetic and/or post-transcriptional levels. The family of long ncRNAs (lncRNAs) is highly diversified; recent studies accumulated a large body of evidence regarding lncRNAs abundance and functions. By most estimates, the number of human lncRNAs outstrips the number of protein-coding genes (Djebali et al., 2012). Function of lncRNAs is known only for few cases, such as X inactive specific transcript (XIST; in X chromosome inactivation), HOX transcript antisense RNA (HOTAIR; in positional identity) and telomerase RNA component (TERC; in telomere elongation). Nevertheless, the list of characterized lncRNAs is growing. LncRNAs may act as scaffolds, decoys or signals and can act through genomic targeting, regulation in cis or trans, and antisense interference. LncRNAs can be categorized according to their role: non-functional lncRNAs that are probably transcriptional noise; lncRNAs that function indirectly through their transcription; and functional lncRNAs acting in cis and/or in trans (Quinn and Chang, 2016). The majority of ncRNAs present a linear structure, but recent evidence reported abundance of circular RNA (circRNAs) forms (see Box "circular RNAs"), whose regulatory functions are still under investigation.

1.1.1 MICRORNAS AND MIRNA-OFFSET RNAS

MicroRNAs (miRNAs) are small endogenous non-coding RNAs of about 22 nucleotides discovered about 20 years ago (Lee et al., 1993), which act as post-transcriptional regulators of gene expression. MiRNAs are highly conserved and present in nearly all eukaryotes, supporting the idea that they play critical roles for the cell physiology. Mature miRNAs (miRs) expression patterns can differ by tissue type and conditions and, since they are involved in cell development, cell differentiation, and regulation of cell cycle, many research investigated their role in disease and cancer (Kong et al., 2012), as well as their use as biomarkers and diagnostics (Wang et al., 2016). In the canonical pathway of miRNA genesis (Figure 2) miRNA genes are transcribed by RNA polymerase II or RNA polymerase III into primary-miRNA transcripts (pri-miRNA). Almost half of miRNA genes are organized in polycistronic clusters and are therefore coexpressed (Kim et al., 2009). In the nucleus, pri-miRNAs fold in a hairpin-like structure with a double-stranded stem of 33 base-pairs, a terminal loop, and two single stranded regions flanking the hairpin. The pri-miRNAs undergo Drosha-mediated cleavage of the single stranded flanking sequences at the base of the hairpin stem, to form the precursor-miRNAs (pre-miRNAs). Non-canonical pathways of miRNA genesis have been identified in recent studies (Winter et al., 2009), including miRNAs derived from introns released by spliced transcripts (mirtrons) (Okamura et al., 2007; Ruby et al., 2007) (Figure 2), and from other transcripts such as snoRNAs (Ender et al., 2008), tRNAs (Maute et al., 2013) and lncRNAs (Keniry et al., 2012). Then, pre-miRNAs are exported in the cytoplasm by Exportin-5 in complex with Ran-GTP, which also protect pre-miRNAs from degradation in the nucleus. In the cytoplasm, pre-miRNAs is processed by the RNase III endonuclease Dicer, which cleaves off the loop of the pre-miRNA and generates a roughly 22-nucleotide miRNA duplex with two nucleotides protruding as overhangs at each 3p end. One strand of the miRNA duplex, the guide strand, is subsequently incorporated into the RNA-induced silencing complex (RISC), a multiprotein assembly including Argonaute-2 (Ago2) proteins, which mediates target gene expression. The other strand (passenger strand) is degraded. Although both the strands could give rise to mature miRNAs, usually only the strand with less thermodynamically stable base pair at its 5p is loaded into RISC. Thus, the activated RISC is guided to the target mRNA by sequence complementarity of the incorporated miRNA. Targeting of mRNA takes place primarily by base pairing at the 5p end of the miRNA within a region of as few as 6 nucleotides called *seed*. Binding to mRNAs' 3'-UTR can occur either with perfect complementarity, causing the mRNA degradation, or with imperfect complementarity, causing reversible inhibition of the mRNA translation(Saxena et al., 2003), but recent evidence suggest that miRNAs can target also coding regions (Hausser et al., 2013) and mRNAs' 5'-UTR (Ørom et al., 2008). Recent studies revealed that miRNA

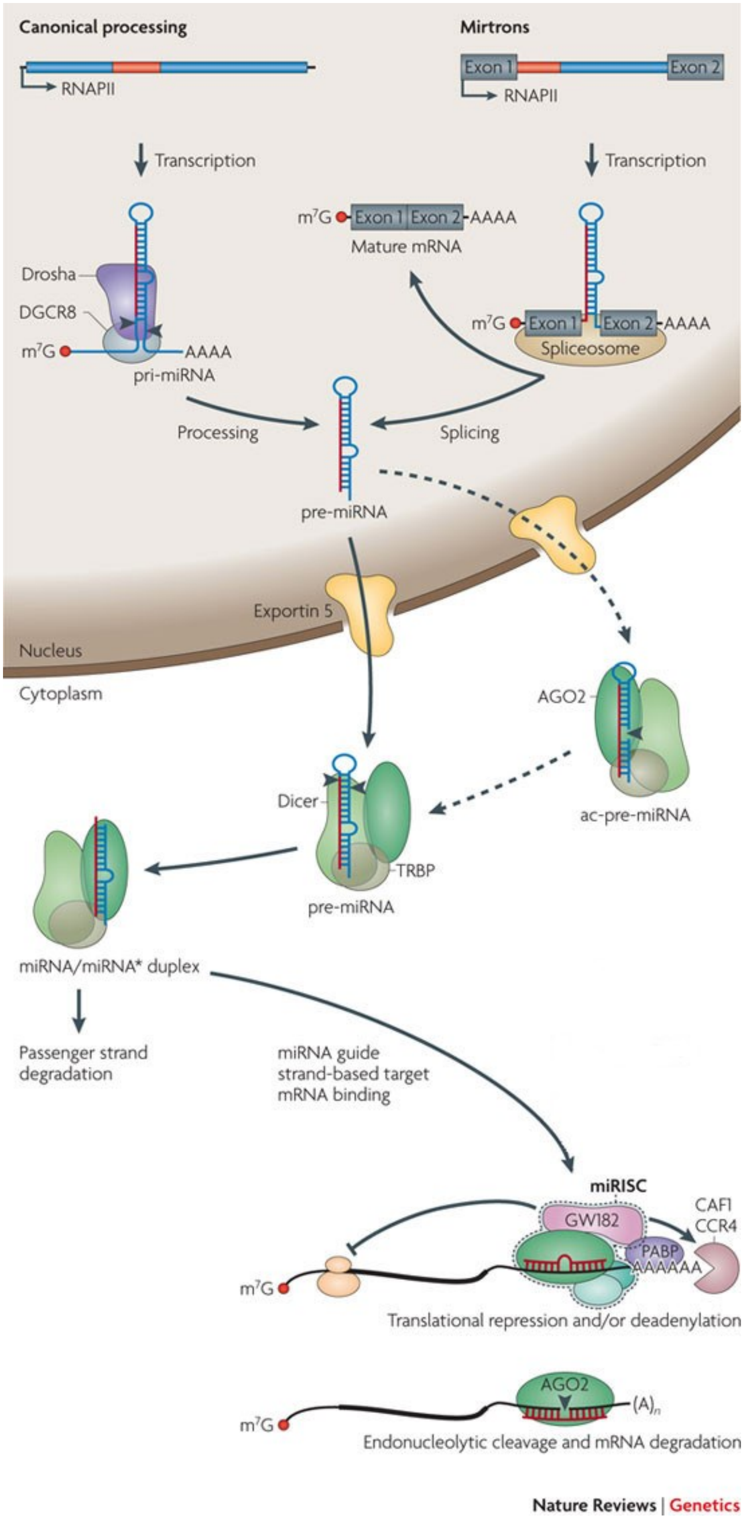


Figure 2. Canonical miRNA biogenesis. Primary precursor (pri-miRNA) processing occurs in two steps, catalyzed by two RNase III enzymes, Drosha and Dicer, operating in complexes with dsRNA-binding proteins (dsRBPs), for example DGCR8 and transactivation-responsive (TAR) RNA-binding protein (TRBP) in mammals. In the first nuclear step, the Drosha–DGCR8 complex processes pri-miRNA into an ~70-nucleotide precursor hairpin (pre-miRNA), which is exported to the cytoplasm. Some pre-miRNAs are produced from very short introns (mirtrons) as a result of splicing and debranching, thereby bypassing the Drosha–DGCR8 step. Cleavage by Dicer, assisted by TRBP, in the cytoplasm yields an ~20-bp duplex. In mammals, Argonaute 2 (AGO2) can support Dicer processing. Following processing, the guide strand of the miRNA/miRNA* duplex is incorporated into a miRNA-induced silencing complex (miRISC), whereas the other strand (passenger or miRNA*) is released and degraded. Adapted from (Krol et al., 2010).

may also function as direct positive or negative regulators of gene transcription by targeting gene promoters in the nucleus (Salmanidis et al., 2014). Individual miRNAs only moderately repress their targets. Besides, under normal physiological conditions, multiple miRNAs seem to act synergistically (Gennarino et al., 2012; Tsang et al., 2010) and through feedback and feed-forward loops that can amplify or reduce their effects (Tsang et al., 2007). Given that, expression of families and/or clusters of miRNAs might have increased effects on a single pathway containing multiple targets of the miRNAs.

Deep sequencing studies has revealed isomiRs (Morin et al., 2008), which are variant miRNA sequences that all appear to derive from the same gene but vary in sequence due to post-transcriptional processing. Because isomiRs contain different sequences, they may have different targets and thus different cell functions. Despite this source of sequence variation may impact cell biology, isomiRs are so far under-studied (Desvignes et al., 2015).

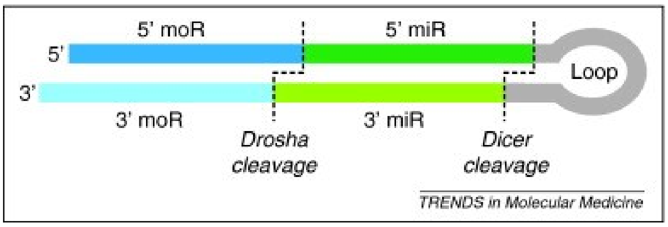


Figure 3. MoRNAs in miRNA hairpin. Each single microRNA precursor hairpin can be processed by Drosha and Dicer to generate up to four small RNAs, namely two miRNA and two miRNA-offset RNAs. The latter derive from the terminal parts of the hairpin. Figure from (Bortoluzzi et al., 2011).

High sequencing depth and careful data mining also revealed miRNA-offset RNAs (moRNAs). First reported in *Ciona intestinalis* (Shi et al., 2009), moRNAs are ~20 nt miRNA-like non-coding RNAs that are believed to be generated by the Drosha processing of miRNA precursors. In pre-miRNAs, moRNAs are located adjacent or overlapping to the 5' and 3' miRNA sequences (Figure 3). MoRNAs were recently reported in few RNA-seq studies carried out in different human cell conditions (Asikainen et al., 2015; Bortoluzzi et al., 2012; Langenberger et al., 2009), including solid tumors (Meiri et al., 2010), and other organisms (Babiarz et al., 2008; Gaffo et al., 2014; Shi et al., 2009; Zhou et al., 2012). They are hypothesized to act as regulatory elements like miRNAs, guiding RISC to complementary target mRNAs (Asikainen et al., 2015), but their function remains unknown. Although moRNAs' abundance is lower than most miRNAs, they are developmentally expressed and their expression seems not to correlate with the corresponding miRNA expression, which in some cases can also be lower (Umbach et al., 2010). MoRNAs prevalently arise from the 5' arm (Gaffo et al., 2014; Shi et al., 2009), even thou Asikainen et al. (2015) (Asikainen et al., 2015) reported a predominance of 3' moRNAs in human embryonic stem cells; but many precursors generate moRNAs also from the minor miRNA arm (Gaffo et al., 2014). This evidence suggests that moRNAs may represent a distinct class of functional miRNA co-product, instead of miRNA by-products.

Box: circular RNAs

Circular RNAs (circRNAs) are a class of non-coding RNAs that present a circular, non-linear structure. CircRNAs are highly stable RNA with important regulatory roles that are abundantly and cell differentiation-dependently expressed in both normal physiology and disease. CircRNAs form covalently closed continuous loop (Li et al., 2015a) generated from immature RNA joined in a non-co-linear way by a process called back-splicing (Figure 4). RNA binding proteins (RBPs) such as Muscleblind (Ashwal-Fluss et al., 2014) and Quaking (Conn et al., 2015) were shown to bridge two flanking introns to induce backsplicing, resulting in circRNA formation. CircRNAs were identified decades ago (Capel et al., 1993), but only recently RNA-seq projects and bioinformatics analysis reported circRNAs to be present in animals (Li et al., 2015a) with developmental stage- and tissue-specific expression (Salzman et al., 2013). Also evolutionary preservation of circRNAs supports important functions. Both paralogous and orthologous gene pairs were reported (Jeck et al., 2013) to express circular transcripts beyond apparent sequence conservation. Several lines of evidences indicate that the majority of circRNAs have limited coding potential (You et al., 2015). Besides, circRNAs can be competitors during pre-mRNA splicing (Ashwal-Fluss et al., 2014). In addition, circRNAs with multiple microRNA (miRNA) binding sites might function as miRNA sponges thus regulating specific pathways (Hansen et al., 2013; Li et al., 2015b), also in cancer (Tay et al., 2015). Furthermore, circRNAs can be competitors during pre-mRNA splicing (Ashwal-Fluss et al., 2014). CircRNAs show independent expression with respect to linear transcripts from the same gene, implying regulated expression(Chen and Yang, 2015). They are characterized by high stability and appear to accumulate in particular in cells with a low proliferation rate. Groundwork of circRNA biology still needs to be done as fundamental research. Several features of circRNAs as richness of functions, regulatory potential, pervasiveness, stability and detectability in body fluids clearly make circRNAs extremely interesting for fundamental research and push scientists to investigate their physiological functions, their impact in disease and their usefulness as biomarkers.

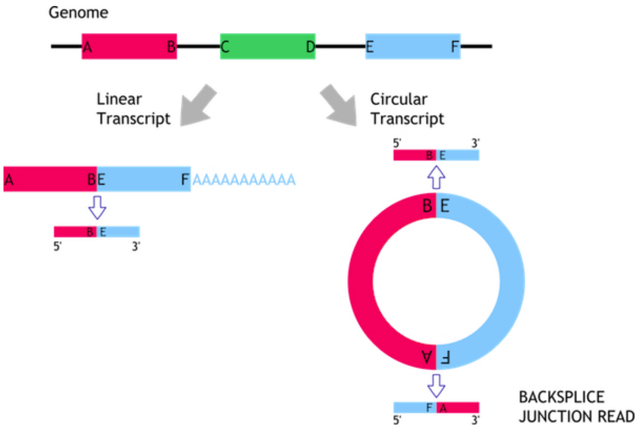


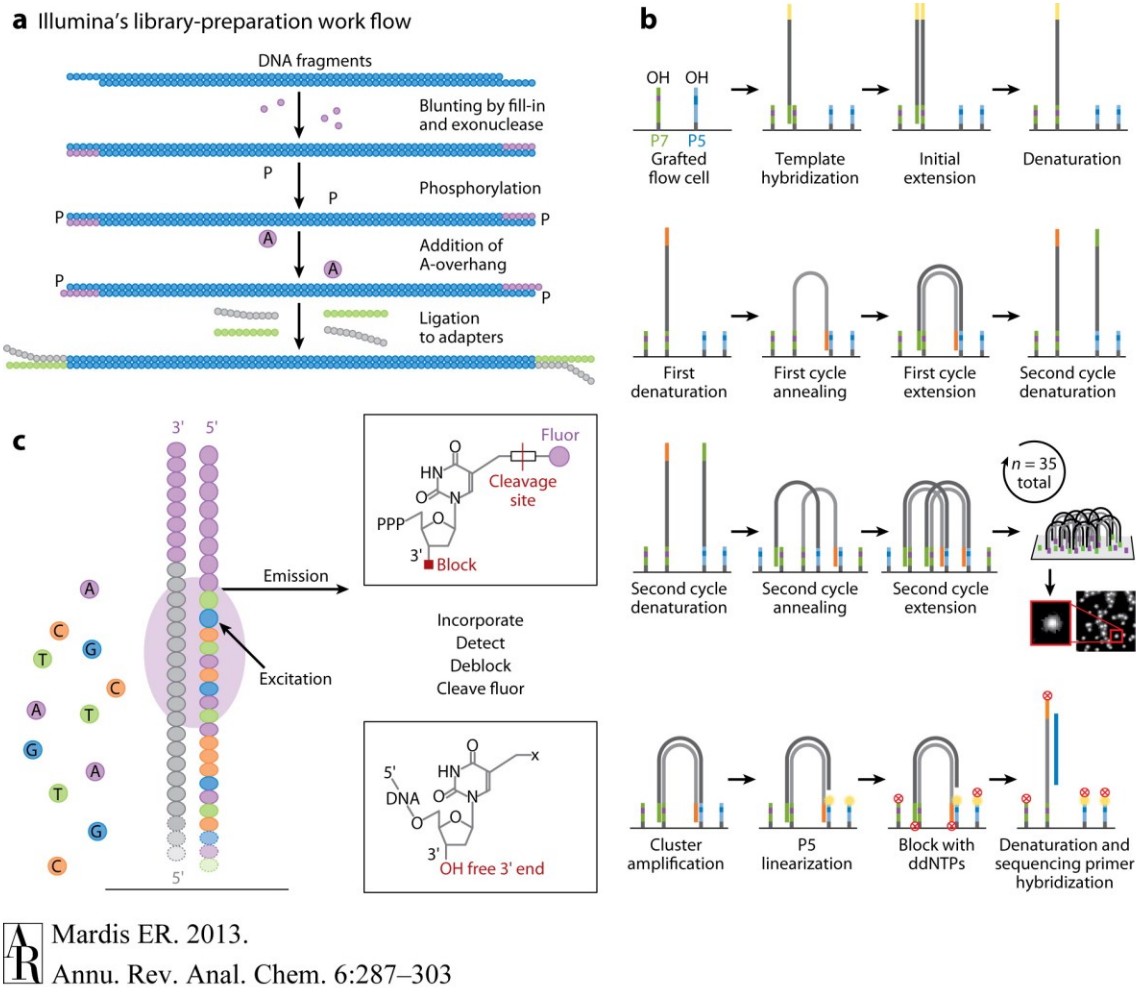
Figure 4. CircRNA structure. Circular RNAs are generated by backsplice events in which exons, instead of being processed in the canonical linear way (AB-EF), undergoes a circularization by the joining of the 5' and 3' ends (B-EF-A).

1.2 HIGH THROUGHPUT TRANSCRIPTOME SEQUENCING: RNA-SEQ

DNA sequencing technology to discover the order of nucleic acids in polynucleotide chains has dramatically improved since the early implementation of the Sanger sequencing method (Sanger et al., 1977). Despite this first-generation Sanger sequencing was improved by automating the process (Hunkapiller et al., 1991), and with which the first assembly of the human genome was accomplished (Consortium, 2004), the real revolution took place with the development of the next-generation sequencing (NGS) technologies. NGS, or second-generation sequencing, methods are based on a shotgun massively parallel high-throughput approach that results in millions, or even billions, of short (from 35 nt to 500 nt, depending on the platform) DNA sequences (reads). NGS can be applied to a wide variety of experiment types (Buermans and den Dunnen, 2014), including transcriptomic, in which case it is referred to as RNA-sequencing (RNA-seq) (Wang et al., 2009).

NGS includes several phases as template (library) preparation, sequencing, imaging or signal processing, and data analysis. The unique combination of specific protocols and underlying biochemistry distinguish one technology from another, which have been realized in commercial products by companies like Illumina, Roche, and Life Technologies.

In particular, the Solexa/Illumina approach is currently the most used (Greenleaf and Sidow, 2014). Illumina technology is based on the sequencing by synthesis (SBS) of complementary DNA (cDNA) fragments attached to a glass slide. Regarding RNA-seq experiments, RNA sequences are reverse transcribed into cDNA fragments in the sample preparation step. Alternative protocols can be used for different type of experiments, such as sequencing of small RNAs, sequencing of multiple samples, and other ones (see Box "Sample and library preparation for RNA-seq studies").The library preparation is accomplished by random fragmentation of cDNA, followed by *in vitro* ligation of specific nucleotide sequences (adaptors) to the ends of each library fragment (Figure 5a). Library fragments are then hybridized to the flow cell, a planar optically transparent surface similar to a microscope slide, which contains a lawn of oligonucleotide anchors tethered to its surface (Figure 5b). Here, the fragments undergo solid-phase PCR amplification to generate clusters of ~1,000 cloned templates (amplicons) for about 200 million distinguishable spots in each of the eight flow cell channels (lanes). This step is named "bridge amplification", because the DNA strands have to arch over to prime the next round of polymerization off neighboring surface-bound oligonucleotides (Figure 5b). This amplification step is required to provide enough signals to be detected by the image sensor during the sequencing phase. Sequencing itself is achieved by synthesis using fluorescent "reversible-terminators" dNTPs (Figure 5c) having four different colors for the



Mardis ER. 2013.
Annu. Rev. Anal. Chem. 6:287–303

Figure 5. Illumina sequencing procedure. (a) Illumina library-construction process. (b) Illumina cluster generation by bridge amplification. (c) Sequencing by synthesis with reversible dye terminators. Figure from (Mardis, 2013).

different bases. The amplicons are single-stranded and after a primer is hybridized to the adapter, the template is extended by a modified DNA polymerase and a mixture of the reversible terminator dNTPs. In each cycle of sequencing a single base is incorporated thanks to a cleavable moiety at the 3' hydroxyl position. The presence of the blocking group allows a synchronized process. The remaining bases are washed away and two lasers interrogate the fluorescent labels of the attached base to get an image in which each cluster will have a different color representing the inserted nucleotide (Figure 5c). These raw image files represent terabytes of data and require substantial storage resources. The images are then processed in order to extract numerical signals for every base at every synthesis event from all the parallel reactions. These signals are used for base calling. Then, the terminating group and the florescent dye are cleaved, and after an additional washing, the machine is ready for the next sequencing cycle. The number of cycles, corresponding to the read lengths, is limited by multiple factors that cause signal decay and dephasing. After image and signal processing, data consist of a list of short sequences together with their base call qualities. The output to the user can be encoded with the

FASTQ format, given by the Illumina CASAVA or BaseSpace (basespace.illumina.com) software. It is a plain text file containing four rows per read. The first row, beginning with the '@' character, is an header uniquely identifying the read (usually the cluster position in the flow-cell is used as ID number) and an optional description; the second row reports the read sequence with 'A', 'G', 'T', 'C' characters representing the bases, or 'N's when the base calling failed; the third row begins with the '+' character, optionally followed by the sequence identifier and description; the fourth row reports the sequencing quality for each base, encoded in a way such that each single character corresponds to the quality of the base in that position. An example of a FASTQ read is given below:

```
@HWI-ST1296:58:D1T0GACXX:1:1101:1243:2227 1:N:0:AGTCAA
CGGCAGTGTCGTAAAATATTCAGTATCACATGAAACCTCTTGTCAACTTTCAAAGCN
+
BCCFFFDFHHHFHJJJJIIJJJJHIJJJJJJIJJJJJJJJJIIJJJIJJJJIJJII
```

A quality value Q is an integer mapping of the probability p that the corresponding base call is incorrect. The equation used for the standard Sanger encoding, known as Phred quality score is:

$$Q = -10 \log_{10} p$$

For example, if Phred assigns a quality score of 30 to a base, the chances that this base is called incorrectly are 1 in 1000 (99.9% accuracy). Sanger format can encode a Phred quality score from 0 to 93 using ASCII 33 ('!') to 126 ('~'). For this reason the encoding is sometimes called *Phred+33*, to distinguish from other offsets, like the old Illumina/Solexa formats like CASAVA version 1.3 to 1.8 that set ASCII offset to 64 instead of 33, expecting Phred scores not greater than 40.

In addition, the Illumina sequencing technology can produce paired-end data, namely each DNA cluster can be sequenced at both ends. After the first round of sequencing, the single stranded flow-cell bound DNA undergo again bridge amplification, but this time the forward strand is washed away, leaving clusters of the reverse strand, which can be sequenced as before. Paired-end reads improve the accuracy of alignments because it reduces aspecific mapping of reads and is informative for transcript reconstruction.

The number of studies using RNA-seq technology has increased significantly over the past few years as evidenced by the number of RNA-seq data set stored in Short Read Archive (SRA) (Figure 6). RNA-seq presents key advantages over traditional methods like Sanger sequencing and microarrays. In particular, the sequencing cost per base is dramatically lower compared to the Sanger sequencing. Having very low background signal, the range of expression levels is greater than in microarrays, potentially spanning six orders of magnitude depending on the sequencing depth. The detection is not limited to known

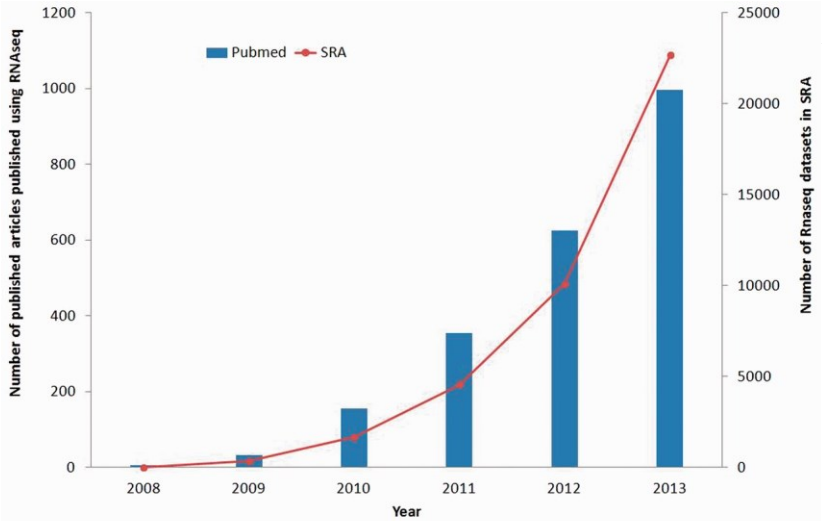


Figure 6. Distribution of the number of RNA-seq archived data sets and publications. SRA, Sequenced Read Archive. Figure from (Han et al., 2015).

sequences and novel splicing junction, isoforms, or unknown variations can be identified with single base precision, which is a great improvement for studies on non-model organisms. However, RNA-seq suffers of lower raw accuracy compared to the traditional Sanger sequencing. In the Illumina approach, the average raw error is in the order of 1-1.5% and the dominant error is substitution, occurring more frequently when the previous incorporated nucleotide is a 'G' base. Moreover, the reads are short and far from the ~1Kb reads of Sanger sequencing and NGS techniques need additional amplification steps that may bias the results.

Box: Sample and library preparation for RNA-seq studies

In RNA-seq studies the first step, sample preparation, is fundamental and should be planned according to the nature of the experiment. The RNA extraction method and the library selection schemes can influence the data (Raz et al., 2011; Sultan et al., 2014). In NGS protocols, the total RNA extracted from cells undergoes a further selection phase, since high proportion of cellular RNA arises from ribosomal and mitochondrial sources. Transcriptome analysis studies focusing on mature coding transcripts assume that the most known mature mRNAs are polyadenylated. Thus, total RNA is processed with oligo-d(T) tagged beads to isolate the poly(A)+ fraction, with optional modifications if transcript strand orientation information has to be maintained (Sultan et al., 2012). Poly(A) enrichment can capture also non-coding RNAs that can be polyadenylated, including microRNAs, snoRNAs, lncRNAs and pseudogenes. However, the recent introduction of ribosomal RNA (rRNA) depletion protocols, which can remove the 80% rRNA constituting the total RNA pool, extended the view of the transcriptome to the poly(A)- fraction of the RNA, facilitating the simultaneous characterization of polyadenylated and non-polyadenylated (e.g. rRNA and other transcripts generated by RNA polymerase I and III, many lncRNAs, and also circRNAs) RNAs. Then, RNA is size selected, usually discriminating between long (>200 nt) and small (<200 nt) RNAs. Large RNA molecules must be fragmented into smaller pieces (200-500 bp) prior to library preparation. Small

RNAs instead do not need fragmentation and adapters specific for small RNA libraries can be ligated directly. Another variation of library preparation protocol derives from the “multiplexing” technique. A single flow cell is partitioned into eight channels, or “lanes” that are independent from each other. Through the use of different indexes in the adapters (multiplexing), a single flow cell could run more than one different sample per lane. The sample source is then distinguished post-sequencing by checking the index nucleotide sequence in each read. This strategy is useful to lower the sequencing cost per experiment when high coverage is not required.

1.3 GENOME ANNOTATION RESOURCES

Nowadays, genome nucleotide sequences and genomic annotation are freely available from online web services, which allow exploring and downloading of different organisms’ data. There are some major resources that are continuously updated and curated, supported by research institutes and organizations. Ensembl (http://www.ensembl.org), NCBI (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov), and the UCSC Genome Browser (University of California, Santa Cruz; http://genome.ucsc.edu) are some central resources for genomic data.

Table 1. Ensembl genome assembly statistics

Species	***Homo sapiens* (Human)**	***Sus scrofa* (Pig)**
Assembly	GRCh38.p5 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.20, Dec 2013	Sscrofa10.2, INSDC Assembly GCA_000003025.4, Aug 2011
Database version	83.38	83.102
Base Pairs	3,547,121,844	3,024,658,544
Golden Path Length	3,096,649,726	2,808,525,991
Coding genes	20,313 (incl 512 readthrough)	21,630 (incl 10 readthrough)
Non coding genes	25,180	3,124
Small non coding genes	7,703	2,804
Long non coding genes	14,896 (incl 197 readthrough)	135 (incl 1 readthrough)
Misc non coding genes	2,307	185
Pseudogenes	14,453 (incl 4 readthrough)	568
Gene transcripts	199,184	30,585
Genscan gene predictions	50,766	52,372
Short Variants	149,490,457	60,359,717
Structural variants	4,149,389	85

In particular, Ensembl (Cunningham et al., 2015) is a joint scientific project, launched in 1999, between the European Bioinformatics Institute (EMBL-EBI) and the Wellcome Trust Sanger Institute. Ensembl is a centralized resource providing the most up-to-date genomic annotations, querying tools and access methods for chordates and key model organisms. Its annotations describe gene and transcript locations, gene sequence evolution, genome evolution, sequence and structural variants and regulatory elements. Ensembl includes full support for 69 species on the main website, plus partial support for 10 additional species on the Ensembl Pre! website (http://pre.ensembl.org). Some species have different level of annotation, reflecting the knowledge of the research community. For instance, comparing the statistics of the genome assembly of human and pig (**Error! Reference ource not found.**) we can infer that the pig genome annotation is probably lacking information about pig non-coding genes, transcripts, and variants.

Other databases are more specific for certain class of biological entities. One important example for small noncoding RNAs is the miRBase database (www.mirbase.org) (Kozomara and Griffiths-Jones, 2013). New miRNAs of many species are continuously discovered thanks to RNA-seq experiments (Friedländer et al., 2014; Londin et al., 2015). Their sequences can be deposited in miRBase, which is the major database resource for microRNA information. The current miRBase release (v.21) accounts 28,645 pre-miRNAs expressing 35,828 mature miRNA products, in 223 species. MiRNA names present a three letter prefix to designate the species followed by a numeric name in sequential order by date of discovery and classification. Orthologous or identical miR sequences are assigned the same numeric value (e.g. *ssc-mir-21* is the pig orthologous of the human *hsa-mir-21*). Paralogues are assigned with the same numeric value followed by a single letter suffix (e.g. *ssc-mir-199a* has one paralogue *ssc-mir-199b*) (Griffiths-Jones et al., 2006).

The growth of sequence and expression data derived from high-throughput technologies set the challenge of storing and sharing these data for scientific records, together with metadata about the specific experiments. Moreover, these repositories provide treasure resources for the experiment reproducibility and re-analysis, comparison with custom data, assessment and development of computational methods. Regarding high-throughput technologies, such as RNA-seq, of particular interest are the GEO (Barrett et al., 2013) and SRA (Leinonen et al., 2011) repositories.

The Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) is an international public repository for high-throughput microarray and next-generation sequence functional genomic data sets submitted by the research community. The resource supports archiving of raw data, processed data and metadata which are indexed, cross-linked and searchable. All data are freely available for download in a variety of formats. GEO also provides several web-based tools and strategies to assist users to query, analyze and visualize data. GEO accepts studies concerning quantitative gene expression, gene regulation, epigenetics, or other functional genomic studies.

The Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra/) is a public repository for the preservation of experimental data, in particular next-generation sequencing (NGS) data. The SRA is operated by the International Nucleotide Sequence Database Collaboration (INSDC). INSDC partners include the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI) and the DNA Data Bank of Japan (DDBJ). The SRA is accessible at http://www.ncbi.nlm.nih.gov/Traces/sra from NCBI, at http://www.ebi.ac.uk/ena from EBI and at http://trace.ddbj.nig.ac.jp from DDBJ.

GEO and SRA are tightly linked since GEO uploads to SRA the original raw data files containing sequence reads and quality scores. However, neither GEO nor SRA process transcriptome or transcript assemblies, which are reported in the Transcriptome Shotgun Assembly Database (http://www.ncbi.nlm.nih.gov/genbank/tsa) (Benson et al., 2013).

Genome annotations, like exon genome positions, transcript and gene exon structure, can be represented with the Genomic Feature Format (GFF) (http://www.ensembl.org/info/website/upload/gff.html). GFF files are plain text files in which each row represent a genomic feature and is composed by nine tab-separated fields. Also, all but the final field in each feature line must contain a value; "empty" columns should be denoted with a '.' character. The fields are:

1. *seqname* - name of the chromosome or scaffold; chromosome names can be given with or without the 'chr' prefix. Important note: the seqname must be one used within Ensembl, i.e. a standard chromosome name or an Ensembl identifier such as a scaffold ID, without any additional content such as species or assembly. See the example GFF output below.
2. *source* - name of the program that generated this feature, or the data source (database or project name)
3. *feature* - feature type name, e.g. Gene, Variation, Similarity
4. *start* - Start position of the feature, with sequence numbering starting at 1.
5. *end* - End position of the feature, with sequence numbering starting at 1.
6. *score* - A floating point value.
7. *strand* - defined as + (forward) or - (reverse).
8. *frame* - One of '0', '1' or '2'. '0' indicates that the first base of the feature is the first base of a codon, '1' that the second base is the first base of a codon, and so on.
9. *attribute* - A semicolon-separated list of tag-value pairs, providing additional information about each feature.

The following is an example line from the Ensembl pig genome annotation file that report annotation of the PLIN2 gene, located in the genomic scaffold GL892718.2 (N.B. the attribute field is split in two lines for layout reasons):

```
GL892718.2     ensembl gene    6383    13103   .       +       .       gene_id
"ENSSSCG00000026749";  gene_version  "1";  gene_name  "PLIN2";  gene_source  "ensembl";
gene_biotype "protein_coding";
```

1.4 THE ADIPOSE TISSUE AND PIG BACKFAT

1.4.1 ADIPOSE TISSUE FEATURES

Adipose tissue is a remarkably complex organ with important physiological role and pathophysiological relevance. Adipose tissue plays a major role in nutrient homeostasis, serving as the site of calorie storage after feeding and as the source of circulating free fatty acids during fasting. In addition, nowadays it is also regarded as an endocrine organ at the center of energy homeostasis maintenance processes. All eukaryotes from yeast to man are able to store calories in the form of lipid droplets, but only vertebrates have specialized cells that are recognizable as adipocytes (Ottaviani et al., 2011).

Two general types of adipose tissue exist in humans, white (WAT) and brown (BAT). White adipocytes store triglycerides and cholesterol in a single large lipid droplet (unilocular appearance), while brown adipocytes, which are present mainly in infants, contain several smaller lipid droplets (multilocular appearance). Recently a third type of adipocyte has been characterized and termed 'brite', for its 'brown-in-white' phenotype (Petrovic et al., 2010). Brite cells (also referred to as beige) are interspersed in white adipose tissue. These adipose tissue types share numerous attributes but also differ in critical ways that include aspects of their gene expression profile and secretome, their developmental origin, and their therapeutic potential.

WAT is a highly dynamic tissue capable of rapidly changing its mass according to the body's energy status; no other nonneoplastic tissue can change its dimensions to the same degree. This feature can be accomplished by increasing the size of individual cells (hypertrophy) or by recruiting new adipocytes from the resident pool of progenitors (hyperplasia). Adipocyte number in WAT in man is remarkably stable in adulthood. In the face of overnutrition, adipose depots expand first by hypertrophy until a critical threshold is reached (~0.7–0.8 ug/cell), upon which signals are released that induce the proliferation and/or differentiation of preadipocytes (Krotkiewski et al., 1983). Besides, ~10% of human subcutaneous adipocytes turn over each year, with birth and death rates matched to result in little change in total cell number (Spalding et al., 2008).

The major function of WAT is to store and release energy-rich lipids, forming a single droplet within a fat cell that constitutes >90% of its volume. The lipid droplet itself is a highly dynamic organelle with more than 200 droplet-associated proteins, most of which are also found associated with droplets in other mammalian tissues as well as in lower organisms (Konige et al., 2014). Fatty acids are metabolized to triglycerides within adipocytes in response to caloric intake. Depending on energy demands, the triglycerides are hydrolyzed to fatty acids (lipolysis), which are released to the blood stream to be used for oxidation in muscle and other body tissues.

WAT is not the same in all body depots. Distinct WAT depots differ substantially in their gene expression profiles, cell size and response to physiological factors such as hormones (Gil et al., 2011). Differences between visceral adipose tissue and subcutaneous adipose tissue are known and research is ongoing in this field since increased visceral adiposity and insulin resistance are strictly related. With respect to subcutaneous adipocytes, visceral adipocytes secrete more inflammatory factors such as tumor necrosis factor α (TNF-α) and leptin (Wronska and Kmiec, 2012). It appears that depot-specific differences in preadipocyte phenotype are established early during development and that each depot has its own unique gene expression signature. In fact, preadipocytes express gene signatures that are specific for their depot of origin even after isolation and prolonged passage under identical conditions(Macotela et al., 2012; Tchkonia et al., 2013).

1.4.1 ADIPOGENESIS

Adipocytes develop from preadipocytes, which themselves derive from precursor cells (Cawthorn et al., 2012). Adipocytes develop from mesenchyme, which is primarily of mesodermal origin. It has also been proposed that some adipocytes derive from hematopoietic precursors (McCullough, 1944), but it appears that this is not a major pathway (Berry and Rodeheffer, 2013; Koh et al., 2007).

Adipogenesis progresses through two main phases: determination and terminal differentiation. During determination, possible alternate fates of an adipose precursor cell become progressively restricted such that it becomes “committed” to the adipose lineage and becomes a preadipocyte. During terminal differentiation the preadipocyte acquires the characteristics of the mature adipocyte. We know much more about terminal differentiation and adipogenesis in vitro because of the use in studies of cellular models already committed to the adipose lineage.

Several signaling pathways are involved in the terminal differentiation phase. The Wnt and hedgehog pathways, studied in the frame of “bone-fat switch”, inhibit adipogenesis in favor of osteogenesis by inhibiting proadipogenic transcription factors like PPARγ (peroxisome proliferator-activated receptor gamma) and C/EBPα (CCAAT-enhancer-binding protein alpha) (Okamura et al., 2007; Xu et al., 2008). Non-canonical signaling via Wnt5b tends to promote adipogenesis, at least in part by blocking β-catenin-mediated signals from classic Wnt signals (Kanazawa et al., 2005). The IGF/insulin signaling is strongly proadipogenic (Garten et al., 2012). For many other pathways, for instance the TGFβ/BMP (transforming growth factor beta / bone morphogenetic protein) superfamily, it has been difficult to draw general conclusions because results depend on the specific ligand, cell type, stage of differentiation, or other experimental conditions. TGFβ and its downstream effector Smad3 have been shown to exert both pro- and anti-adipogenic actions in different in vitro and ex vivo models (Choy et al., 2000; Yadav et al., 2011). Among the BMPs, BMP2 and BMP4 have been shown to increase both osteogenesis and adipogenesis, depending upon other components of the differentiation cocktail, whereas BMP7 promotes brown adipogenesis specifically (Zamani and Brown, 2010).

Both hypertrophy and hyperplasia are tightly controlled, negatively or positively, by a combination of multiple transcription factors (Gregoire et al., 1998). Transcriptional cascades in adipogenesis see PPARγ as the "master regulator" of fat cell formation, as it is both necessary and sufficient for adipogenesis; PPARγ is so potent an adipogenic factor that it can drive non-adipogenic cells like fibroblasts and myoblasts to become adipocytes (Hu et al., 1995; Tontonoz et al., 1994). Other important inducers of adipogenesis are the bZIP factors C/EBPα, C/EBPβ, and C/EBPδ, with C/EBPβ and δ acting early in terminal differentiation. Differentiation is "locked in" by a positive feedback loop between PPARγ and C/EBPα (Rosen et al., 2002; Wu et al., 1999); a second positive feedback loop between PPARγ and C/EBPβ reinforces the decision to differentiate (Park et al., 2012). Many of these factors bind at common genomic "hot spots" with early factors establishing chromatin accessibility at the same locations that will later be bound by downstream factors (Siersbæk et al., 2012). Other transcription factors are known to promote or inhibit adipogenesis, in part by inducing or repressing expression of PPARγ (Cristancho and Lazar, 2011; Rosen and MacDougald, 2006). PPARγ, in turn, directly binds to and regulates a huge number of genes that control virtually all aspects of adipocyte metabolism. Additional transcription factors involved in adipose determination are Zfp423 (Gupta et al., 2010), which induces adipose lineage commitment by amplifying the effects of BMPs via a SMAD-interaction domain; Zfp521, which inhibits adipogenesis interacting with Ebf1 (Festa et al., 2011; Kang et al., 2012) and represses Zfp423; and Tcf7l1, which responds to confluency and mediates changes in structural proteins that regulate differentiation (Cristancho et al., 2011).

1.4.1 LIPOLYSIS

Lipolysis is the process that is required for fatty acids to be liberated from triglyceride so that they can be oxidized locally or by other organs. Lipolysis is driven by β-adrenergic signaling in the adipocyte, but other inducers (such as TNF-α) exist and may have physiological relevance (Rydén and Arner, 2007). The lipolytic machinery consists of at least three major enzymes and associated cofactors. The primary cleavage of triacylglycerol to diacylglycerols is performed by adipose triglyceride lipase (ATGL). The second enzyme in the pathway is hormone-sensitive lipase (HSL) that is the major diglyceride lipase in adipocytes. Monoglyceride lipase (MGL) completes the process by generating glycerol and free fatty acids. Together, these three enzymes account for >90% of the lipolytic activity in the adipocyte (Young and Zechner, 2013). ATGL, in particular, is highly regulated at both the transcriptional and posttranscriptional levels, including multiple phosphorylation events and translocation to the surface of the lipid droplet. It is activated by a protein cofactor called CGI-58, which is normally bound in an inactive state by the lipid droplet protein perilipin-1 (Plin1). PKA-dependent phosphorylation of Plin1 releases CGI-58, allowing it to bind and activate ATGL (Granneman et al., 2009). Conversely, ATGL is inhibited by a protein called G0S2, though its importance in vivo is still unclear (Yang et al., 2010).

Insulin is the major physiological suppressor of lipolysis, a process that becomes impaired in obesity even though insulin levels are high. Insulin blocks lipolysis in different ways. First, it activates phosphodiesterase 3b (PDE3b) via Akt-mediated phosphorylation; this has the effect of reducing intracellular cAMP levels and blocking PKA activation (Degerman et al., 1998; Kitamura et al., 1999). More recently, a non-canonical pathway has been described in which insulin blocks activation of PKA selectively on Plin1 through a PI3K-mediated, Akt-independent pathway (Choi et al., 2010). Over a slightly longer timescale, insulin also represses lipolysis by transcriptionally silencing lipase genes via repression of the transcription factors FoxO1 and IRF4 (Chakrabarti and Kandror, 2009; Eguchi et al., 2011). Interestingly, lipolysis is required for the generation of endogenous PPARα ligands.

1.4.1 ADIPOSE TISSUE INTERACTION WITH IMMUNE CELLS

Although mature adipocytes constitute >90% of WAT mass/volume, they account for less than 20% of the total cells in WAT (Eto et al., 2009). The other cells, collectively referred to as the stromal–vascular fraction, are a heterogeneous population of endothelial cells, macrophages, fibroblasts, stem cells and lymphocytes. Every gram of adipose tissue contains 1–2 million adipocytes and 4–6 million stromal-vascular cells, of which more than half are leukocytes (Kanneganti and Dixit, 2012). This composition makes WAT an important endocrine organ secreting adipokines, for instance leptin and adipsin, which regulate important physiological functions such as appetite, energy expenditure, insulin sensitivity, inflammation and coagulation (Hauner, 2005). Macrophages within the fat pad produce TNF-α and other proinflammatory cytokines(Weisberg et al., 2003; Xu et al., 2003a), an effect magnified by overnutrition (Hotamisligil et al., 1993), that significantly impair the insulin sensitivity of local adipocytes and also liver and muscle.

Hypertrophic obesity is associated with a chronic state of low-grade inflammation, characterized by high serum IL-6, TNF-α and C-reactive protein (CRP) levels and low adiponectin levels (Bahceci et al., 2007). The increased local production of inflammatory proteins can alter the lipolytic activity and adipokine functions of fat cells (Arner and Langin, 2014; Hotamisligil, 2006; Johnson and Olefsky, 2013). Adipose tissue macrophages not only arise from the recruitment of blood monocytes, but recently it has been shown that local macrophage proliferation contributes significantly to obesity-induced increases in macrophage number (Amano et al., 2014; Hashimoto et al., 2013; Jenkins et al., 2011; Qiu et al., 2014; Yona et al., 2013). Obese WAT is characterized by increased infiltration of macrophages and increased release of inflammatory adipokines, such as CCL2, TNF and IL-6 (Hotamisligil, 2006). The inflammatory environment within WAT impairs insulin signaling and induces oxidative stress and endothelial dysfunction, which leads to systemic insulin resistance (Maury and Brichard, 2010).

During the last decade, pig transcriptomic data have been obtained initially by expressed sequence tag sequencing (Chen et al., 2006; Gorodkin et al., 2007; Mikawa et al., 2004; Uenishi et al., 2004, 2007) and microarrays (Ferraz et al., 2008; Hornshøj et al., 2007; Moon et al., 2009; Zhou et al., 2013), which allowed the comparison of gene expression levels in several pig tissues. More recently, the RNA-seq approach was used to compare the transcription profile of different pig fat tissues or different pig breeds (Chen et al., 2011; Corominas et al., 2013; Jiang et al., 2013; Li et al., 2012b; Sodhi et al., 2014; Toedebusch et al., 2014; Wang et al., 2013a; Zhou et al., 2013). The differentially expressed genes reported in these studies are useful for investigating the metabolic pathways activated by or associated with increased fat deposition in the pig. However, the large amount of data produced and the results reported in literature are often barely comparable because of differences in the studied breeds, ages of the animals and fat deposition stages. Moreover, these studies identified several new genes and transcripts not reported in swine or other species.

1.4.1 MICRORNAS IN ADIPOSE TISSUE

miRNAs regulate adipogenesis and cell-specific functions of fat cells. In addition, miRNAs have been associated with impaired adipogenesis, insulin resistance and obesity-related inflammation. The role of miRNAs in lipid metabolism was first reported in Drosophila, where deletion of mir-14 increased the accumulation of triacylglycerol and diacylglycerol (Xu et al., 2003b). To date, most research evaluating the role of miRNA in adipose tissue has focused on human and mouse cell lines. Esau et al. (Esau et al., 2004) first identified a potential role for miR-143 in adipogenesis of human pre-adipocytes, and showed that inhibition of miR-143 decreased adipocyte differentiation. Other studies showed that miR-103 and the miRNA cluster miR-17–92 enhanced adipogenesis (Wang et al., 2008b; Xie et al., 2009), while the let-7 and miR-27 family of genes impaired adipogenic differentiation (Karbiener et al., 2009; Kim et al., 2010; Sun et al., 2009b). More recent human studies on the expression of miRNAs in adipose tissue found that the expression of miRNAs was adipose depot-specific (Klöting et al., 2009; Ortega et al., 2010) and that some miRNAs correlated with the morphology of adipose tissue, adipocyte size (Klöting et al., 2009) and metabolic (fasting glucose and/or triglycerides) parameters (Ortega et al., 2010). Inhibition of Drosha and Dicer repressed the differentiation of human mesenchymal stem cells into adipocytes (Oskowitz et al., 2008), supporting a role for miRNAs in adipocyte development. Dicer ablation from mouse preadipocytes blocks adipogenesis with impaired lipogenesis and downregulated expression of adipocyte markers such as PPARγ, TNF receptor superfamily member 6 (also known as FAS), GLUT4 (solute carrier family 2 facilitated glucose transporter member 4) and FABP4 (fatty acid-binding protein, adipocyte) (Mudhasani et al., 2010).

MiRNAs have been studied in adipose differentiation; at least 20 miRNAs have been shown to affect adipogenesis, though some are not specific for fat and appear to be required for mesenchymal cell differentiation generally (Oskowitz et al., 2008). Some miRNAs affecting adipogenesis target transcription factors like PPARγ and C/EBPα directly, whereas others regulate important signaling pathways like insulin-Akt, TGFβ, and Wnt (Chen et al., 2013). Moreover, miRNAs might regulate insulin sensitivity through actions on lipolysis, adipokines and insulin signaling pathways. miRNAs such as let-7 (Sun et al., 2009b; Yong and Dutta, 2007), miR-21 (Jeong Kim et al., 2009), miR-22 (Huang et al., 2012), miR-27 (Jeong Kim et al., 2009; Karbiener et al., 2009; Lin et al., 2009), miR-31 (Sun et al., 2009a; Tang et al., 2009), miR-130 (Lee et al., 2011), miR-138 (Yang et al., 2011), miR-145 (Guo et al., 2012a), miR-155 (Liu et al., 2011), miR-221/222 (Meerson et al., 2013; Parra et al., 2010), miR-224-

5p (Peng et al., 2013), miR-369-5p(Bork et al., 2011) and miR-448 (Kinoshita et al., 2010) inhibit adipogenesis in human, mouse and porcine cells.

A large number of miRNAs are expressed in human WAT (Keller et al., 2011; Ortega et al., 2010), but few seem to be differentially expressed (with reduced expression) in obese WAT compared with lean WAT. Nevertheless, there is lack of congruency between findings reported by different studies of miRNAs in obesity.

Table 2. Important miRNAs with dysregulated expression in white adipose tissue (WAT) of humans with obesity, and miRNAs associated with inflammation in human adipose tissue. Adapted from (Arner and Kulyté, 2015).

Up in obese WAT	**Down in obese WAT**		
miR-1229	let-7a	miR-145	miR-221
miR-125b	let-7d	miR-150	miR-26a
miR-146b	let-7i	miR-151-5p	miR-30c
miR-199a-5p	miR-125a	miR-16	miR-378
miR-21	miR-126	miR-17-5p	miR-484
miR-221	miR-130b	miR-185	miR-484-5p
miR-222	miR-132	miR-193a-3p/5p	miR-520
miR-342-3p	miR-139-5p	miR-193b/5p	miR-652
miR-519d	miR-141	miR-197	miR-659
miR-99a	miR-143	miR-200a/b	miR-92a
Inflammation associated			
let-7a	miR-150	miR-26a	miR-652
let-7d	miR-155	miR-26b	miR-671
miR-125a–5p	miR-17-5p	miR-28-3p	miR-883b-5p
miR-126	miR-181a	miR-325	miR-92a
miR-132	miR-193a/5p	miR-335	miR-95
miR-143	miR-193b	miR-433	miR-99a
miR-145	miR-221	miR-511	
miR-146b-5p	miR-222	miR-517a	

miRNAs, either directly or indirectly through regulatory elements such as transcription factors, influence the expression and secretion of inflammatory proteins involved in signal transduction networks that regulate the chronic low-grade inflammation in obesity (Klöting et al., 2009; Ortega et al., 2010; Sonkoly and Pivarcsi, 2009).Several individual miRNAs that control inflammation in WAT have been described (Ge et al., 2014; Hulsmans et al., 2011). Further, specific miRNAs regulate inflammation in human WAT through effects on CCL2 secretion (Ortega et al., 2010).

miRNAs in adipose tissue development, lipid metabolism and adipogenesis have been studied also in farm animals (Civelek et al., 2013; Guo et al., 2012b; Liu et al., 2010;

McDaneld et al., 2009; Wang et al., 2013c). However, investigations of adipogenic miRNAs of porcine backfat are scarce. A few works (Chen et al., 2012; Cho et al., 2010; Li et al., 2011, 2012b) have reported the identification and characterization of miRNAs from porcine subcutaneous adipose tissue in pigs of different breeds, ages and developmental stages. Nevertheless, there are a small number of matches among the results.

1.4.1 PIG ADIPOSE TISSUE ECONOMIC ASPECTS

Pigs (*Sus scrofa*) provide relevant biomedical models to dissect complex diseases, such as obesity and diabetes, due to their anatomical, genetic, and physiological similarities with humans (Koopmans and Schuurman, 2015; Schachtschneider et al., 2015; Walters et al., 2012). Besides, pigs are mostly important for the meat industry. The pig is among the best animals with respect to adipose accumulation and the adipose tissue is directly associated with the yield and the quality of meat. In particular, backfat deposition and fat traits are among the most important characteristics studied in pigs due to their strong relationship with the human nutritional value of pig products (Wood et al., 2008). The Italian pig breeding industry has peculiar aspects as more than 70% of the Italian pig production is dedicated to the processing of high quality Protected Origin Designation (POD) dry cured hams, together with a large number of other cured products that are recognized all around the world as superior products. Furthermore, they represent the basis of important economic districts in several Italian regions. For such high quality productions, meat (and cuts) with an excelled aptitude for salting and seasoning is needed (Bosi and Russo, 2010; Russo and Nanni Costa, 1995). The relevance of the quality of the raw material is underlined by the fact that Consortia of the PDO hams indicate the required characteristics for the fresh thigh and rules for the curing processing but also dictates rules for the genotypes (breeds and crosses) that are allowed, the age, the slaughtering weight of pigs and the feed that can be used. Among the requested meat and carcass traits, fatness (composition, quantity, and distribution in the carcass and hams) plays a key role, and influence all production chain. For example, an adequate and uniform fat coverage of the carcass is needed in order to allow the proper salt concentration within the muscles in typical dried cured hams and to reduce as much as possible the seasoning loss (Bosi and Russo, 2010; Čandek-Potokar and Škrlep, 2012). The strong negative correlation between backfat and unsaturated fatty acid content, together with the excessive decreasing of backfat thickness results in increasing of the unsaturated lipids and contributes to cause problems during ham seasoning.

Generally, between 10% and 30% of the variation in meat quality traits and meat products (ham), such as ultimate pH, color, water-holding capacity, drip loss, tenderness, marbling, etc. is determined by the genetic background of the animal (Sellier, 1998). The traits determining meat quality are difficult to improve by traditional selection, because the heritability of quality traits in pigs is quite low (Sellier, 1998), the measure for the quality traits is expensive and only possible after slaughter, selection is devoted to maintain a correct fat coverage level lowering genetic progress for other performance traits, and in Italian Large White race the phenotypes associated with adipose characteristics are expressed late during production life. Moreover, in order to account for the processing industry needs, pigs are slaughtered at about 160 Kg live weight (heavy pigs) and only after 9 months of age. These limits, together with feeding rules, long seasoning period of POD hams (at least 12 months), the need to reduce the environmental impacts of pig farms and slaughtering plants, represent constrains that the Italian pig industry have to face. Additional limit for the complete exploitation of the opportunities of a selection plan is the lack of knowledge on the genes, gene effects, and gene interactions (including miRNAs and regulatory interactions) affecting single qualitative characteristics of meat.

To date, the number of studies carried out on a homogeneous sample of individuals of the same breed reared on the same environmental conditions is poorly represented in the literature. Therefore, there is need for studies to identify gene expression profiles in porcine fat tissues and to gain insight into the gene regulatory relations in pig adipose tissue biology.

2 MATERIALS AND METHODS

2.1 SOFTWARE FOR RNA-SEQ DATA ANALYSIS

Different steps for the analysis include raw reads filtering (optional), alignment to a reference genome (if available, de novo aligning otherwise), expressed genes identification, transcript reconstruction (optional), gene/transcript expression quantification, differential gene expression assessment (optional). Methods for each step are described in the following.

2.1.1 READ PREPROCESSING

Reads resulting from sequencing experiments may contain sequencing errors like bases called incorrectly or with high uncertainty. Some sequencing technologies, such as Illumina, provide for each base the quality of the read signal encoding it in the FASTQ format. There are patterns of the qualities along the sequenced fragments typical of the technologies. For instance, Illumina sequencing shows decreased qualities in the last nucleotides sequenced (Figure 7) because of noisy signals derives from the subsequent washes of nucleotides at each cycle. Low quality of the reads may result in the impossibility of finding the corresponding region on the reference genome in the mapping step (see below), thus increasing the number of reads unmapped, or even been mapped in a wrong position if the bases are actually miscalled. A solution is to trim the reads of the low quality region, according to some a priori quality threshold, or up to a very low quality called base. With respect to just ignoring read quality under the assumption that incorrectly called bases will not map on the reference genome, this approach has the advantage of speeding up the subsequent processes reducing the time spent in trying to map artifacts, and increase the confidence on downstream analysis (Del Fabbro et al., 2013). Moreover, removing poor quality bases and/or reads can improve sequence assembly (Cox et al., 2010). In Illumina/Solexa sequencing, also the beginning of the read presents lower quality with respect to the central part. Hence, trimming should be performed on both ends of the fragment. There are many tools implementing the trimming step, like DynamicTrim (Cox et al., 2010), FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/download.html), Trimmomatic (Bolger et al., 2014). Average quality of the trimmed read and length of the trimmed region (or conversely length of the untrimmed part) could be used as a summarizing measure of the fragment reliability.

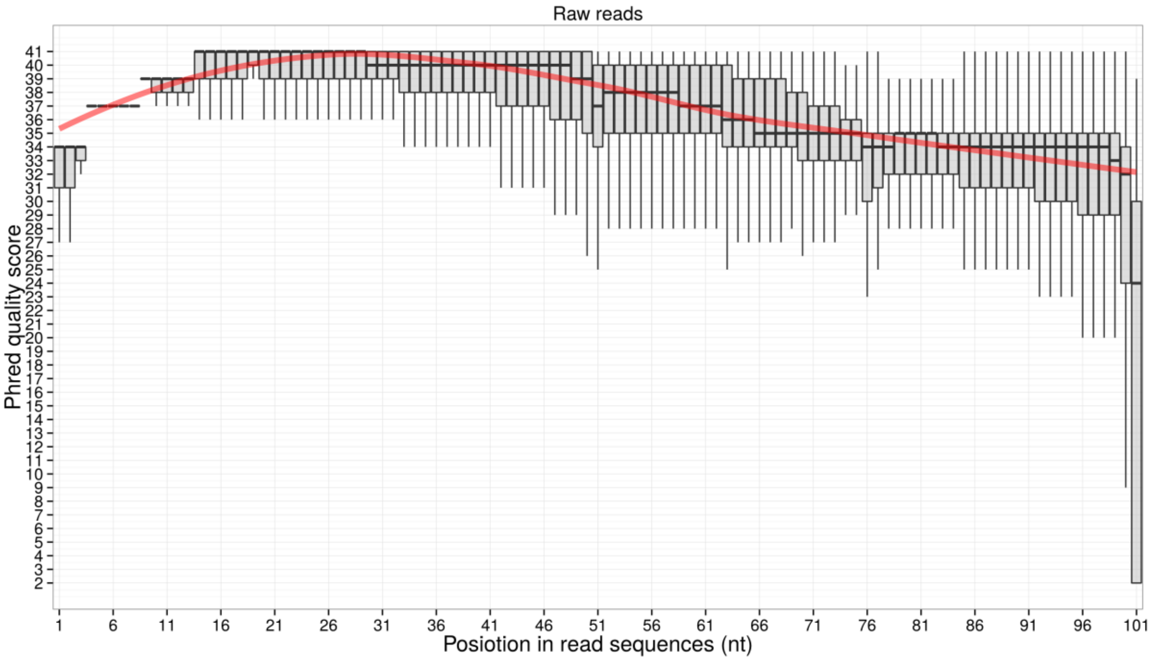


Figure 7. Illumina per-base read qualities. Phred qualities (vertical axis) of each base call (horizontal axis) are reported for a set of 100nt long reads sequenced with an Illumina HiSeq2000. A boxplot represent the range of quality values for each read position along the whole set of reads. The red curve smooths the median values along the read positions. The qualities are best between the 10th to 50th cycle and drops in the last part of the reads.

2.1.2 ALIGNMENT TO REFERENCE GENOME

The most important part in the pipeline is the alignment to the reference genome. We assume we can retrieve the nucleotide sequence for an organism's genome. For instance genome databases like the UCSC Genome Browser (https://genome.ucsc.edu/) and Ensembl (http://www.ensembl.org/) websites provides free download in FASTA format of many species genome sequences, including *Homo sapiens*, *Mus musculus*, *Sus scrofa*, and other vertebrates and eukaryotes.

The alignment of tens or hundreds millions of short reads to a whole genome reference sequence, usually of some billion nucleotides like in the case of human or pig (about three billion bases), is a task computationally different from the "classic" bioinformatics problem of local alignment and is not feasible to perform with methods such as BLAST (Altschul et al., 1990). Methods to cope with this issue have been developed concurrently with the advance of sequencing technologies and today there are at least 60 different tool for mapping short reads from NGS experiments (Fonseca et al., 2012). We can group aligners in two main categories, *unspliced* and *spliced* aligners, according to their ability to deal with transcriptomic data. Dissimilarly to DNA-seq data, RNA-seq reads may contain sequences that are not represented in the reference genome because transcripts from eukaryotic genomes do not include introns, which can span very large segments, transcripts can undergo post-transcriptional editing or polyadenylation, and can be the results of trans-splicing or back-splicing events. Most frequently, the read sequence may derive entirely from one single exon or from two (or more) exons, thus spanning exon-intron junctions (Figure 8 (3)). The *unspliced* aligners, such as BWA (Li and Durbin, 2009) and Bowtie (Langmead et al., 2009), do not consider splicing events occurring in the reads and fail to find a genomic position for the reads spanning splicing junction. On the contrary, the *spliced* aligners, such as TopHat (Trapnell et al., 2009), HISAT (Kim et al., 2015), GSNAP (Wu and Nacu, 2010), and STAR (Dobin et al., 2013), are designed in a way that they can recognize exon spanning reads and map their fragment to the corresponding exons.

One of the most popular methods is TopHat2 (Kim et al., 2013), the successor of TopHat. It is built around the Burrows-Wheeler transform (BWT)-based unspliced aligner Bowtie2 (Langmead and Salzberg, 2012) that allows a fast and memory efficient way to map the reads on an entire genome, with high accuracy also across small insertions and deletions (indels). TopHat2 inherits from TopHat the ability to discover novel splice sites by the read mappings, and can also predict novel fusion transcripts (Kim and Salzberg, 2011). TopHat2 proceeds by aligning the reads against the known transcriptome, if an annotation file is provided, using Bowtie2 (Figure 8(1)). The unmapped reads and the poorly mapped reads are aligned against the genome, indeed placing reads contained entirely in exons and leaving unmapped the reads spanning multiple exons (Figure 8 (2)). A third phase identifies novel splice sites according to known junction signals (GT-AG, GC-AG, and AT-AC) from the 2nd phase unmapped reads. Unmapped reads are split into smaller non-overlapping segments which are then aligned to the genome and re-aligned for the entire read sequence if left and right mapped segments are within the maximum intron size distance (Figure 8 (3)). The genomic flanking sequences of these potential splice sites are concatenated to form a novel transcriptome. Paired-end reads are processed independently and in the final stage the aligned reads are analyzed together to produce paired alignments. TopHat2 allow any read to map to multiple genomic regions and reports up to a user defined number (10 by default) of different alignment genome positions. Reads mapping in more than the threshold multiple alignments are assumed to have low-complexity sequence and are simply discarded. Using an approach similar to the 3rd step, TopHat2 is able also to detect indels and fusion breakpoints. TopHat2 strategy results to be fast and memory efficient with respect to competitors (Kim et al., 2013).

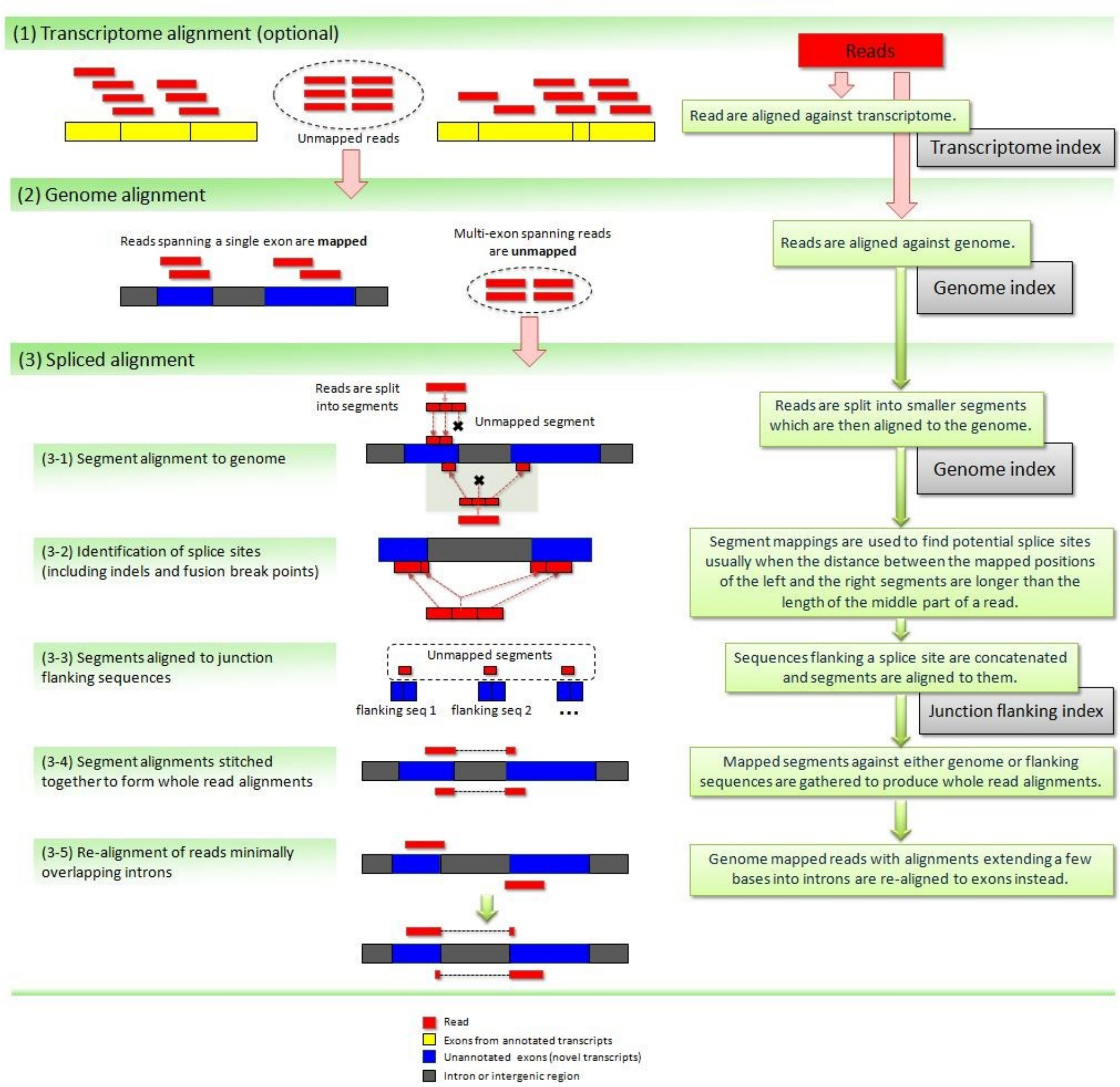

(1) Transcriptome alignment (optional)
Unmapped reads
Reads
Read are aligned against transcriptome.
Transcriptome index
(2) Genome alignment
Reads spanning a single exon are mapped
Multi-exon spanning reads are unmapped
Reads are aligned against genome.
Genome index
(3) Spliced alignment
Reads are split into segments
Unmapped segment
(3-1) Segment alignment to genome
Reads are split into smaller segments which are then aligned to the genome.
Genome index
(3-2) Identification of splice sites (including indels and fusion break points)
Segment mappings are used to find potential splice sites usually when the distance between the mapped positions of the left and the right segments are longer than the length of the middle part of a read.
(3-3) Segments aligned to junction flanking sequences
Unmapped segments
flanking seq 1
flanking seq 2
...
Sequences flanking a splice site are concatenated and segments are aligned to them.
Junction flanking index
(3-4) Segment alignments stitched together to form whole read alignments
Mapped segments against either genome or flanking sequences are gathered to produce whole read alignments.
(3-5) Re-alignment of reads minimally overlapping introns
Genome mapped reads with alignments extending a few bases into introns are re-aligned to exons instead.
Read
Exons from annotated transcripts
Unannotated exons (novel transcripts)
Intron or intergenic region

Figure 8. Tophat2 workflow. (1) Reads are first mapped to the transcriptome in an unspliced manner. (2) Transcriptome unmapped reads are aligned to the genome in unspliced manner, again. (3) Genome unmapped reads probably contains exon junctions and undergo spliced alignment. Figure from (Kim et al., 2013).

Box: Burrows-Wheeler transform

The Burrows-Wheeler transform (BWT) is a reversible permutation of the characters in a text (T). It is constructed adding a special character, say “$”, that is not present in T and that is lexicographically less than the characters in T. The matrix of cyclic rotations of characters in T$ is sorted lexicographically by row. The last column of the matrix is the BWT of T, BWT(T) (Figure 9a). The permutation matrix has the “last to first (LF) mapping” property, for which the characters in the last column occur in the same order of the first column. For instance, in Figure 9a the “a”s of the first column occur in the same order as they occur in the last column. This property can be exploited for the matching through the exact-matching in an FM-index algorithm (Ferragina and Manzini, 2001) and for the

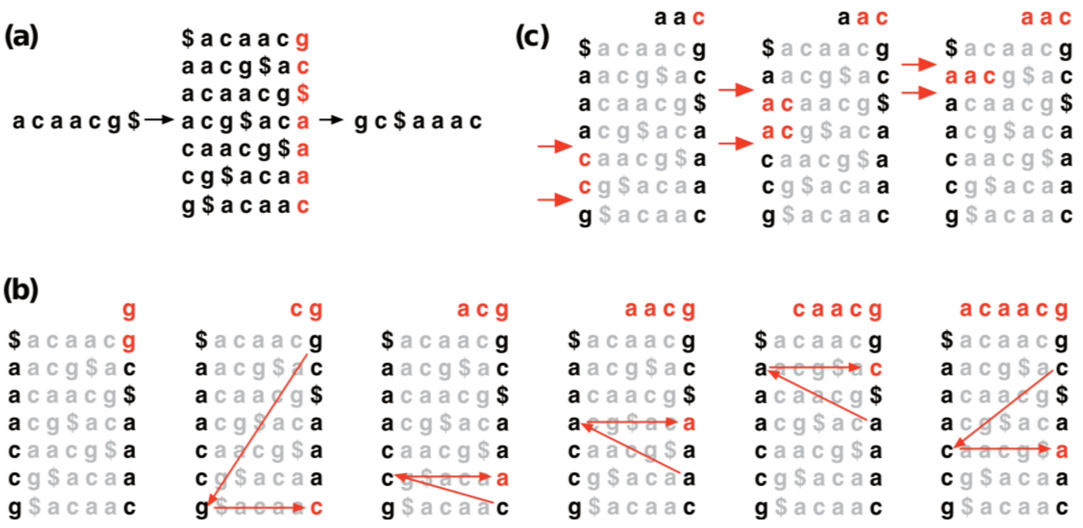


Figure 9. Burrows-Wheeler transform. (a) The Burrows-Wheeler matrix and transformation for 'acaacg'. (b) Steps taken to identify the range of rows, and thus the set of reference suffixes, prefixed by 'aac'. (c) application of the last first (LF) mapping strategy to recover the original text (in red on the top line) from the Burrows-Wheeler transform (in black in the rightmost column). Figure from (Langmead et al., 2009).

reconstruction of the original text from the transformation (Langmead et al., 2009; Li and Durbin, 2009). Indexing the reference genome by this approach results to be an efficient strategy for mapping reads that do not require large memory as for hash-based aligners.

2.1.3 TRANSCRIPTOME RECONSTRUCTION

Transcriptome reconstruction is the process of inferring the transcript isoforms from the fragmented reads. Most accurate methods rely on the mapping of reads to the reference genome (see Box "de novo transcriptome assembly" for non-genome-guided assemblers).

Cufflinks (Trapnell et al., 2010) is one of the most widely used transcript assembler (Pertea et al., 2015). It relies on splice junctions identified from the mapping of the reads to the reference genome to infer transcript exon chains. It is able to reconstruct transcript sequences also without gene annotation, reaching high recall (Steijger et al., 2013). However, better results are achieved if reference annotation is provided, also for the organisms where deep annotations do not already exist (Roberts et al., 2011a).

Cufflinks clusters the reads and build a graph model to represent all possible isoforms for each gene. Its approach is to generate the minimum number of transcripts that will explain all reads (splice events detected) in the graph (Figure 10b-e). From the read mappings (Figure 10a), it first identifies the 'incompatible' read pairs (fragments) that could be originated from distinct transcript isoforms (Figure 10b). Then, it builds a graph connecting the compatible fragments and assembles isoforms from the overlap graph

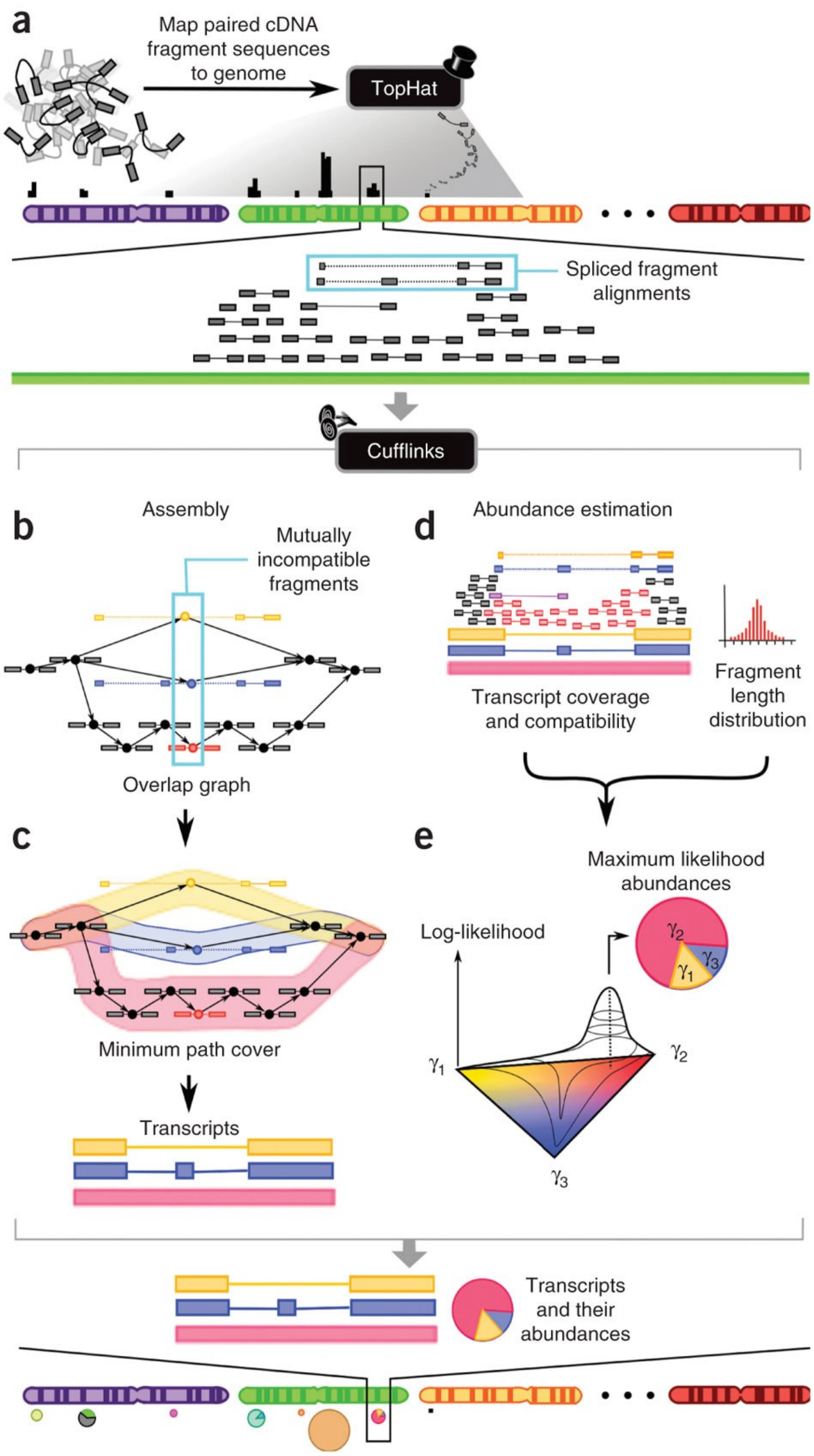

a
Map paired cDNA fragment sequences to genome
TopHat
Spliced fragment alignments
Cufflinks
b
Assembly
Mutually incompatible fragments
Overlap graph
c
Minimum path cover
Transcripts
d
Abundance estimation
Transcript coverage and compatibility
Fragment length distribution
e
Maximum likelihood abundances
Log-likelihood
γ_1
γ_2
γ_3
Transcripts and their abundances

Figure 10. Cufflinks workflow. Details in main text. Figure from (Trapnell et al., 2010)

(Figure 10c). In addition to the transcript reconstruction, Cufflinks quantifies expression level for the transcript isoforms with a statistical model by estimating the probability for each fragment to be originated by an isoform (Figure 10d) and maximizing the likelihood of the possible sets of relative abundances of the isoforms (Figure 10e). Gene expression is defined as the sum of all the transcript abundances for the gene, and is reported in fragments per kilobase of transcript per million fragments mapped (FPKM) units which should reflect the relative abundances of transcripts in terms of the expected biological objects (fragments) observed from the RNA-seq experiment.

Box: *de novo* transcriptome assembly

D*e novo* transcriptome assembly, or "genome-independent" reconstruction, is performed without a reference genome guiding the assembly. It is more challenging than guided assembly and generally less accurate. For this, it is mostly used when the organism genome sequence is not available. Methods performing *de novo* assembly use the reads to build consensus transcripts. A common strategy is to model overlapping subsequences (k-mers) by a de Bruijn graph, reducing the complexity of handling millions of reads, and then traverse the graph to assemble each isoform directly [Velvet (Zerbino and Birney, 2008); Trinity (Grabherr et al., 2011); Oases (Schulz et al., 2012)], or post-processing the assembly merging the contigs (Trans-ABySS (Birol et al., 2009)). See Martin et al. 2011 (Martin and Wang, 2011) for a review of transcriptome assemblers.

2.1.4 GENE EXPRESSION QUANTIFICATION AND DIFFERENTIAL EXPRESSION ASSESSMENT

Expression level estimation and differential expression assessment from RNA-seq data present the problem that more reads will map to longer genes/transcripts even when expression level is the same. Another challenge is given by the comparison of different sequencing depth in different samples, which require normalization of the raw counts.

The first method suggested to compare expression levels within the same experiment genes and among different experiments was the RPKM (Reads Per Kilobases per Million mapped reads) measure (Mortazavi et al., 2008). However, this still widely-used approach has proven ineffective and more beneficial procedures have been proposed (Anders and Huber, 2010; Bullard et al., 2010; Hansen et al., 2012; Robinson and Oshlack, 2010). Trapnell et al. (2010) proposed the FPKM (Fragments Per Kilobase per Million mapped reads) measure, which is a generalization of the RPKM and that considers as count units mapped transcript fragments (like paired-end reads) instead of single mapped reads. In small RNA sequencing the reads cover the full length of the transcripts, so the expression can be represented as simple read counts. Other biases in gene expression estimation can depend upon the technologies used. For instance, Illumina sequencing is sensible to GC content, which can alter read counts in comparisons between genomic regions for a given sample (Risso et al., 2011). Moreover, quantification depends on the mapping ability of the aligner (repetitive sequences are difficult to align) and whether multiple mapped reads (derived from repeat regions and paralogs genes that result to be ambiguous; usually aligners set a threshold for the maximum number of loci in which a read is mapped) are considered or not for quantification. In differential expression analysis these biases are usually ignored since they are assumed to affect all samples similarly. Methods to assess differential expression from RNA-seq data principally represent expression data with simple count-based probability distribution, such as Poisson (Marioni et al., 2008). The Poisson distribution comes naturally for count data like RNA-seq but considers the variance equal to the mean, an assumption that can result to be too restrictive and may predict smaller variations than what is seen in the data, condition that is called overdispersion. Recent methods, like *edgeR* (Robinson et al., 2010) and *DESeq* (Anders and Huber, 2010) to cite the most popular ones, tackle this issue by representing the number of reads assigned to each gene using a negative binomial (NB) distribution. With respect to the Poisson distribution, the NB allows a larger variance, specified with an additional parameter. Despite sharing the same distribution to model the data, these two methods differ for the way they estimate the dispersion for each gene. *edgeR* estimates dispersion by relating the variance to the mean proportionally with a constant, *DESeq* estimates the overdispersion parameters from the data with a local regression (GLM gamma family). In particular, *DESeq* describes the read count with a generalized linear model (GLM) of the NB family with a logarithmic link. For the differential expression test DESeq uses a GLM, which allows complex experimental designs.

DESeq2 (Love et al., 2014) is an improvement of the *DESeq* method in which the authors included shrink dispersion estimates according to an empirical Bayes approach driven by the data to account for gene-specific variation, indeed reducing overestimation of the dispersions (Figure 11). *DESeq2* also shrinks the estimates of logarithmic fold changes to reduce the typical bias observed in low count genes (Figure 12). Moreover, the differential expression is assayed using a Wald test on the shrunken estimates of logarithmic fold changes (LFCs), obtaining P-values and the relative Benjamini-Hochberg (BH) corrected values.

Cuffdiff2 (Trapnell et al., 2013) is a methods belonging to the *Tuxedo* (Trapnell et al., 2012) software suite as part of *Cufflinks2*. In its most recent version (v2.1 at the time of this writing), it models the read counts with a beta-negative binomial distribution fitting a GLM, and the dispersion estimate is performed similarly to *DESeq*. The mixture distribution has the advantage of modeling the uncertainty of multiple mapped read fragments that are shared among different isoform transcripts. This occurs because in higher eukaryotes alternative isoforms of most genes share large amounts of sequence, and many genes have paralogs with high sequence similarity. Indeed, by calculating the

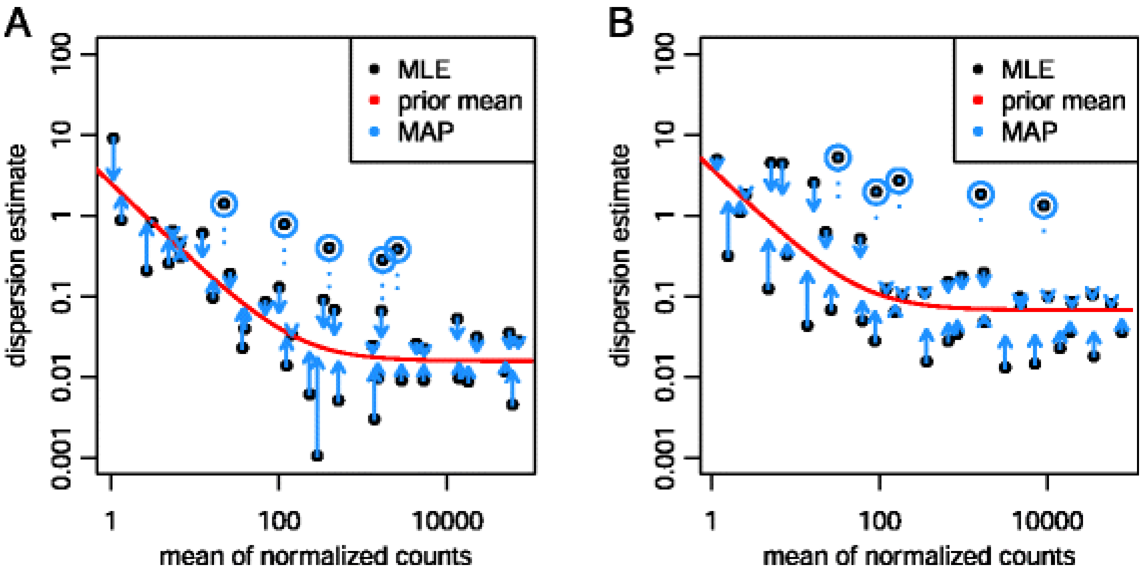


Figure 11. Shrinkage estimation of dispersion. Plot of dispersion estimates over the average expression strength (A) dataset with six samples across two groups and (B) for five samples from another dataset, fitting only an intercept term. First, gene-wiseMLEs are obtained using only the respective gene's data (black dots). Then, a curve (red) is fit to the MLEs to capture the overall trend of dispersion-mean dependence. This fit is used as a prior mean for a second estimation round, which results in the final MAP estimates of dispersion (arrow heads). This can be understood as a shrinkage (along the blue arrows) of the noisy gene-wise estimates toward the consensus represented by the red line. The black points circled in blue are detected as dispersion outliers and not shrunk toward the prior (shrinkage would follow the dotted line). MAP, maximum a posteriori; MLE, maximum-likelihood estimate. Figure from (Love et al., 2014).

confidence that each fragment is correctly assigned to the transcript that generated it (transcripts with more shared exons and few uniquely assigned fragments will have greater uncertainty) *Cuffdiff2* can estimate expression of transcripts and represent gene expression as the sum of the gene isoforms' expression. *Cuffdiff2* then uses t-tests to compute P-values and uses BH correction to quantify differential expression significance.

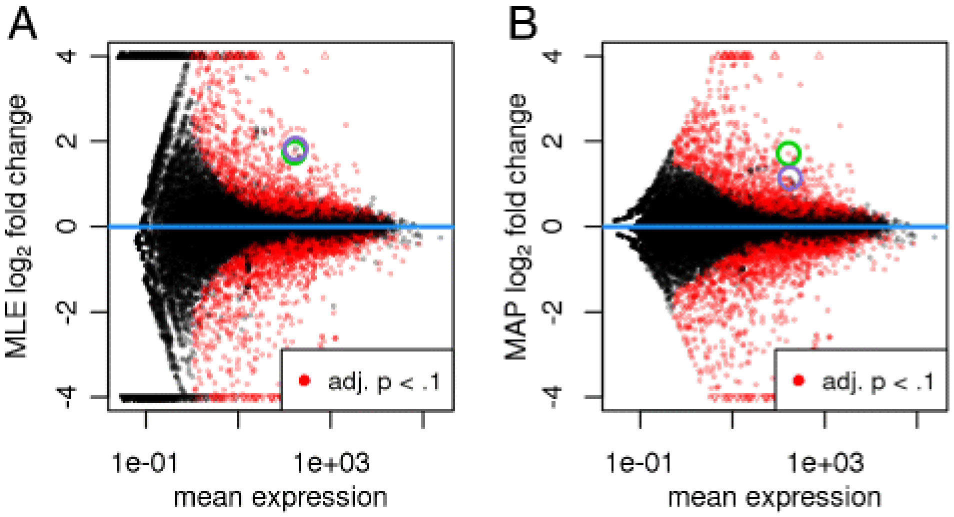


Figure 12. Effect of shrinkage on logarithmic fold change estimates. Plots of the (A) MLE (i.e., no shrinkage) and (B) MAP estimate (i.e., with shrinkage) for the LFCs, over the average expression strength for a ten vs eleven sample comparison. Small triangles at the top and bottom of the plots indicate points that would fall outside of the plotting window. Two genes with similar mean count and MLE logarithmic fold change are highlighted with green and purple circles. Figure from (Love et al., 2014).

2.2 SOFTWARE FOR SMALL RNAS' RNA-SEQ DATA ANALYSIS

2.2.1 SMALL RNAS DETECTION

MiR&moRe is a pipeline for the analysis of small, miRNA-like, RNAs' RNA-seq data described in (Bortoluzzi et al., 2012). *MiR&moRe* can detect known miRNAs and quantify their expression. Moreover, it allows the investigation on expressed sequence variations and the detection of microRNA offset RNAs (moRNAs). The pipeline's steps (Figure 13) consider a preprocessing of the raw reads; their mapping to the human reference genome; the mapping also to human miRNA precursors including flanking bases; the identification of unknown sister miRNAs, moRNAs, and known-miRNA isomiRs; and the quantification of expression for the detected small RNAs.

The preprocessing step removes the adapter from raw sequences, discards the reads in which the adapter sequence was not detected and the clipped reads that are not within miRNA-like sizes (longer than 30nt or shorter than 18nt as default), plus those reads having average quality below a user defined threshold. The reads passing the filters are likely to represent full miRNA/moRNA sequences and are mapped with Bowtie to the human genome. Reads mapping to more than five positions in the genome are further discarded. The remaining reads constitute the clean set of reads representing the small RNAs expressed. The clean reads are mapped to miRNA precursors' sequences that are extended 30 nucleotides upstream and downstream the precursor boundaries, in order to allow the detection of potential moRNAs. The alignments allows for mismatches on the known mature miRNAs, restricting to one mismatch on the seed region (the conserved sequence ~6nt long that is the major driver of the targeting) or two mismatches on the 3' end at maximum to account for post-transcriptional editing of the molecule. Other variations to identify isomiRs consider longer or shorter sequences of the known miRNAs.

miRDeep2 (Friedländer et al., 2011) is popular software with functionalities similar to miR&moRe. It processes RNA-seq data of miRNA sequencing experiments to identify and quantify the expressed molecules. Contrary to *miR&moRe*, it was applicable also to non-human data and it is able to predict novel miRNA precursors and mature miRNAs from the organism's genome. However, miRDeep2 is limited to miRNA identification and quantification, isomiRs identification and quantification are reported in a format that is not easy to manipulate for downstream analysis and moRNAs' analysis is not conceived.

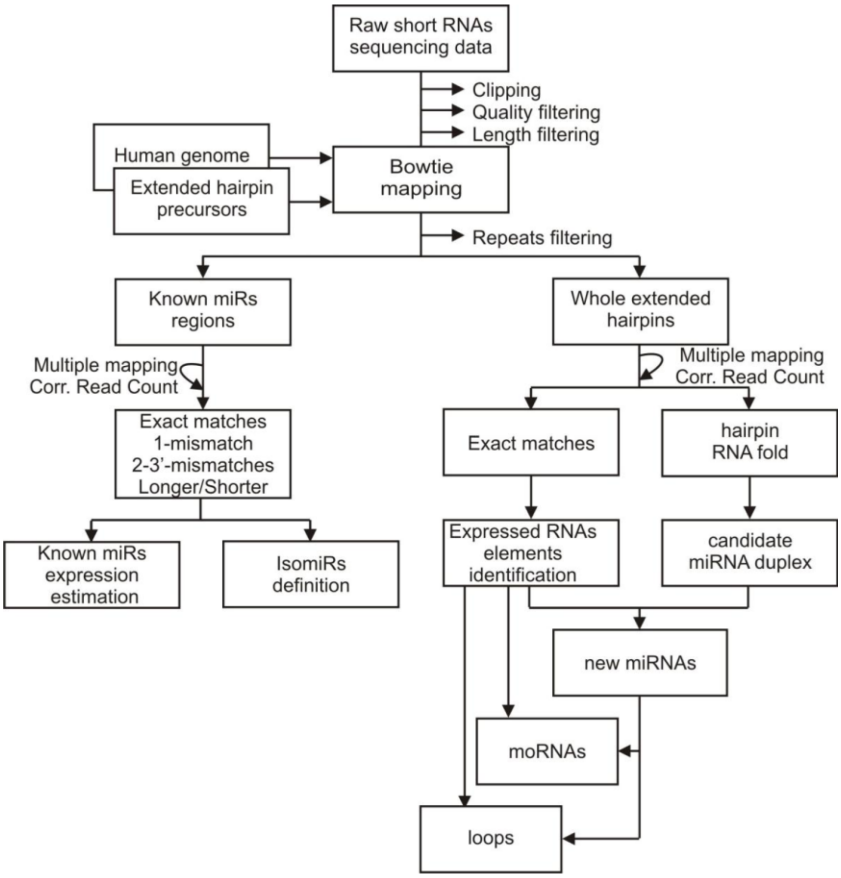

Raw short RNAs sequencing data
Clipping
Quality filtering
Length filtering
Human genome
Extended hairpin precursors
Bowtie mapping
Repeats filtering
Known miRs regions
Whole extended hairpins
Multiple mapping Corr. Read Count
Multiple mapping Corr. Read Count
Exact matches 1-mismatch 2-3'-mismatches Longer/Shorter
Exact matches
hairpin RNA fold
Known miRs expression estimation
IsomiRs definition
Expressed RNAs elements identification
candidate miRNA duplex
new miRNAs
moRNAs
loops

Figure 13. *MiR&moRe* pipeline workflow. The pipeline performs a preprocessing of raw small RNA-seq reads before the mapping to the human reference genome and in parallel to the miRNA hairpin precursor sequences, extended with flanking bases. The mappings with no more than five (default value) multi-mapped loci are then processed in two "branches": one for the characterization of isomiRs (left branch), the other for identification and quantification, from known hairpins, of known miRNAs, new sister miRNAs, moRNAs and loops (right branch). Figure from (Bortoluzzi et al., 2012).

2.2.2 SMALL RNA-TRANSCRIPT PUTATIVE INTERACTIONS

A single miRNA can target potentially hundreds or even thousands of mRNAs. Since validating a potential target in the laboratory is time consuming and costly, the use of a computational approach that narrows down the selection of the miRNA-transcript target interactions is a critical initial step before experimental validation. Currently there are several miRNA target prediction methods and software tools (Peterson et al., 2014), whose models share common features derived by considerations on the molecular mechanisms characteristic of the miRNA targeting process, such as seed matching, conservation, binding free energy, and binding-site accessibility. Table 3 reports some of the most popular tools.

Table 3. Popular miRNA target prediction tools, with name, website for software download or web server online use, and article reference.

Tool name	Website	Publication references
miRanda	http://www.microrna.org/	Enright et al., 2003; John et al., 2004
miRanda-mirSVR	http://www.microrna.org/	Betel et al., 2010
TargetScan	http://www.targetscan.org	Agarwal et al., 2015
RNA22-GUI	https://cm.jefferson.edu/rna22/	Miranda et al., 2006
PITA	http://genie.weizmann.ac.il/pubs/mir07/	Kertesz et al., 2007
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/	Krüger and Rehmsmeier, 2006

One of the earlier tools for miRNA target prediction is miRanda (Enright et al., 2003; John et al., 2004). *MiRanda* is a method based on three properties, sequence complementarity with the target sequence, binding energy, and evolutionary conservation of the target sites. The sequence complementarity stage uses a Smith-Waterman-like dynamic programming algorithm that takes into account G-U wobble pairs, allows moderate insertions and deletions, and uses a weighting scheme that rewards complementarity at the 5p end of the miRNA. The strength of binding is calculated as the free energy of optimal strand-strand interaction using the Vienna secondary structure programming library (Wuchty et al., 1999). The conservation step is not embedded in the main algorithm, but is performed as a post-processing filtering of the interactions predicted by the previous steps, and is species dependent. The *miRanda* model does not consider any additional protein interaction, such as with RISC. For practical use, the selection of a target prediction tool to use has to take into account additional characteristics like the tool maintenance, user-friendliness, and adaptability to the user needs. The *miRanda* software (latest update 8/2010, current version v3.3a) and source code can be downloaded (www.microrna.org) and used from the command line. As input it requires the sequences of both miRNAs and (UTR) transcripts, and it allows the specification of some parameters like the free energy threshold, alignment threshold, weight of seed region, and gap penalty. *MiRanda* is also available online as part of the *miRanda-mirSVR* tool (Betel et al., 2010) providing pre-computed targets for humans, rats, mice, flies, and worms.

2.2.2 ENRICHMENT OF TRUE MIRNA-TRANSCRIPT PREDICTED RELATIONS

The regulatory network of a miRNA is probably dynamic. Only a proportion of the miRNA complementary sites that are annotated transcriptome wide will be present and relevant in any given cell. For this reason, pruning the search of miRNA target to the set of transcripts expressed in specific condition of a specific tissue will reduce the amount of false miRNA-mRNA interactions predicted. In addition, methods for target prediction are affected by high rates of false positive predictions, and combining different methods does not give reliable results (Ritchie et al., 2009). As suggested by (Huang et al., 2007; Lionetti et al., 2009), an approach to refine predicted interactions is to select on the strength of correlation between the miRNA's expression and its gene/transcript targets' expression. Considering the silencing effect of miRNAs on their target transcripts' expression, a conservative cut off could be to restrict only to the negative correlated targets. The correlation computation requires the samples to be "matched" for transcript and small RNA sequencing experiments i.e. both the transcript and the small RNA expression levels were estimated in each sample.

MiRanda is a tool suitable for this approach because it provides non-pre-computed target predictions. In fact, many other target prediction approaches are released only as repositories of miRNA target predictions pre-computed for restricted number of species, commonly *Homo sapiens*, *Mus musculus* and *Drosophila melanogaster*. Instead, the *miRanda* algorithm implementation is available and can be used with custom input sequences of miRNA and potential target transcripts. In addition, a conservation filter is not applied in the *miRanda* executable script, allowing for a wider application of the software. To reduce the false positive interaction number, the target prediction can be filtered *a posteriori.* A filtering based on evolutionary conservation was applied in several studies (Enright et al., 2003) but also criticized by (Betel et al., 2010), as several non-conserved target sites are functional.

2.3 PIG SUBJECT SELECTION AND BACKFAT SAMPLE COLLECTION

RNA-seq data from Italian Large White (ILW) pig backfat samples were provided by Prof. R. Davoli and Dr. P. Zambonelli (DISTAL - University of Bologna), in the frame of a research project in pig genomics for the Italian heavy pig production chain. Backfat samples of 20 ILW pigs were collected from a purebred population of 949 ILW sib-tested pigs provided by the Italian National Association of Pig Breeders (Associazione Nazionale Allevatori Suini, ANAS; http://www.anas.it). Animals were selected to compose two groups (LEAN and FAT) of 10 pigs showing extreme and divergent characteristics for the backfat thickness (BFT) estimated breeding value (EBV) (see Box "Pig rearing and sample collection procedure"). With respect to the larger population that presented EBVs for BFT ranging from -10.64 mm to 7.28 mm (mean value -1.96 mm and standard deviation (SD) 3.01), BFT mean values of LEAN and FAT groups were outside (plus or minus) two standard deviations from the population mean value (-7.98 mm to 4.06 mm range). Specifically, FAT and LEAN animals were associated with average BFT values of +5.22 mm

(± 1.30 SD) and -8.63 mm (± 1.40 SD), as indicated in Table 4. The 20 animals were slaughtered on 12 dates, with five dates common to both groups (Table 4). The animals were selected also according to their pedigree to avoid the presence of full sibs in the considered groups and with a 1:1 sex ratio within each group. The collected samples were immediately frozen in liquid nitrogen and stored at 80 °C until RNA extraction.

Table 4. Genetic indexes and phenotypes for BFT and hot carcass weight of the pigs selected for the transcriptome analysis.

Group	Sample ID	Sex	Day of slaughter	Slaughter weight (kg) (*)	BFT phenotype (mm)	BFT EBV		
							Mean	SD
FAT	477	M	6	120	43	7.36	5.22	1.3
	476	F	6	119	37	7.17		
	474	M	2	113	38	6.03		
	482	F	9	-	42	5.75		
	478	F	7	118	33	5.05		
	516	F	3	115	36	4.88		
	479	M	8	-	41	4.76		
	483	F	10	119	38	4.41		
	489	M	18	108	35	3.54		
	484	M	15	128	35	3.27		
LEAN	490	M	19	113	24	-6.46	-8.63	1.4
	473	F	2	132	23	-7.54		
	487	M	18	110	23	-7.61		
	517	M	4	117	20	-7.71		
	485	F	17	126	20	-7.82		
	475	M	5	119	20	-8.03		
	481	M	9	-	22	-9.91		
	486	F	17	123	19	-10.27		
	488	F	18	128	19	-10.37		
	480	F	9	-	16	-10.59		

BFT, backfat thickness; EBV, estimated breeding value.
(*) Slaughter weight: the hot carcass slaughter weight is reported. For four animals the weight was not available due to a problem of the automatic recording system at the slaughterhouse.

Total RNA was extracted with Trizol (Invitrogen) according to the manufacturer's instructions. Results of the extraction were quantified using a Nanodrop ND-1000 spectrophotometer, and the quality of the extracted RNA was assayed using an Agilent 2100 BioAnalyzer. The long RNA libraries were prepared from total RNA using the TruSeq RNA sample preparation kit (Illumina) and version 3 of the reagents, following the manufacturer's suggested protocol. Pairs of libraries were run on a single lane of an Illumina HiSeq2000. Reads were 100 nucleotide (nt) paired-end, represented in FASTQ format. Small RNA libraries were prepared from total RNA using the TruSeq Small RNA kit (Illumina) and version 3 of the reagents following the manufacturer's suggested protocol.

Two libraries were run on a single lane of an Illumina GAII; the other 18 small RNA libraries were run on an Illumina HiSeq2000.

Box: Pig rearing and sample collection procedure

All animals were kept according to Italian and European law for pig production, and all procedures described were in compliance with national and European Union regulations for animal care and slaughtering. The animals were reared on the ANAS Sib-Test genetic station from about 30 kg live weight to at least 155 kg live weight. For the genetic evaluation of a boar, full sib triplets (two females and one castrated male) were farmed on the genetic station to be performance tested. The formula and amount of the ration was the same for all and was based mainly on cereals and soybean, given in excess calculated using the *quasi ad libitum* rule (a ration sufficiently abundant that 60% of pigs were able to ingest the full supplied food). At the end of the tests, animals were transported to a commercial abattoir located about 25 km from the test station according to Council Rule (EC) No 1/2005 on the protection of animals during transport and related operations and amending Directives 64/432/EEC and 93/119/EC and Regulation (EC) No 1255/97. At the slaughterhouse, the pigs were electrically stunned and bled in a lying position in agreement with Council Regulation (EC) No 1099/2009 on the protection of animals at the time of killing. All slaughter procedures were monitored by the veterinary team appointed by the Italian Ministry of Health. Estimated breeding values (EBV) can be defined as a genetic merit for a phenotypic trait, one half of which will be passed to the progeny; EBVs are expressed in unit of measurement of the specific trait and refer to difference from a fixed average value. Backfat thickness EBVs were calculated by ANAS for the animals as described by (Russo et al., 2000, 2008). EBVs were determined through a BLUP multiple-trait animal model procedure (Henderson and Quaas, 1976) using the BFT, measured in mm, recorded post-mortem in correspondence with the *gluteus medius* muscle. The model included fixed effects of batch in test, sex, age at beginning of test, age of sow, weight at slaughter, age at slaughter and inbreeding coefficient as well as the random effects of litter, individual permanent environment and animal. Pigs' genetic merit for the BFT trait was calculated taking into account the additive relationship matrix. EBVs were expressed as differences from the genetic mean value for the considered trait in the year 1993. The BFT genetic index may present as a negative value because the value of the trait refers to the fixed genetic base defined by ANAS as mean values of the pigs born in 1993 which is considered as 'zero', so the more negative values indicate lower values of BFT. After slaughter, backfat samples were collected from 949 ILW pigs slaughtered at an average hot carcass weight of 118.97 kg (0.29 SEM) and at an average age of 8 months during the years 2011 and 2012 on 27 different slaughtering days.

3 RESULTS

The present work has both methodological and applicative results. The computational analysis workflow (Figure 32) could be seen as a method for the investigation of the transcriptome in a tissue, to characterize long and short transcripts expressed in terms of their sequence variation, abundance, and putative regulatory interactions. Moreover, given the comparative design of the experiment, this approach can highlight gene and transcript expression differences between two sample groups representing different physiological and/or disease conditions. At present there are different approaches for RNA-seq data analysis, leaving researchers with the burden of design and implementation of their computational pipeline. Even already existing pipelines might not fulfill the requirements for the specific experiment. This was the case for our study of small RNAs, in which the existing pipelines, *miR&moRe* and *miRDeep2*, were implemented only for human small RNAs in one case, or did not analyzed isomiRNAs and moRNAs in the other. For this reason, we adapted the pipeline software and improved it providing new features. The methods developed were applied to RNA-seq experiments on Italian Large White backfat, resulting in new findings about swine adipose tissue that are presented in published research articles (also reported here) and in the last section of this chapter (manuscript in preparation).

3.1 METHODOLOGICAL RESULTS

The method development activity regarded the implementation of an automated and computationally efficient pipeline for the analysis of long transcript RNA-seq data; plus, the improvement of the *miR&moRe* pipeline for the analysis of small transcript RNA-seq data. One further pipeline for the detection of circRNAs was developed and is described here. Details follow.

3.1.1 AN AUTOMATED ANALYSIS PIPELINE FOR LONG TRANSCRIPT RNA-SEQ DATA

To characterize the transcriptome from RNA-seq data in terms of sequence and expression level variations, and to compare gene expression from different sample, we composed a custom computational pipeline that can automatically execute the various steps for all the sample data provided.

As introduced above, different software and tools for the analysis of long RNA RNA-seq data has to be used to compose an analysis pipeline. Our choice of implementation borrows the idea of Bortoluzzi et al. (Bortoluzzi et al., 2012), in which the pipeline steps are composed using Scons (http://www.scons.org/) like for the compilation of software source code. Scons is a software tool written in Python designed to facilitate software development by managing the building and compilation of (large) software projects. Scons' actions are generic and can be implemented in Python scripts that will be custom for the specific application. Scons has been used as the pipeline backbone that links various analysis tools and triggers the computations in the right order. With this choice, we achieved repeatability of the analyses and improved computational performance. In fact, Scons allows automatic and parallel execution of different tasks (for instance, when a set of analysis steps has to be performed sequentially for each sample), and in a parsimonious way in the sense that a task is not rebuilt if not necessary (i.e. when input parameters has not changed between repeated runs). In addition, Scons computes a dependency tree of the various tasks and can executes independent tasks in parallel. These features make Scons scalable even to large projects.

The two pipelines have similar conceptual steps (raw data preprocessing, read mapping to the reference genome, characterization of RNA sequences, expression quantification). Yet, they are different in the tools utilized and the in the data features that are examined. The methods chosen to implement the long RNA pipeline steps follow.

The read preprocessing step is performed by means of DynamicTrim, which is an additional program in the SolexaQA package. It trims each read to its longest contiguous read segment (from either or both ends) where quality scores exceed a user-defined threshold, and writes this information to a standard FASTQ file. In combination with a length filter, custom written, it was used to remove poor quality bases and/or reads from high throughput sequence data. We chose to filter reads according to untrimmed read length and untrimmed read average quality. Reads having the mate discarded were discarded, too. A step reporting read quality for each preprocessing step was introduced using the FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for the raw reads, while the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) package has been used to report results of the trimming and length filter phases, as well as custom scripts either in Python and R. Subsequent steps of the pipeline resemble the protocol developed by Trapnell et al.(Trapnell et al., 2012), which uses *TopHat2* as read aligner, *Cufflink2* as transcript sequence inference and expression estimation tool, and *Cuffdiff2* as gene/transcript differential expression method. Additional methods were integrated for the gene expression level estimation and gene differential expression assessment, namely the *htseq-count* tool from the *HTSeq* framework (Anders et al., 2015) and *DESeq2*, respectively.

The pipeline output can be used in further downstream analysis, like for instance novel transcript characterization by coding potential prediction and/or sequence comparative annotation.

3.1.2 IMPROVEMENTS TO THE MIR&MORE PIPELINE

The *miR&moRe* pipeline was originally implemented to handle specifically data from human cells. The critical point was the use of human reference genome and human miRNA annotations, including the human miRNA-hairpin reference sequences from the miRBase repository. We generalized the scripts' code allowing small RNA sequencing data analysis for all the species genomes for which a reference genome and miRNA annotation are available. In addition, custom genomes and annotation could be provided as long as they comply with the required formats (Figure 14). Basically, the reference genome sequence must be in FASTA file (single file or split in chromosomes), and miRNA annotation in GFF formats like those provided by miRBase. An additional feature added to the processing pipeline was the prediction of novel miRNA precursors and mature miRNAs. In fact, the *miR&moRe* pipeline discovery power is limited to the prediction of sister miRNAs from already annotated precursors. Other tools such as *miRDeep2* provide the possibility of inferring novel miRNA hairpins from the reference genome, giving the genomic coordinates and the putative mature miRNA positions in the predicted hairpins. The modification carried out to *miR&moRe* allowed the integration of miRNA prediction from *miRDeep2* software and also the characterization of miRNAs from novel precursors (Figure 14). These adjustments introduced powerful features to *miR&moRe*, improving its capability, breath of application, and increasing its discovery skills.

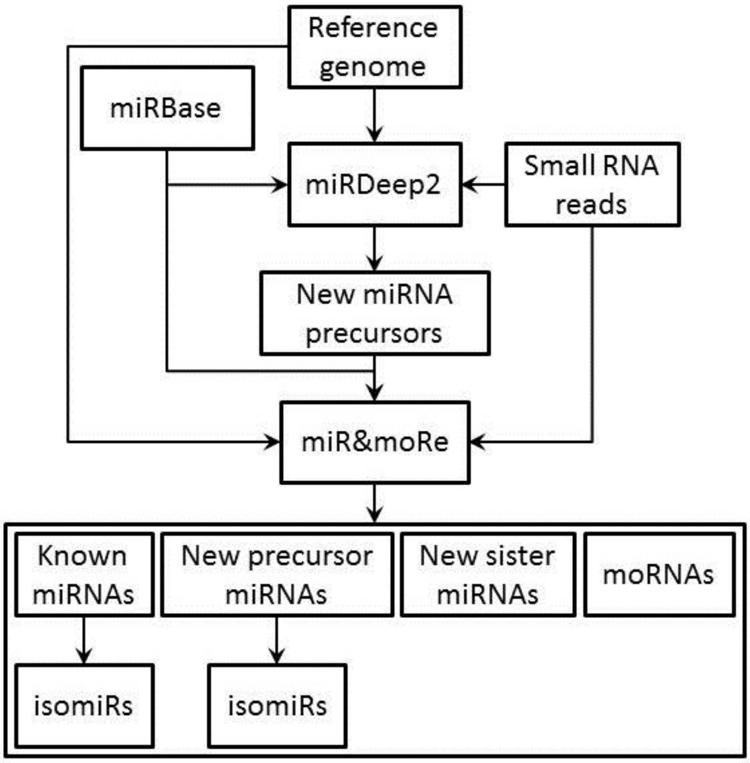


Figure 14. New miR&moRe pipeline workflow. The input reference genome can be from any species and miRNA annotation can be retrieved from miRBase; miRDeep2 is executed with the same reference genome, annotation and sequencing reads inputs given to miR&moRe. miRDeep2 predictions of new pre-miRNAs and the relative new miRNAs that are added to the miRBase annotation that is input to miR&moRe. In this way miR&moRe can characterize miRNAs from novel precursors and compute isomiRs also for novel pre-miRNA mature miRNAs.

3.1.3 A COMPUTATIONAL PIPELINE FOR THE DETECTION OF CIRCRNAS FROM RNA-SEQ DATA

As introduced before, circRNAs are circularized RNA molecules in which exon boundaries, several exons or intron and exon sequences are joined in a non-collinear way generating backsplice junctions that are sequences not included in the genome. CircRNA are not polyadenylated. Thus, poly(A) libraries do not capture circRNAs sequences and specific strategies for RNA-seq library preparation are needed, such as implementing depletion of ribosomal RNA without poly(A) enrichment. From the computational point of view, specific methods are needed for NRA-seq reads mapping to detect backsplices.

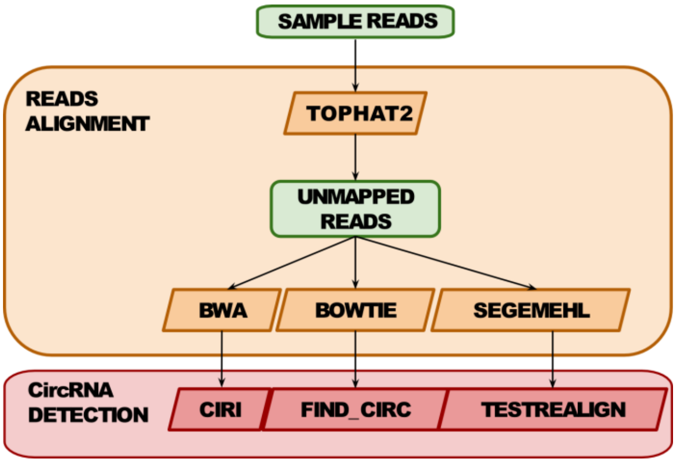


Figure 15. Pipeline for detection and quantification of circular RNAs.

In parallel with the ascertainment of experimental protocols for circRNA sequencing, algorithms for circRNA detection were set, such as *find_circ* (Memczak et al., 2013), *seghemel/testrealign* (Hoffmann et al., 2014), and *CIRI* (Gao et al., 2015). These methods essentially detect of back-splice junctions from mapping data; yet differ for the alignment strategy and format that they can process, providing heterogeneous predictions. Aiming at comparing different circRNAs detection method results, we applied the concepts of automation, modularity, and parallelization that have been employed in the development of the pipeline for the characterization of long RNAs (see chapter "An automated analysis pipeline for long transcript RNA-seq data") to set up a computational pipeline (Figure 15). As before, we used Scons to link and execute the various steps, which are described below.

To remove reads belonging to linear transcripts from subsequent processing for backsplice detection, a preliminary alignment to the reference genome is performed with TopHat (Kim et al., 2013) setting its parameters in a way that backsplice reads are not mapped. The strategy of a preliminary filter on the linear transcript has also been adopted by another circRNAs detection tool, CIRCExplorer (Zhang et al., 2014). Unmapped reads are then used as input for each circRNAs detection tools, *find_circ*, *testrealign*, and *CIRI*, that in turn involve a read alignment phase using, respectively, Bowtie (Langmead and Salzberg, 2012), Segemehl (Hoffmann et al., 2014), and BWA (Li and Durbin, 2009) aligners. In addition to the discovery of putative circRNAs, the pipeline achieves circRNAs expression quantification in terms of backsplice reads. This measure will be the starting point for the comparison of circRNAs expression with linear transcripts expression, in order to evaluate circular to linear expression proportion; and unsupervised analysis based on circRNAs expression data and circ/linear proportion estimation.

3.2 APPLICATIVE RESULTS

The implemented methods were applied to characterize the transcriptome and miRNome of pig adipose tissue and to compare pig backfat transcriptional profiles of different animal with extreme phenotypes of backfat thickness. Further, the transcriptome and miRNome profiles integration allowed the identification of post transcriptional regulatory interactions between miRNA and transcripts expressed in the tissue.

Next sections focus on the results of the application of the methods (see section "Materials and methods"), with details of implementation (software versions and additional tools) for each specific analysis. The first section ("Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness") regards the pig backfat transcriptome profiling and is extracted from the published article Zambonelli et al. (in press). Next two sections regard the characterization of pig backfat miRNome. The former ("miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs") is extracted from the published article (Gaffo et al., 2014). The latter section, "Differentially expressed small RNAs in Italian Large White pig adipose tissue", present the results of the comparison of miRNome profiles of two groups of ILW pigs yielding extreme and divergent backfat thickness phenotypes, with additional focus on the reconstruction of the putative small RNA-transcript interactions occurring in the tissue. Some supplementary tables are not reported here because of limited space and we remand to the relative article electronic supplementary material.

3.2.1 TRANSCRIPTIONAL PROFILING OF SUBCUTANEOUS ADIPOSE TISSUE IN ITALIAN LARGE WHITE PIGS DIVERGENT FOR BACKFAT THICKNESS

The objective of this research was to investigate the transcription profile of Italian Large White (ILW) pig backfat tissue and to compare the transcriptome of animals reared in the same herd and farming conditions showing high (FAT) vs. low (LEAN) backfat thickness (BFT). Moreover, a first functional characterization of DEGs has been obtained to provide new insights on genes, pathways and processes influencing the divergent aptitude of subcutaneous adipose tissue deposition in ILW pigs.

3.2.1 MATERIALS AND METHODS

RNA-seq data pre-processing and mapping to the swine genome

Exploratory analyses on the raw reads quality were carried out using FASTQC v0.10.1 software http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which generates an HTML report for each sample read set. Read fragments with a quality Phred score lower than 30 were trimmed using the DynamicTrim script of SOLEXAQA v2.1 (Cox et al., 2010). The FASTX-TOOLKIT v0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/) was used for trimming the result report. A custom Python script using the HTSEQ package (Anders et al., 2015) filtered out the trimmed reads shorter than 50 nt. To maintain a consistent paired-end read set, discarded read mates were also filtered out, despite their length and quality. Each sample paired-end clean read set was mapped to the swine genome (Sscrofa10.2.70) by TOPHAT v2.0.8 (Kim et al., 2013) using default parameters with transcriptome inference from the Ensembl annotation (TOPHAT2 used BOWTIE v2.1.0.0 (Langmead and Salzberg, 2012)) and SAMTOOLS v0.1.19 (Li et al., 2009)).

Gene/transcript expression evaluation and transcript reconstruction

Gene annotation for the reference genome was retrieved from Ensembl (BioMart) using the BIOMART R package (Durinck et al., 2009)(Durinck et al. 2009). Read alignments were processed by CUFFLINKS v2.1.1 (Roberts et al., 2011a, 2011b; Trapnell et al., 2010) to identify and discover expressed genes and transcripts, and to quantify their expression. Expression data were indicated as fragments per kilobase of transcript per million mapped reads (FPKM). CUFFLINKS was applied to each sample alignment; then, we merged the transcript predictions in a non-redundant reference using the CUFFMERGE tool from the CUFFLINKS package. To reduce artefacts deriving from the transcript prediction and normalisation strategies, only predicted transcripts at least 200 nt long and with minimal expression of 100 (CUFFLINKS normalised) reads in at least one of the two groups were considered for transcriptome reconstruction and for the following analyses.

Gene and transcript differential expression assessment The samples were inspected by principal components analysis to examine their similarities. The read counts of each gene in the 20 considered samples were transformed with the variance stabilising transformation function provided by the DESEQ2 package (Anders and Huber, 2010) and used to compute the principal components.

The genes identified by CUFFLINKS were assessed for differential expression (DE) between the LEAN and FAT groups by means of two strategies, namely CUFFDIFF2 (v2.1.1 from the CUFFLINKS package; Trapnell et al., 2012) and DESEQ2 v1.2.1 (Anders and Huber, 2010). Transcript DE was assayed only with CUFFDIFF2. To represent gene expression, the two methods use similar statistical approaches based on a generalised linear model (GLM) of the negative binomial family. CUFFDIFF2 extends the model using a beta negative binomial distribution to handle uncertainty of multimapped reads. On the contrary, DESEQ2 considers only uniquely mapped reads (counted by means of the htseq-count script of the HTSEQ package; (Anders et al., 2015)) but facilitates the specification in the statistical model of additional factors affecting the fit of the GLM. In this study, the statistical model included sex effect as a potential conditioning factor. Gene and transcript DE test-computed P-values were corrected according to the Benjamini–Hochberg procedure. DEGs and transcripts were considered statistically significant according to a false discovery rate 0.05.

Transcript characterisation

Using custom scripts, including BEDTOOLS v2.17.0 software (Quinlan and Hall, 2010), we retrieved the nucleotide sequences of the transcripts extracting from the *Sus scrofa* genome, the stretches of nucleotides according to the annotation generated by the RNA-seq analysis tools. Transcripts were identified or characterised by sequence similarity using BLASTN and BLAST2 from the NCBI BLASTN suite (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) using the Megablast algorithm (Morgulis et al., 2008). To assign a gene name, the sequences' IDs obtained with this comparison were used to query the NCBI Gene and the UniGene databases (http://www.ncbi.nlm.nih.gov/unigene/). We used two strategies for transcript annotation. DE transcripts (DETs) and DEGs were annotated by similarity using nr/nt nucleotide collection. The threshold considered for the identification of our transcripts was identity ~80% in at least 70% of the sequence length of a transcript present in the database. Transcripts from new genes were characterised using a comparative genomics approach. We compared the new transcripts from intergenic regions with known human transcripts (RefSeq Release 72) by aligning with

BLASTN (NCBI BLAST 2.2.29+). For each transcript, the best hit was considered, and then, alignments with E-value greater than 10E-6, identity <60% and length <100 nt were discarded.

Prediction of coding/non-coding potential

The transcript coding potential was predicted by CPC (Coding Potential Calculator;(Kong et al., 2007)). CPC is a support vector machine-based classifier of transcript protein-coding potential grounded on six features of sequence. Three features assess the extent and quality of the predicted transcript ORF (open reading frame): FRAMEFINDER software identifies the longest ORF in the three forward and the three reverse frames, then the coverage and the integrity of the predicted ORF are evaluated. Another three features derive from results of the BLASTX search against UniProt Reference Clusters. All the features together contribute to a final score and to the classification of transcripts as coding or non-coding. Only transcripts not including uncalled bases were considered for CPC analysis.

Validation by quantitative real-time PCR

The validation of selected transcripts was performed using a quantitative real-time PCR (qPCR) approach using 18 of the 20 samples used for the RNA-seq analysis. Two samples, one in the FAT group and one in the LEAN group, were not considered because the total RNA extracted was used completely for the RNA-seq analysis. qPCR validation was carried out using a Rotor-Gene TM 6000 (Qiagen,Corbett Research). After DNase treatment (TURBO DNA-freeTM, Ambion, Applied Biosystems), 1 lg of total RNA was reverse transcribed using the iScript cDNA Synthesis kit (Bio-Rad), according to the manufacturers' instructions.

The samples were first used to analyse four candidate normalising genes: beta-2-microglobulin (B2M); polymerase (RNA) II (DNA directed) polypeptide A, 220 kDa (POLR2A); hypoxanthine phosphoribosyltransferase 1 (HPRT1); and tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta (YWHAZ). The primer pairs and the PCR conditions used are reported in Table S1. The expression levels of these four genes were evaluated using NORMFINDER, and B2M and HPRT1, the two most stably expressed normalising genes, were utilised as reference genes. For each gene selected for validation, we designed an external primer pair to obtain the amplicon for the standard curve construction and an internal primer pair for the qPCR on the Rotor-Gene 6000 (Table S1). Standard curves for each gene were generated from 10 to 12 serial dilutions (from 109 to 25 molecules/ll) of the PCR amplicons obtained with the external primer pairs and containing the internal primers used in the qPCR analysis. Amplifications were performed in a total volume of 10 ll containing 5 ll of the SYBR Premix Ex TaqTM (Takara Bio Inc.), 0.5 ll of each primer and about 100 ng of cDNA. The Premix Ex TaqTM was optimised for a two-step cycling, and the amplification conditions for the tested genes are reported in Table S1. The PCR efficiency was calculated as E = 10 exp(-1/slope), with a range between -2.7 and -4.3, indicating a good PCR efficiency result. All the PCR products were checked on a polyacrylamide gel, and the specificity of the amplification was checked by a final melting curve analysis.

Threshold cycles obtained for the samples were converted by Rotor-Gene 6000 to mRNA molecules/ll using the relative standard curve for each gene (Bustin and Nolan, 2004). Moreover, the average mRNA molecules/ll for each sample was normalised by dividing the mRNA molecules of a gene/ll by the geometric average of B2M and HPRT1 mRNA molecules/ll in the given sample, as suggested by Bustin & Nolan (2004) and Vandesompele et al. (2002). Differences in the expression level calculated for FAT and LEAN samples were tested by a two-tailed Student's t-test. Statistical analyses were performed with SAS version 9.3 (SAS Institute, Inc.), and a nominal P-value 0.05 was considered a significance threshold.

Functional characterisation

Functional annotation, classification and clustering of selected gene sets were carried out by DAVID TOOLS 6.7 (Huang et al., 2008) using Biological Processes and Molecular Function Gene Ontology categories and KEGG pathways. A threshold for significance of P < 0.01 and P < 0.05 after Benjamini correction was considered for the selection of the functional categories, respectively, in the characterisation of the most expressed transcripts and for the selection of the functional categories of DEGs.

3.2.1 RESULTS

Sequencing, reads pre-processing and mapping

Pairs of samples were run together, after barcoding, on a single lane of an Illumina HiSeq 2000 apparatus, obtaining a total of 3 917 123 414 raw reads for the 20 considered samples, with an average of 195 856 171 raw reads per sample (Table S2; GEO accession GSE68007). After trimming and length filtering, the number of clean reads per sample was on average 113 934 264 (58.04%) and was used for read-to-genome mapping (Figure 16a). On average, 91.07% of the mapped reads aligned on a single genome locus (uniquely mapped reads) (Table S2). On average, 72.42% of the uniquely mapped reads (n = 72219306.45) aligned to annotated exons, 19.15% mapped to intergenic regions and

8.43% mapped to introns of annotated genes. The deep sequencing allowed for the identification of genes expressed at low levels and relatively rare alternatively spliced transcripts. We observed splicing events in 21.19% of the reads on average, providing useful information for the reconstruction of alternative transcript isoforms (Figure 16b).

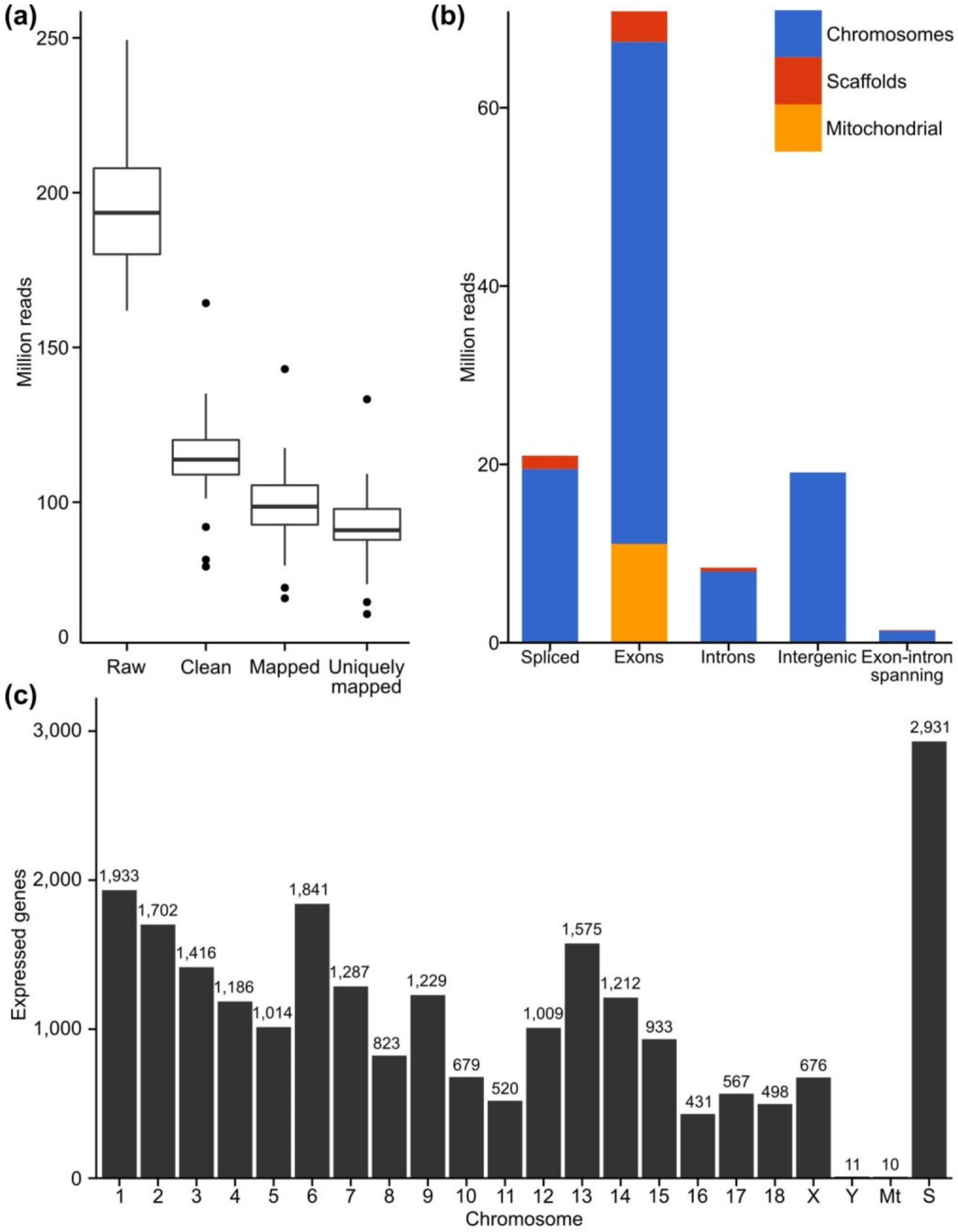


Figure 16. Read processing and alignment results. (A) The boxplots show the distribution of the reads considered in different steps and filters of the computational analysis pipeline, in the 20 considered samples. From left to right we show the number of raw reads sequenced, of clean reads resulted from the filtering steps, of reads successfully mapped to the reference genome, and of reads with unique alignment in the genome. (B) From the left, the bars show the average amounts, in the 20 considered samples, of reads spliced, aligned to an exon, to an intron, to intergenic regions (according to the *Sus scrofa* 10.2 genome annotation), or spanning exon-intron borders. Different colors indicate the proportion of read aligning to chromosomes (blue), genome scaffolds (red) or mitochondrial genome (yellow). (C) Number of expressed genes detected in different chromosomes, in mitochondrial genome (Mt) or in genome scaffolds (S).

Transcripts and genes expressed in backfat samples

The deep sequencing analysis of backfat transcripts performed on the two groups of pigs divergent for fat deposition in this tissue allowed the detection of 63 418 transcripts. Many of them have not yet been annotated in the porcine genome, thus providing new consistent resources for pig genome annotation and studies of adipose tissue biology. We identified the expression of genes on all porcine autosomes, sex chromosomes and mitochondrial genome. Chromosome 1 had the largest number of expressed genes (8.23%), followed by chromosomes 6 (7.84%) and 2 (7.25%). Furthermore, a non-negligible part (12.48%) of the expressed genes was located in genomic scaffolds (Figure 16c), as about 7.5% of the genome has no assigned location yet, as described in the Ensembl annotation of the pig genome (at the time of the analysis, database version 78; http://www.ensembl.org/Sus_scrofa/Location/Genome). In terms of genes, we identified 23 483 expressed pig genes: 12 707 known and 10 776 putative new genes.

Transcripts were split into different classes according to how they matched with the genome annotations (Figure 18a, Table S3). Transcripts exactly matching the reference annotation were indicated as 'known' transcripts; annotated transcripts' new isoforms or those overlapping with an annotated transcript were indicated as 'novel' isoforms; and all other transcripts, such as those expressed from extragenic regions, were referred to as 'new' transcripts and might represent putative new genes. The majority of expressed

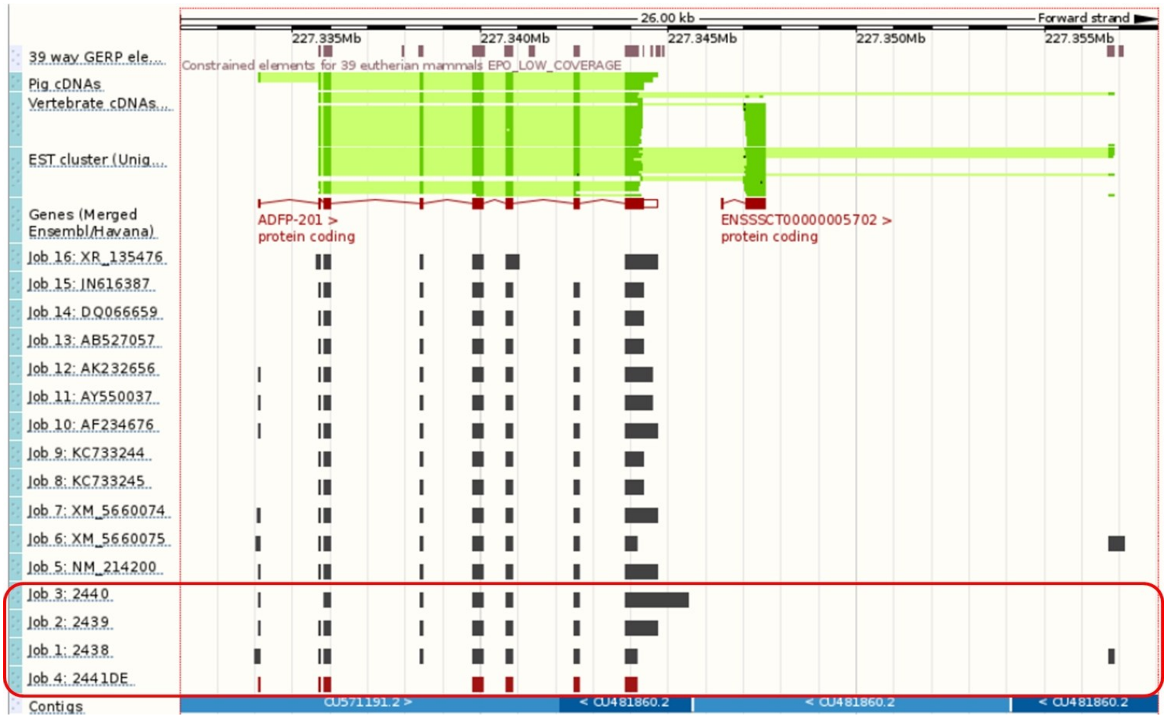


Figure 17. Alignment of the four detected isoforms of PLIN2 gene (red box) with the porcine and vertebrates transcripts present in Ensembl.

transcripts were novel isoforms (35 030; 55.2%) or known transcripts (12 969; 20.5%) that are prevalently annotated as protein coding (12 883; 99.3%); 15 419 (24.3%) were expressed new transcripts.

Transcript lengths ranged from 200 to 50 610 nt, with median and average values of 3224 and 3979 respectively. The average size exceeded the 2-kb mean pig transcript size that can be estimated according to Ensembl pig coding transcript annotations. We observed that the novel isoforms reconstructed were longer than 'known' pig transcripts (Figure 18b). Sequences longer than 5 kb comprised 25% of the expressed transcripts. Of note, we detected two transcripts overlapping the ZBTB16 gene and two new transcripts from chromosome 16 that were longer than 40 kb.

Considering transcript expression, we observed that new transcripts were less expressed in fat tissue than were known transcripts (Figure 18c). Nevertheless, all three transcript categories spanned a considerably large range of expression values.

The majority of the expressed genes (12 138; 52%) presented only one transcript isoform expressed in fat tissue (Figure 18d); 27.0% and 18.3% of the genes presented two and three expressed isoforms, respectively, whereas the remaining 12.7% of the genes were each associated with four to 31 different isoforms. We identified 31 isoforms for the gene MAP4K4, for which a complex expression pattern is reported in humans: Ensembl release 79 lists 20 MAP4K4 transcripts, generated by at least three different promoters by complex alternative splicing and by polyadenylation patterns, whereas five protein isoforms are reported in UniProt release 2015_3.

Regarding isoform types, as shown in Figure 18e many genes expressing only one transcript (first bar from the left) in fat tissue were putative new genes (green portion). Interestingly, some genes expressing only one transcript in fat tissue were represented only by a novel isoform (first bar, red shading). The proportion of novel isoforms (red portion) increased along with the number of expressed transcripts per gene. Moreover, the transcript classes showing exonic overlap compared to a reference transcript were found in genes with a varying number of transcripts and were particularly abundant in genes with up to three isoforms. The remaining transcript classes were very rare.

Interesting new isoforms derived from the known gene perilipin 2 (PLIN2; also known as ADFP, adipofilin), an important gene for fat metabolism in pigs (Davoli et al. 2010; Gandolfi et al. 2011) whose expression in humans correlates positively with cytosolic triacylglycerol levels (Conte et al. 2013). Only one transcript for pig PLIN2 is currently annotated in Ensembl (ENSSSCT00000005701), whereas according to our results, PLIN2 is expressed four different isoforms. The most expressed PLIN2 transcript (expressed two times more in FAT than in LEAN pigs) was a non-annotated isoform (TCONS_00002441 in Table 5; 2441DE in Figure 17) and was characterized by the skipping of the fourth exon. The same transcript also has a shorter 30 sequence with respect to the canonical PLIN2/ADFP form, probably due to the use of an alternative polyadenylation site. Importantly, the skipping of the 83-nt-long exon four introduces a downstream shift in the reading frame and a premature stop codon. Thus, this transcript encodes a truncated

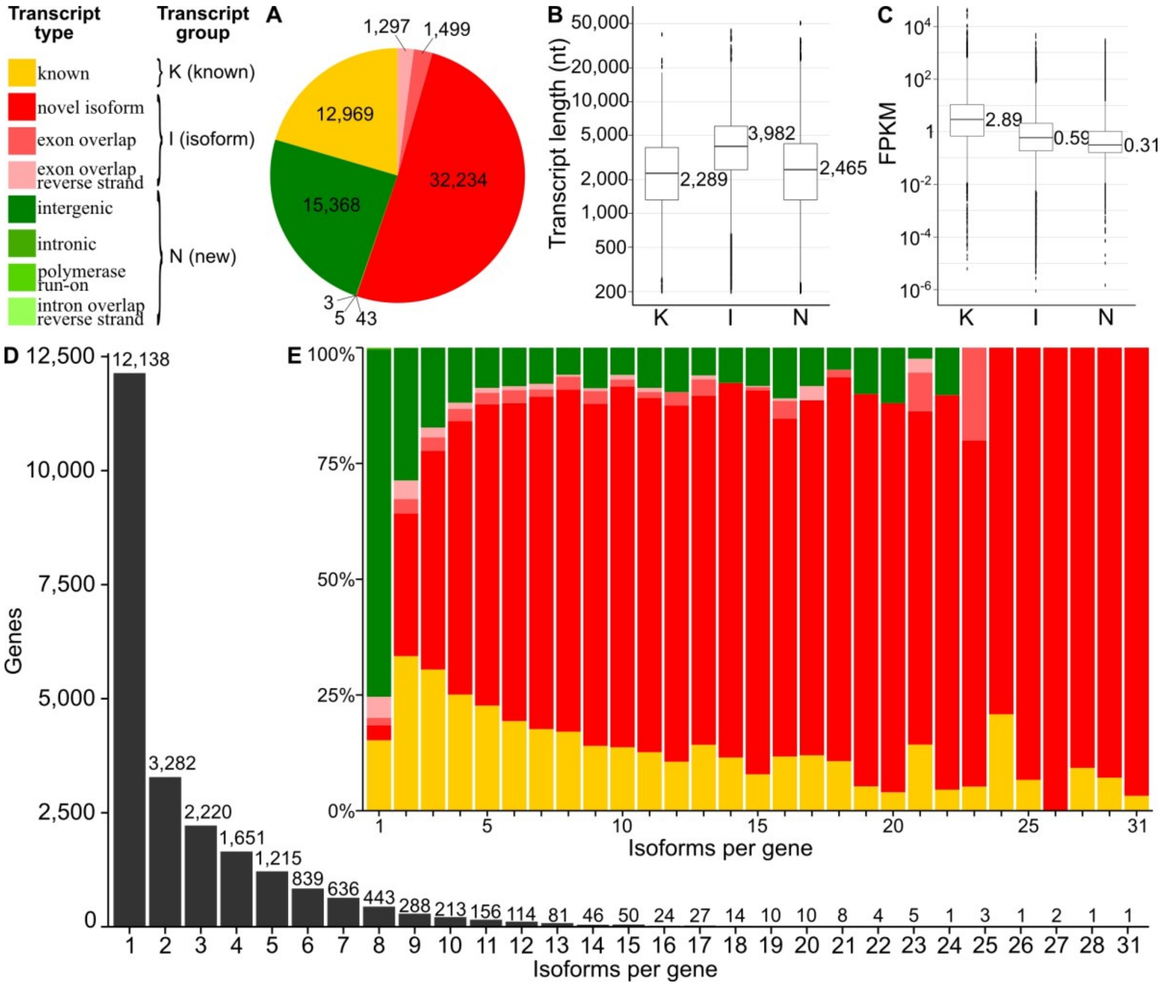


Figure 18. Transcripts and isoforms classification. (A) Expressed transcript were classified, according to current gene annotations, into 8 types, reported with different colors (see legend) and grouped in three categories: K (known) collects transcripts found in reference annotation (yellow); I (isoform) collects alternative forms of transcripts (red shades); N collects new transcripts from not-annotated loci (green shades). The pie chart shows the number of transcripts detected, for each type, and their mutual proportions. Three transcript types of the N group have few elements (43 intronic; 5 possible polymerase run-on fragments; 3 transcript intron overlap a reference intron on the opposite strand) and are barely visible in the chart. (B) Transcript length distributions in the three categories. (C) Transcript expression level distribution for the three categories. (D) Number of genes (vertical axis) with their number of transcript isoforms detected (horizontal axis). Genes with only one transcript isoforms detected are the most frequent; however, genes with up to 31 different isoforms were detected. (E) The proportion of each transcript type for the transcript isoforms grouped as in (D). Genes with only one isoform (first bar) are mainly intergenic genes (green part). For genes having more than one isoform expressed, the proportion of novel isoforms detected increases along with the number of different isoforms for a gene (red part).

protein (only 80 amino acids) corresponding to the N-terminal region and of the perilipin domain of the annotated PLIN2 protein isoform (463 amino acids). The other two new transcripts differ from the annotated isoform, one by the skipping of exon 2 and the other with a longer first exon, probably due to alternative transcription start site usage by different promoters. The four expressed isoforms are also heterogeneous in the length of their 30-UTR regions.

Coding and non-coding transcripts from new genes

We obtained a characterization of intergenic transcripts from new genes first, both by similarity, comparing them against human transcripts, and by predicting their coding potential. New pig transcripts with an assigned human best hit numbered 10 020 (65%), expressed by 7099 genes (66%) and corresponding to 4633 human Refseq sequences (3882 unique gene symbols; Table S4).

We considered 12 702 intergenic transcripts for protein-coding potential analysis. For each transcript, the coding potential of both the forward and the reverse complement sequences was evaluated. According to CPC results, we classified 35.8% (n = 4551) of transcripts as coding and 64.2% (n = 8151) as non-coding. Following Zhou et al. (2014), we considered proper non-coding only those transcripts classified as non-coding and having a CPC score lower than -1 for both the forward and the reverse sequences. A portion of the non-coding transcripts (37.5%) resulted with a CPC score <-1 for both the forward and the reverse complement sequences. We referred to these transcripts as a 'reliable non-coding' class, which represented 24% (n = 3,056) of the intergenic transcripts (Figure 19a). We observed that intergenic coding transcripts were on average longer than intergenic non-coding transcripts (4149 and 3083 nt respectively) and that the reliable non-coding fraction had an even shorter average length (2571 nt; Figure 19b and Table S5). Reportedly, non-coding transcripts tend to be shorter and to have fewer exons than do coding transcripts in mammalian genomes (Iyer et al. 2015).

Coding transcripts had an average expression in fat tissue higher than did the non-coding transcripts (5.32 vs. 2.28 FPKM, respectively, and 3.23 FPKM for the reliable non-coding group; Figure 19c). One reliable non-coding transcript was ranked within the 100 most expressed transcripts detected in backfat tissue; 15 reliable non-coding transcripts were within the 1000 most expressed transcripts; and 98 were within the 10% most expressed transcripts (Table S6). In agreement with previous results showing that coding transcripts tend to present higher expression than do non-coding ones (Cabili et al., 2011; Iyer et al., 2015), we observed that intergenic transcripts ranking in the 10% most expressed in backfat tissue were enriched in the coding category (55%) and particularly if compared

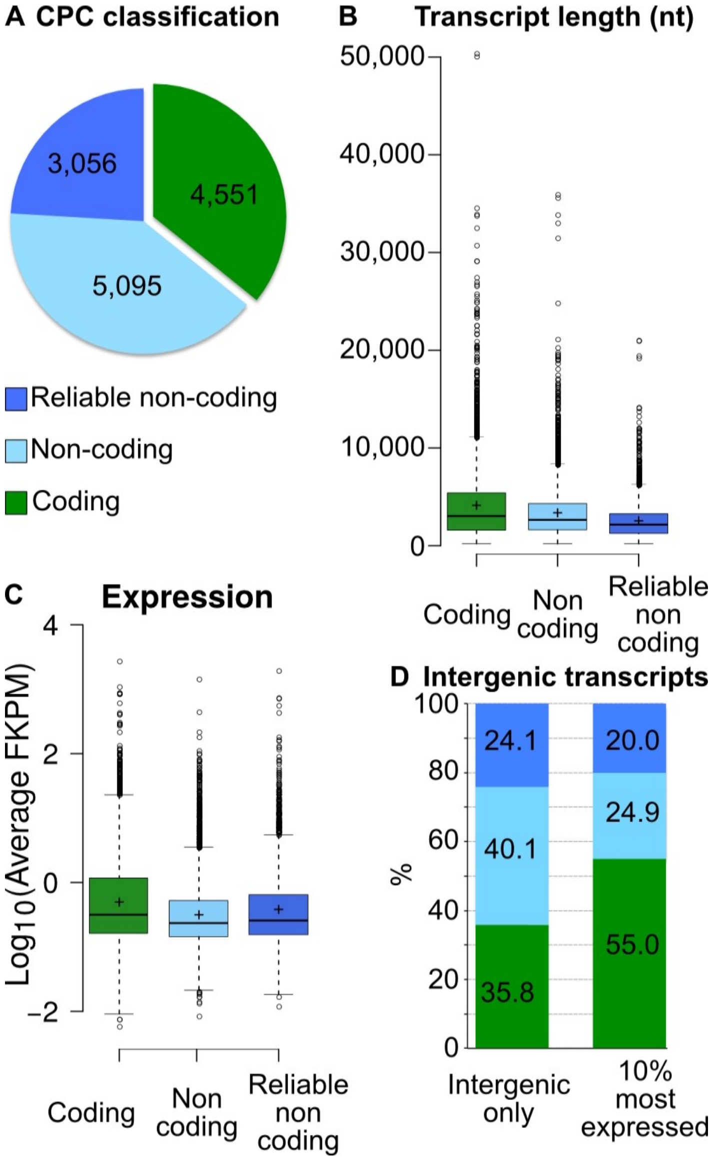


Figure 19. Coding potential of new intergenic transcripts. According to CPC scores, calculated both for the forward and for the reverse complement sequence, the intergenic transcripts were classified as "coding", "non-coding" and "reliable non-coding". (A) The pie chart shows numbers and proportions of intergenic transcripts falling in each category and provides the color code for the figure panels. (B) and (C) show respectively the distribution of lengths and of expression levels of intergenic transcripts, binned in the three categories. (D) Percentages of transcripts per category are compared, considering all the intergenic transcripts and the subset of the intergenic transcripts ranked within the 10% most expressed transcripts considering the whole transcriptome.

with the proportion of the coding category within the set of intergenic transcripts (35.8%; Figure 19d, green portions).

Function of most expressed transcripts

A global view of the transcription profile of porcine backfat tissue was obtained by averaging the FPKM values of all 20 analyzed samples. The 1411 most expressed transcripts, together accounting for 75% of expression, were chosen to extract the most expressed genes (Table S6). Among these genes, 59 were indicated as reliable non-coding (CPC score <1) and 66 showing a positive CPC score were indicated as putative coding.

According to DAVID functional annotation and clustering, we characterized the biological processes (Table S7) associated with the most expressed genes. Results showed ribosomal activity, oxidative phosphorylation, protein metabolic processes, intracellular protein transport, regulation of translation initiation, fatty acid metabolism, and response to oxidative stress to be the biological processes more represented in subcutaneous adipose tissue of the analyzed samples.

Gene/transcript differential expression

Unsupervised analysis of gene expression profiles was carried out to inspect similarities among the samples. Principal components analysis revealed a clear separation of the LEAN and FAT samples according to the first two most informative components (Figure 20a), which, notably, did not separate the samples by sex (Figure 20b).

Average gene expression values for FAT and LEAN groups were 32.46 and 33.63 FPKM. In both groups, few highly expressed genes contributed to the majority of the cumulative expression. For instance, roughly 25% of expressed genes (5908 and 5728 in FAT and LEAN respectively) constituted 95% of the total detected expression (Figure 22). As expected, transcript expression distribution was similar to the gene expression distribution, being positively skewed with a mean and median of 11.84 and 0.64 FPKM respectively. The transcripts' average expression values were lower than were the genes' expression values because the latter was computed as the sum of transcript expression of each gene.

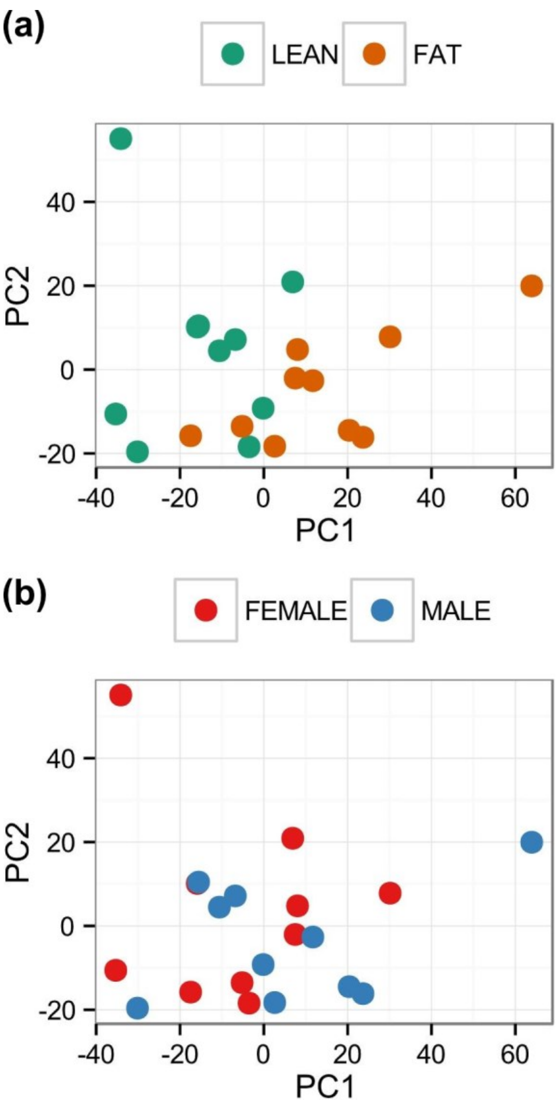


Figure 20. Principal components analysis (PCA) based on gene expression profiles.

To identify a set of robust DEGs and DETs, the transcription profiles of FAT and LEAN samples were compared with the integration of two methods applied at the gene and transcript levels. CUFFDIFF2 identified 414 DEGs between the FAT and LEAN groups, corresponding to 1187 transcripts: 266 DEGs were more highly expressed and 148 DEGs were expressed less in FAT samples. Fold changes in the base two logarithmic scale of DEGs ranged from 0.46 to 8.95 for the more highly expressed genes, and from -6.19 to -0.47 for the lower expressed ones (Table S8). DESEQ2 identified 586 DEGs (185 in common with the DEGs identified by CUFFDIFF2) corresponding to 1504 transcripts: 358 genes were up-regulated and 228 genes were less expressed in FAT samples. DEGs base two logarithmic scale transformed fold changes (Log_2 FC) ranged from -1.13 to -0.20 for the lower expressed

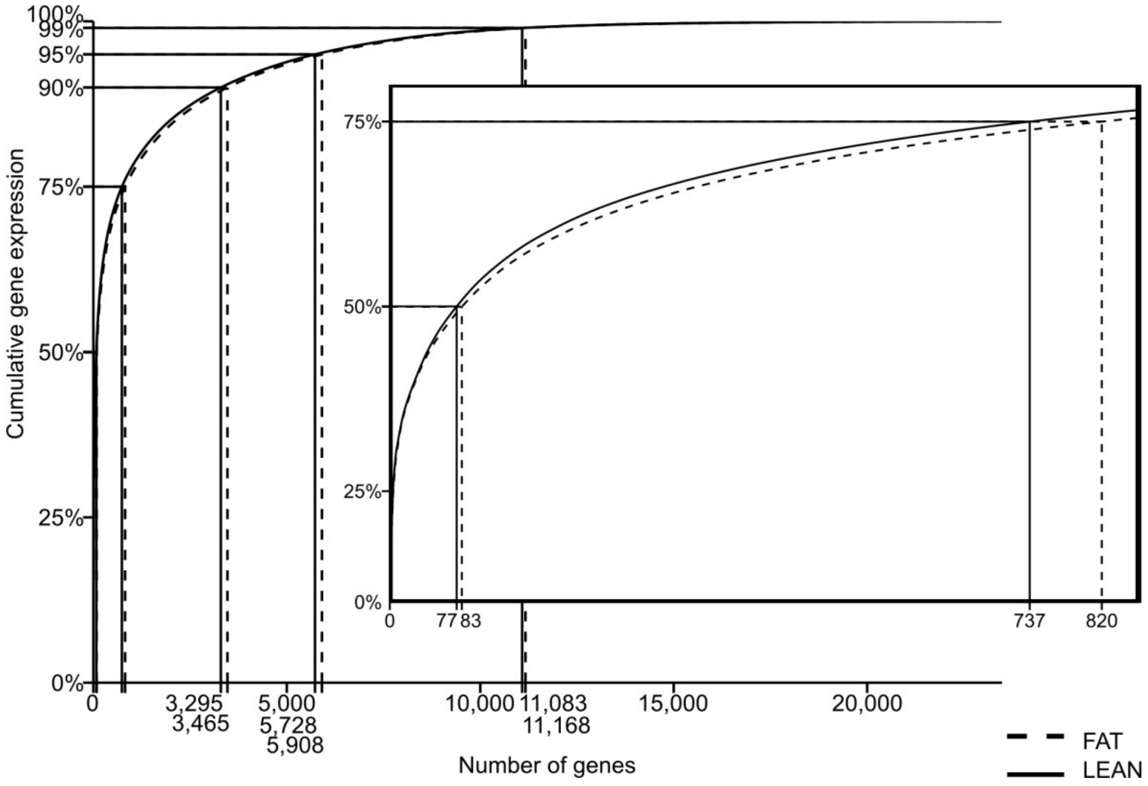
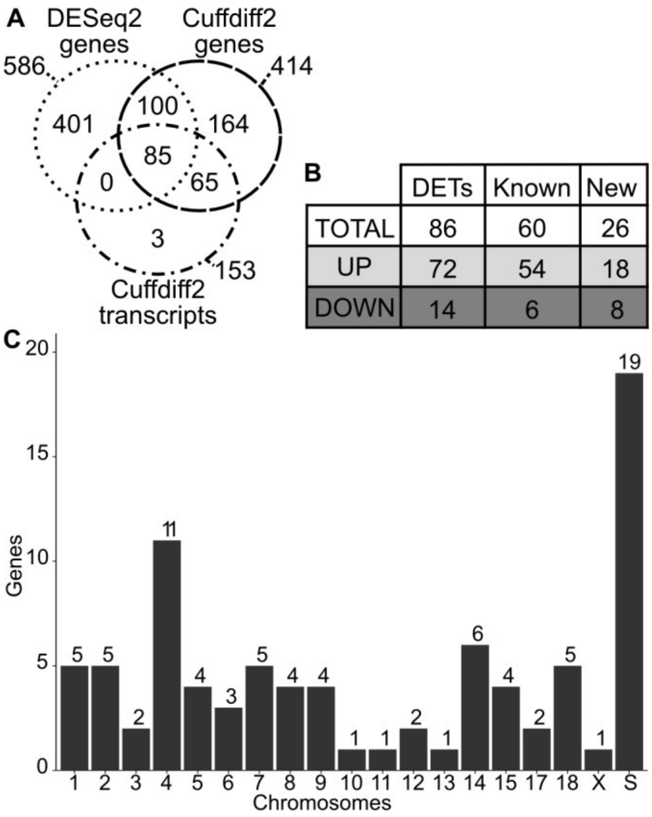


Figure 22. Cumulative gene expression in FAT and LEAN groups.

genes and from 0.21 to 1.18 for the more highly expressed genes (Table S9). CUFFDIFF2 differential expression analysis at the transcript-level identified 154 DETs (corresponding to 153 genes): 48 had a lower expression level and 106 transcripts were more highly expressed in FAT samples, with the Log$_2$ FC ranging from -3.44 to -0.54 and from 0.64 to 3.66, respectively (Table S10). On the whole, 818 genes were detected as being DE, or associated with at least one DET, according to at least one method (Figure 21a).

The overlapping of the different lists



	DETs	Known	New
TOTAL	86	60	26
UP	72	54	18
DOWN	14	6	8

Figure 21. Differentially expressed genes and transcripts identified. (a) Intersection of genes differentially expressed (DE) according to DESEQ2 and CUFFDIFF2 analysis, and genes with at least one transcript DE according to the transcript-level CUFFDIFF2 analysis. We focused on the transcripts belonging to the 85 genes identifiers commonly identified by all the methods, which corresponded to 78 gene symbols (b) Proportions of the new and known DE transcripts resulting in higher and lower expression in FAT vs. LEAN samples. (c) Number of DE genes mapping to chromosomes or to genome scaffolds (S).

of DEGs and DETs evidenced a group of 86 DETs that were identified by all the approaches, from now on referred to as ‘common DETs’ or cDETs. These DETs belonged to 78 DEGs, from now on referred to as ‘common DEGs’ or cDEGs, given that five genes were represented by more than one isoform (Table 5).

Table 5. List of the DE genes and transcripts.

Cufflinks transcript ID	Cufflinks gene ID	Gene locus	Gene symbol	Cuffdiff2 gene $\log_2$(FAT/ LEAN)	Transcript group	Coding potential
TCONS_00102010	XLOC_040987	JH118612.1:113132-140205	*DSC2*	2.55	Known	-
TCONS_00061823	XLOC_023331	4:78928264-78930654	-	2.46	New	NON CODING
TCONS_00033774	XLOC_013001	15:140797584-140847461	*NYAP2*	2.38	New	CODING
TCONS_00061359	XLOC_023211	4:35670339-35685878	*DCSTAMP*	2.23	Novel isoform	CODING
TCONS_00095554	XLOC_036823	GL893451.1:11131-27485	*CRLF2*	2.21	Known	-
TCONS_00093244	XLOC_035190	9:50996895-51001264	-	2.17	New	NON CODING
TCONS_00087029	XLOC_032796	8:140307937-140315415	*SPP1*	2.09	Known	-
TCONS_00003007	XLOC_000806	1:283547172-283552108	-	2.07	New	CODING
TCONS_00095549	XLOC_036822	GL893451.1:7060-10625	-	2.03	New	NON CODING
TCONS_00067029	XLOC_025404	5:36179189-36186325	*LYZ*	2.03	Known	-
TCONS_00042581	XLOC_016514	18:6731368-6733669	*GIMAP2*	1.98	Known	-
TCONS_00061600	XLOC_023265	4:55660234-55715444	*ATP6V0D2*	1.96	Novel isoform	CODING
TCONS_00039556	XLOC_015432	17:53815353-53827092	*MMP9*	1.92	Known	-
TCONS_00039900	XLOC_015518	17:4110395-4192029	*MSR1*	1.92	Known	-
TCONS_00061643	XLOC_023283	4:62172539-62226917	*STMN2*	1.85	Known	-
TCONS_00034645	XLOC_013236	15:62409564-62414328	-	1.84	New	RELIABLE NON CODING
TCONS_00091509	XLOC_034399	9:63158999-63198155	*ST14*	1.79	Novel isoform	CODING
TCONS_00098750	XLOC_038994	GL895411.1:0-1073	*INHBB*	1.65	New	CODING
TCONS_00022322	XLOC_008474	13:32323641-32330286	*CCR1*	1.63	Known	-
TCONS_00044383	XLOC_017319	2:11807281-11850646	*MPEG1*	1.63	Known	-
TCONS_00075056	XLOC_028007	6:70039585-70099223	*PADI2*	1.6	Known	-
TCONS_00095875	XLOC_037025	GL893645.1:0-307	-	1.57	New	RELIABLE NON CODING
TCONS_00084869	XLOC_032187	8:71288921-71302169	*AMBN*	1.56	Known	-
TCONS_00033691	XLOC_012975	15:133452328-133456736	*SLC11A1*	1.56	Known	-
TCONS_00089513	XLOC_033895	9:90266412-90348498	*SCIN*	1.55	Known	-
TCONS_00042660	XLOC_016535	18:8306789-8313120	-	1.52	New	CODING
TCONS_00059834	XLOC_022860	4:99905518-	*CD1A*	1.52	Novel	CODING

Cufflinks transcript ID	Cufflinks gene ID	Gene locus	Gene symbol	Cuffdiff2 gene $\log_2$(FAT/ LEAN)	Transcript group	Coding potential
		99915176			isoform	
TCONS_00059837	XLOC_022860	4:99905518-99915176	*CD1A*	1.52	Known	-
TCONS_00093519	XLOC_035465	9:101443296-101443885	*GPNMB*	1.46	New	NON CODING
TCONS_00098157	XLOC_038614	GL894967.1:126-517	*GPNMB*	1.42	New	CODING
TCONS_00018804	XLOC_007247	12:23439824-23441829	-	1.4	New	CODING
TCONS_00103084	XLOC_041497	X:37303173-37393818	*CYBB*	1.38	Known	-
TCONS_00065337	XLOC_024931	5:52504178-52625145	*BCAT1*	1.37	Novel isoform	CODING
TCONS_00098113	XLOC_038589	GL894923.1:47-563	*GPNMB*	1.36	New	CODING
TCONS_00002441	XLOC_000664	1:227333991-227356844	*PLIN2*	1.32	Novel isoform	CODING
TCONS_00044392	XLOC_017322	2:12191483-12243400	*LPXN*	1.31	Known	-
TCONS_00084565	XLOC_032101	8:33970571-33982450	*UCHL1*	1.27	Novel isoform	CODING
TCONS_00067389	XLOC_025495	5:64579162-64590512	*OLR1*	1.26	Known	-
TCONS_00059747	XLOC_022835	4:97720982-97736619	*CD48*	1.25	Known	-
TCONS_00028769	XLOC_011055	14:143745489-143752509	*GMFG*	1.23	Known	-
TCONS_00029056	XLOC_011139	14:8804077-8816800	*STC1*	1.23	Novel isoform	CODING
TCONS_00098643	XLOC_038938	GL895339.1:13269-61205	*COTL1*	1.15	Known	-
TCONS_00100592	XLOC_040068	GL896326.1:1999-3913	*ACP5*	1.13	Known	-
TCONS_00096837	XLOC_037668	GL894123.1:0-400	*CD163*	1.13	New	CODING
TCONS_00097297	XLOC_037990	GL894401.1:0-471	*CD163*	1.13	New	CODING
TCONS_00005002	XLOC_001331	1:125897935-125953413	*AQP9*	1.09	Known	-
TCONS_00096863	XLOC_037686	GL894145.1:0-401	*CD163*	1.09	New	CODING
TCONS_00071337	XLOC_027094	6:74616232-74621248	*C1QC*	1.08	Known	-
TCONS_00012469	XLOC_005058	11:21534980-21685851	*LCP1*	1.07	Novel isoform	CODING
TCONS_00079920	XLOC_030238	7:94900207-94906867	*AKAP5, LOC100153460*	1.06	Novel isoform	CODING
TCONS_00041537	XLOC_016257	18:6613761-6621027	*GIMAP4*	1.06	Known	-
TCONS_00097908	XLOC_038444	GL894747.1:3047-10617	*HMOX1*	1.06	Novel isoform	CODING
TCONS_00030401	XLOC_011444	14:71516962-71521335	*EGR2*	1.05	Known	-
TCONS_00030878	XLOC_011579	14:117265093-117349965	*BLNK*	1.04	Known	-
TCONS_00056578	XLOC_021190	3:77408776-77439119	*PLEK*	1.04	Known	-
TCONS_00071335	XLOC_027093	6:74609911-74612993	*C1QA*	1.02	Known	-
TCONS_00081915	XLOC_030757	7:54395230-54406136	*BCL2A1*	1.01	Known	-
TCONS_00041554	XLOC_016261	18:6872940-6875292	*GIMAP1*	1	Known	-
TCONS_00085005	XLOC_032236	8:79743274-79751980	*SFRP2*	0.99	Known	-
TCONS_00098919	XLOC_039115	GL895590.1:0-1327	*GPNMB*	0.91	New	NON

Cufflinks transcript ID	Cufflinks gene ID	Gene locus	Gene symbol	Cuffdiff2 gene $\log_2$(FAT/ LEAN)	Transcript group	Coding potential
						CODING
TCONS_00068526	XLOC_026077	5:52625315-52630242	*BCAT1*	0.89	New	RELIABLE NON CODING
TCONS_00062055	XLOC_023401	4:97099149-97103132	*FCER1G*	0.87	Known	-
TCONS_00009719	XLOC_003695	10:48841010-48961015	*MRC1*	0.86	Novel isoform	CODING
TCONS_00030894	XLOC_011584	14:117670639-117938624	*PIK3AP1*	0.85	Known	-
TCONS_00017526	XLOC_006800	12:36561025-36604089	*CLTC*	0.8	Novel isoform	CODING
TCONS_00062959	XLOC_023614	4:119674090-119703427	*CD53*	0.78	Known	-
TCONS_00081898	XLOC_030753	7:53623061-53644262	*CTSH*	0.78	Known	-
TCONS_00060570	XLOC_023035	4:119013307-119039899	*ADORA3*	0.74	Known	-
TCONS_00052401	XLOC_020144	3:11035819-11055510	*LAT2*	0.71	Known	-
TCONS_00004118	XLOC_001095	1:35133812-35137388	*CTGF*	0.68	Known	-
TCONS_00045043	XLOC_017499	2:59214054-59218018	*IFI30*	0.65	Known	-
TCONS_00004124	XLOC_001096	1:35240242-35281384	*ENPP1*	0.62	Known	-
TCONS_00062884	XLOC_023592	4:116704501-116707235	*OLFML3*	-0.54	Known	-
TCONS_00035484	XLOC_013426	15:131680309-131684630	*IGFBP5*	-0.65	Known	-
TCONS_00101718	XLOC_040809	JH118426.1:306724-312138	-	-0.77	New	RELIABLE NON CODING
TCONS_00063805	XLOC_024145	4:77261119-77264781	-	-0.77	New	NON CODING
TCONS_00050164	XLOC_018733	2:124815021-124828122	*CDO1*	-0.9	Novel isoform	CODING
TCONS_00101559	XLOC_040715	GL896532.1:212-2567	*ADSSL1*	-1.02	New	NON CODING
TCONS_00079927	XLOC_030240	7:94987617-94990126	*HSPA2*	-1.1	Known	-
TCONS_00083805	XLOC_031620	7:66542203-66555641	-	-1.18	New	CODING
TCONS_00041725	XLOC_016313	18:15292592-15295178	-	-1.61	New	CODING
TCONS_00048853	XLOC_018425	2:65175406-65180520	*DNAJB1*	-1.66	Novel isoform	CODING
TCONS_00029533	XLOC_011248	14:35688332-35701411	*HSPB8*	-1.81	Known	-
TCONS_00094194	XLOC_036009	GL892492.1:0-3540	*HSPA1B*	-2.32	New	NON CODING
TCONS_00101505	XLOC_040677	GL896522.1:9039-10877	*HSPA1A*	-2.57	New	RELIABLE NON CODING
TCONS_00098059	XLOC_038555	GL894890.1:5-696	*HSP70*	-3.44	New	NON CODING

The cDETs present the same fold change sign of the corresponding cDEG (Figure 21b): 72 DETs were more highly expressed in FAT (max CUFFDIFF2 gene-level Log2 FC = 2.55 for the DSC2 gene) and 14 DETs had lower expression levels in FAT (minimum Log2 FC = -3.44 for an intergenic gene located in the GL894890.1 scaffold). Among the 86 cDETs, 44 were known transcripts, 16 were novel isoforms and 26 came from intergenic regions.

The cDEGs were found on all chromosomes except for chromosomes 16 and Y, with up to 11 DEGs on chromosome 4 and 19 DEGs in scaffolds (Figure 21c). The most expressed (average FPKM >100) known cDEGs, reported in decreasing FPKM order, were DNAJB1, CTSH, CTGF, C1QC, SPP1 and CDO1.

Coding and non-coding intergenic DET

We considered the 41 novel isoforms or new transcript cDETs for CPC analysis. In 14 of these transcripts, both the forward and reverse sequences were probably non-coding, according to integrated ORF analyses, similarity searches and CPC score thresholds used before. Five cDETs with CPC scores < 1 were scored as reliable non-coding. Of the remaining transcripts, nine presented low coding potential both in the forward and reverse complement sequences but with CPC scores ranging from -1 to 0 (non-coding), and 27 were classified as coding transcripts (Table S11).

qPCR confirmation of DE for selected genes

To validate the results obtained by RNA-seq, 11 cDEGs were chosen according to the absolute value of the Log2 FC between FAT and LEAN pigs or for their functional role and involvement in relevant pathways. As reported in Figure 23, the DE of all selected genes was validated, with high correlation between the fold changes obtained by RNA-seq and by qPCR data.

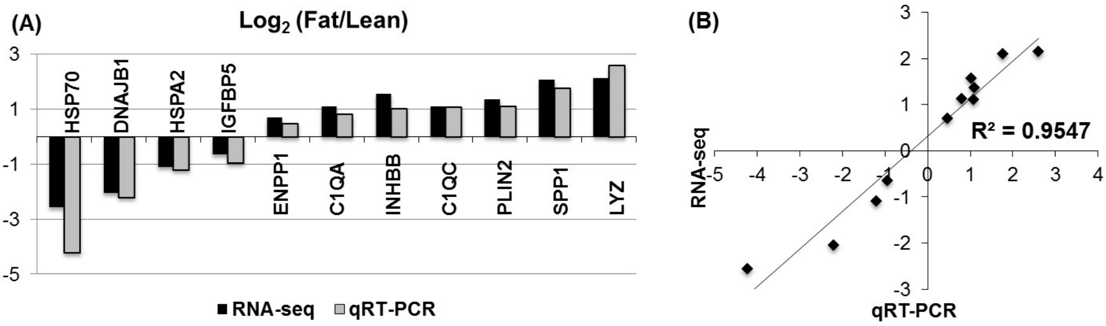


Figure 23. qPCR validation of 11 genes differentially expressed according to RNA-seq data. (a) Log2 FC values obtained from RNA-seq, according to CUFFDIFF2 estimates (black bars), and from qPCR data (grey bars), for the 11 tested genes; (b) scatterplot showing the good correlation between the Log2 FC values calculated with the two experimental methods.

Differentially expressed transcript characterization

We characterized the cDEGs in terms of their functional role in adipose tissue. Using DAVID Bioinformatics Resources, we first identified the functional categories enriched in genes differentially regulated between FAT and LEAN groups.

The Biological Process categories enriched in more highly expressed DEGs were response to stimulus, immune system process and cell activation, and skeletal system development (Table 6). DAVID clustering of the few lower expressed genes detected (ADSSL1, CDO1, DNAJB1, HSPA1A, HSPA1B, HSPA2, HSPB8, IGFBP5, OLFML3) allowed for the identification of the functional categories unfolded protein binding and stress response represented by five heat shock protein genes that are involved in protein stabilization after cellular stress. Apart from the Gene Ontology-based functional characterization of the whole subsets of more highly and lower expressed genes, we considered cDEG function and involvement in specific pathways, according to literature and knowledge bases. Several more highly expressed genes in FAT animals (ACP5, BCL2A1, CD1A, EGR2, ENPP1, GPNMB, INHBB, LYZ, MSR1, OLR1, PIK3AP1, PLIN2, SPP1, STC1) were characterized by a metabolic function related mainly to adipocyte growth regulation, whereas others (CCR1, CD163, SLC11A1) are known to be involved in immune defense of the organism.

Table 6. Table 3 - David functional annotation clustering obtained considering the significant Biological Processes GO terms (Benjamini adjusted P-values <0.05) of genes more expressed in FAT than in LEAN animals.

Annotation Cluster 1		**Enrichment Score: 7.0**
Term	Count	Genes
GO:0006954~inflammatory response	12	*C1QA, SLC11A1, CYBB, ADORA3, OLR1, HMOX1, CCR1, LYZ, C1QC, BLNK, CD163, SPP1*
GO:0006952~defense response	15	*ADORA3, OLR1, CCR1, LYZ, COTL1, C1QC, CD163, INHBB, CD48, C1QA, SLC11A1, CYBB, HMOX1, SPP1, BLNK*
GO:0009611~response to wounding	14	*ADORA3, PLEK, OLR1, CCR1, LYZ, C1QC, CD163, C1QA, SLC11A1, CYBB, CTGF, HMOX1, SPP1, BLNK*
GO:0009605~response to external stimulus	17	*ADORA3, PLEK, OLR1, CCR1, LYZ, C1QC, CD163, INHBB, C1QA, SLC11A1, CYBB, CTGF, SFRP2, HMOX1, STC1, SPP1, BLNK*
GO:0050896~response to stimulus	29	*ADORA3, AQP9, ENPP1, CCR1, UCHL1, ACP5, C1QC, CD48, SLC11A1, PLIN2, CTGF, HMOX1, FCER1G, BLNK, SPP1, EGR2, OLR1, PLEK, LYZ, CD1A, COTL1, CD163, INHBB, C1QA, CYBB, LAT2, SFRP2, STC1, LCP1*
GO:0006950~response to stress	19	*ADORA3, AQP9, PLEK, OLR1, CCR1, UCHL1, LYZ, COTL1, C1QC, CD163, INHBB, CD48, C1QA, SLC11A1, CYBB, CTGF, HMOX1, SPP1, BLNK*
Annotation Cluster 2		Enrichment Score: 2.7
Term	Count	Genes
GO:0001775~cell activation	7	*CD48, SLC11A1, LAT2, PLEK, LCP1, BLNK, GIMAP1*
GO:0002274~myeloid leukocyte activation	4	*CD48, SLC11A1, LAT2, GIMAP1*
GO:0046649~lymphocyte activation	6	*CD48, SLC11A1, LAT2, LCP1, BLNK, GIMAP1*

Annotation Cluster 1	**Enrichment Score: 7.0**	
Annotation Cluster 3	Enrichment Score: 2.4	
Term	Count	Genes
GO:0048583~regulation of response to stimulus	10	*C1QA, SLC11A1, LAT2, PLEK, ENPP1, HMOX1, FCER1G, C1QC, SPP1, GIMAP1*
GO:0050776~regulation of immune response	7	*C1QA, SLC11A1, LAT2, HMOX1, FCER1G, C1QC, GIMAP1*
GO:0050778~positive regulation of immune response	6	*C1QA, SLC11A1, LAT2, FCER1G, C1QC, GIMAP1*
GO:0002443~leukocyte mediated immunity	5	*C1QA, SLC11A1, LAT2, FCER1G, C1QC*
GO:0002682~regulation of immune system process	8	*C1QA, SLC11A1, LAT2, HMOX1, SCIN, FCER1G, C1QC, GIMAP1*
Annotation Cluster 4	Enrichment Score: 2.0	
Term	Count	Genes
GO:0060348~bone development	6	*AMBN, CTGF, ACP5, STC1, GPNMB, SPP1*
GO:0031214~biomineral formation	4	*AMBN, ENPP1, GPNMB, SPP1*
GO:0001503~ossification	5	*AMBN, CTGF, STC1, GPNMB, SPP1*
GO:0001501~skeletal system development	7	*AMBN, CTGF, MMP9, ACP5, STC1, GPNMB, SPP1*
Annotation Cluster 5	Enrichment Score: 1.6	
Term	Count	Genes
GO:0001775~cell activation	7	*CD48, SLC11A1, LAT2, PLEK, LCP1, BLNK, GIMAP1*

3.2.1 DISCUSSION

Transcriptome data highlight the adipose tissue complexity

The deep sequencing analysis of the pig backfat transcriptome allowed for finding thousands of expressed genes and transcripts. In the present study, we applied stringent cleaning and filtering procedures of the sequencing data and, on average, 90 million reads per sample were mapped, obtaining a higher sequencing depth compared to previous studies (Chen et al., 2011; Jiang et al., 2013; Sodhi et al., 2014; Wang et al., 2013b).

The adipose tissue is not only metabolically and transcriptionally active but also has been recognised as an important endocrine organ (Kershaw and Flier, 2004; Trayhurn, 2005). Adipocytes are a dynamic and highly regulated population of cells (Moreno-Navarrete and Fernández-Real, 2012; Rosen and MacDougald, 2006). Our results agree with these data supporting the characterisation of the adipocytes as highly specialised endocrine cells that can play key roles in various physiological processes. The multifunctionality and complexity of the tissue is witnessed also by the high number of transcripts (more than 60 000) found in the present study, including many new transcripts from previously non-annotated loci in the porcine genome. The majority of the reconstructed sequences are novel isoforms of already known genes that express more than two different transcripts each. Similar patterns are observed in human cells (Djebali et al., 2012), and the high quality of the sequenced reads used in our analysis supports the idea that this is more attributable to an incomplete annotation of the transcript isoforms expressed in pig backfat than to transcript reconstruction artefacts. In our analysis for almast half of the expressed genes different isoforms have been found for the same locus. The detected splicing variants may contribute to improve knowledge about the porcine transcriptome and to refine the current swine genome annotation. The new PLIN2 isoforms reported above are an interesting example, especially if compared to the human genome where at least eight PLIN2 transcript isoforms are annotated and only four of them are coding. Remarkably, three human PLIN2 isoforms encode N-terminal truncated amino acid chains that are similar to the truncated isoform we reconstructed in our study and whose function has not yet been elucidated. Furthermore, Russell et al. (2008) identified in a PLIN2-deficient mouse cell line the expression of a PLIN2 C-terminal truncated protein that may partially replace the function of the full-length protein. Additional studies are needed to understand if and how the short transcript we found to be differentially expressed could change the gene functions compared to the wild-type long protein.

Functional characterization of the adipose tissue expression profile

The profile of the subcutaneous adipose tissue transcriptome in pigs was delineated, and the functional analysis of the genes expressed in backfat tissue was performed to understand their metabolic role and to connect them to specific competencies of the tissue. We did not find particular differences between the functional categories of the genes expressed in the backfat tissue of FAT and LEAN pigs. Furthermore, among the most highly expressed genes in the fat tissue, many are involved in metabolic pathways and biological processes related to protein metabolism, oxidoreductase activity for ATP production, regulation of lipid synthesis and degradation.

Genes differentially expressed between LEAN and FAT animals converge and connect to specific functions

The detection of DEGs and DETs was obtained by a stringent procedure grounded on the integration of different methods for expression estimation, and differential expression testing, as conducted in a recent study (Ropka-Molik et al., 2014) focused on muscle tissue gene expression in pigs of different breeds. In the present study, which compared pigs of the same breed and reared under standard conditions, we detected significant gene expression variations. The sensitivity of our approach was supported by the successful validation of all 11 DEGs assayed.

We analysed the biological functions of genes differentially expressed between FAT and LEAN animals (Figure 24). It is interesting to note that the main differences were found for functional categories of genes related to inflammation and immunity that were more highly expressed in FAT pigs. The genes with lower levels of expression in FAT animals include some heat shock protein genes. The biological functions of DEGs show a stronger activation in adipose tissue of FAT pigs of important processes involved in hypertrophy and adipogenesis, such as differentiation and maturation. Supposedly, these biological processes could be altered in adipose tissue of FAT pigs due to dysregulated adipose metabolism and endocrinology, similar to what was hypothesised in humans (Sethi, 2010). On the whole, there is a consistent difference concerning the biological functions characterising the most expressed genes on backfat tissue and those of the genes differentially expressed between FAT and LEAN pigs.

Some genes more highly expressed in FAT animals could modulate backfat physiological processes

Specific DEGs more highly expressed in FAT pigs participate in biochemical pathways related to and involved in adipocyte metabolism and adipose tissue physiology. Ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) encodes a catalytic enzyme involved in adipocyte maturation (Liang et al., 2007). Pan et al. (2011) showed that the over-expression of ENPP1 in a human cell line resulted in adipocyte insulin resistance and demonstrated an association with fatty liver, hyperlipidemia and dysglycaemia. Accordingly, the study by Chandalia et al. (2012) underlined an increased ENPP1 expression in adipose tissue associated with defective adipocyte maturation leading to pathogenesis of insulin resistance and its associated complications for glucose and lipid metabolism in the absence of obesity. In addition, Meyre et al. (2005) reported the presence of three ENPP1 SNPs in human genes associated with adult obesity and increased risk of glucose intolerance and type 2 diabetes. Furthermore, the genes acid phosphatase 5, tartrate resistant (ACP5) and lysozyme (LYZ), which in this research have higher transcriptional levels in FAT pigs, have been reported to be involved in excessive backfat deposition in pigs and in the development of atherosclerosis (Padilla et al., 2013).

In the present research, some genes overexpressed in the adipose tissue of FAT pigs, namely STC1, EGR2 and INHBB, are related to adipocyte differentiation and adipocyte maturation. STC1 (stanniocalcin 1) has been reported in literature to be up-regulated during adipogenesis and to modulate steroidogenesis. Serlachius and Andersson (2004) related STC1 up-regulation to the set of survival genes in adipocyte differentiation, which is also associated with overexpression of the anti-apoptotic protein BCL2 reported to be involved in the inflammation pathway. EGR2 (early growth response 2) is a direct target of mir-224-5p, a negative regulator of adipocyte differentiation that is down-regulated during the early process of mouse adipocyte differentiation, and the expression of EGR2 is increased (Peng et al., 2013). The INHBB (inhibin beta B) gene encodes the activin B subunit, which is part of the inhibins/activins family of proteins with cytokine and hormone activity. In human and mice, INHBB has been associated with physiological and metabolic modifications during adipogenesis when it is highly expressed and is the predominant activin in human adipose tissue (Hoggard et al., 2009). INHBB is member of TGF-protein superfamily of secreted growth factors involved in many biological responses including regulation of apoptosis, proliferation and differentiation of human adipocytes, tissue remodelling and inflammatory immune response (Dani, 2013). It can be

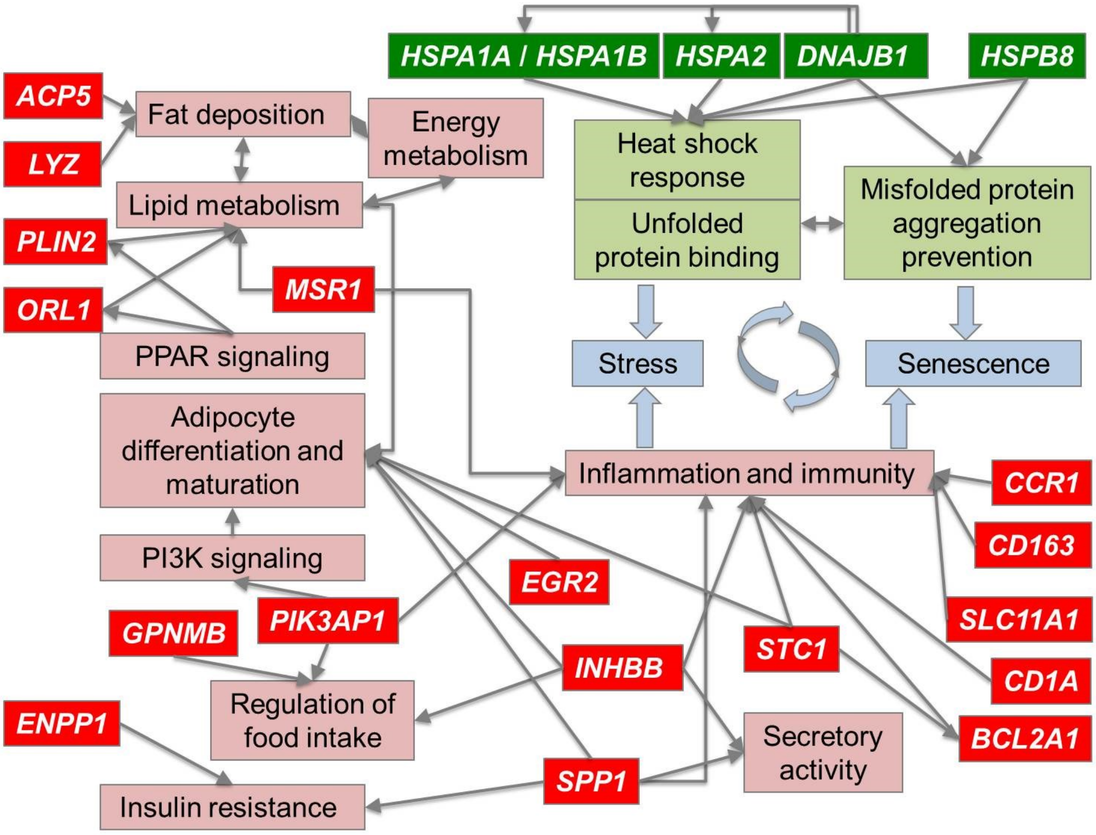


Figure 24. The impact of genes differentially expressed between FAT and LEAN animals on specific and connected biological processes. Genes differentially expressed in FAT vs. LEAN pigs converge on specific functions that are more activated or impaired in FAT pigs. Genes and functions upregulated and downregulated in FAT pigs are shown in red and green respectively. Several genes more highly expressed in FAT pigs are linked to fat deposition and lipid metabolism, to adipocyte differentiation and maturation or to signalling pathways regulating them; FAT pigs also show increased expression of genes involved in inflammation and immunity and increased expression of genes involved in the control of complex behaviour, also by inflammation-mediated secretory activity of adipocytes. Metabolic alterations induce chronic stress in the adipose tissue. FAT pigs show under-expression of several genes involved in stress response by unfolded protein binding and misfolded protein aggregation prevention. The impairment of these functions might in turn augment inflammation and the consequent secretory activity and possibly induce senescence.

hypothesised that in FAT pigs the proadipogenic INHBB gene expression increases, as it is involved in the differentiation of pre-adipocytes into mature adipocytes and that INHBB is involved in many physiological processes including the control of food intake and energy metabolism through the regulation of hypothalamic and pituitary hormone secretions. Another gene overexpressed in FAT pigs related to feeding and pituitary secretions is GPNMB (glycoprotein transmembrane NMB). GPNMB is one of the receptors activated by bombesin-like endogenous peptide ligands, such as the gastrin-releasing peptide (GRP), neuromedin B (NMB) and neuromedin C (GRP18-27). These receptors are involved in the regulation of many biological functions including thermoregulation, feeding, pituitary, gastric and pancreatic secretion. The NMB/NMB-R pathway is involved in the regulation of a wide variety of behaviours, such as spontaneous activity, feeding and anxiety-related behaviour (Yamada et al., 2002).

The OLR1 [oxidised low-density lipoprotein (lectin-like) receptor 1] gene is more highly expressed in FAT pigs compared to LEAN animals. This gene codes for a LDL receptor that belongs to the C-type lectin superfamily and is one of many target genes, including perilipins, of the PPAR signalling pathway, which is involved specifically in lipid metabolism and fatty acid transport. In this way, OLR1 is a receptor that mediates the recognition, internalisation and degradation of oxidatively modified low-density lipoprotein by vascular endothelial cells. OLR1 removes oxidised low-density lipoproteins from the circulation as part of lipid metabolism pathways (Mehta and Li, 2002).

Genes involved in immunity and inflammation are more highly expressed in FAT animals

Some other genes overexpressed in FAT pigs are related to immunity. Links between inflammation and human obesity as well as metabolic diseases are well-known mechanisms based on the recruitment of immune cells into adipose tissue (Kabir et al., 2014). The development of a pre-inflammatory condition in the presence of dysregulated excessive adipogenesis is associated with adipose macrophage infiltration and activation. From our study, we can hypothesise a similar process in backfat tissue of FAT pigs, where we identified the over-expression of the gene macrophage scavenger receptor 1 (MSR1), encoding a membrane glycoprotein that in humans is involved in the pathologic deposition of cholesterol in arterial walls during atherogenesis (Haasken et al., 2013). Additionally, the overexpression of secreted phosphoprotein 1 (SPP1) in FAT pigs can suggest a hypothesis that this gene encodes a protein acting as a pro-inflammatory cytokine that promotes monocyte chemotaxis and cell motility and might link, in pigs as in mice, fat accumulation to the development of insulin resistance by sustaining inflammation and the accumulation of macrophages in adipose tissue (Nomiyama et al., 2007). Interestingly, a porcine SPP1 gene polymorphism was associated with backfat thickness in a Landrace 9 Jeju (Korea) Black pig F2 population (Han et al., 2012). SPP1 might play a key role in the pathway that leads to type I immunity enhancing interferon-gamma and interleukin-12 production and suppressing interleukin-10 (Ashkar et al., 2000). Therefore, these data allow hypothesising SPP1 as a gene associated, in pigs as in human, with the link between obesity, adipose tissue inflammation and insulin resistance. In addition, phosphoinositide-3-kinase adaptor protein 1 (PIK3AP1), more highly expressed in FAT pigs, is a positive regulator of phosphatidylinositol 3-kinase (PI3K) signalling. The PI3K signalling pathway has a key role in the insulin-dependent regulation of adipocyte metabolism (glucose and lipid metabolism). In addition, PI3K participates in obesity-associated inflammatory cell recruitment (neutrophils and macrophages), as well as in the CNS-dependent neurohumoral regulation of food intake/energy expenditure (Beretta et al., 2015; McCurdy and Klemm, 2013).

Other genes found in the present research and related to inflammatory conditions of the adipose tissue in FAT pigs are particularly interesting to mention: CD163, a member of the scavenger receptor cysteine-rich superfamily (Guo et al., 2014; Smith et al., 2014); solute carrier family 11 (proton-coupled divalent metal ion transporter), member 1 (SLC11A1), a gene involved in resistance to Salmonella infection (Kommadath et al., 2014); chemokine (C-C motif) receptor 1 (CCR1), which was previously found to be overexpressed in obese pigs (Kogelman et al., 2014); BCL2-related protein A1 (BCL2A1), a gene found to be overexpressed in pigs with a high obesity index and that is related to immunity, inflammatory pathways and osteoclast differentiation (Kogelman et al., 2014); and CD1a molecule (CD1A; indicated as PCD1A in the cited paper), a surface antigen involved in immunity that was found to be overexpressed in obese pigs by Kogelman et al. (2014). The same authors highlighted a strong connection between fat deposition on the body (obesity), immunity and bone development. They also indicated that the CCR1 gene is a strong candidate for regulation of immune response as it encodes a receptor of pro-inflammatory chemokines in adipose tissue, playing a pivotal role in obesity-associated diseases (Kabir et al., 2014; Lumeng and Saltiel, 2011).

Heat shock response, protein folding and repair are impaired in FAT animals

Considering the 14 genes with lower expression levels in FAT animals, direct relationships with lipid metabolism are not apparent. However, the ‘unfolded protein binding’ function is enriched among these genes, which include five (DNAJB1, HSPA1A, HSPA1B, HSPA2 and HSPB8) encoding functionally linked heat shock proteins. Heat shock proteins are involved in stabilisation of existing proteins against aggregation, mediating the folding of newly translated proteins in the cytosol and in organelles, and also in the ubiquitin–proteasome pathway. DNAJB1, a member of the Hsp40 family, promotes protein folding and prevents misfolded protein aggregation, just as HSPB8, a member of the Hsp20 family, does (Vicario et al., 2014). DNAJB1 also stimulates the ATPase activity of a protein of the Hsp70 family to which other genes with lower expression levels in FAT pigs (HSPA1A, HSPA1B and HSPA2) belong, indicating a possible functional link between these four genes. Our results suggest a general impairment of the protein folding and repair in the fattest animals, in accordance with previous observations of studies carried out on human obesity. Obesity is a pathological human condition in which a chronically positive energy balance induces in adipocytes, the cells in charge of storing the excess of energy in fat depots, a persistent stress activating in turn defence processes such as autophagy or apoptosis.

As reviewed by Newsholme and de Bittencourt (2014), if the heat shock response, a key component of the physiological response to resolve inflammation, is hampered in adipose tissue, the adipocyte metabolic stress triggers fat cell senescence with reduction of the heat shock proteins activity. In this condition, the advance of inflammasome-mediated secretory activity from adipose to other tissues promotes cellular senescence in many other cells of the organism, aggravating obesity-dependent chronic inflammation. This mechanism also could have been activated in the FAT pigs of our experiment (Figure 24) due to a genetic aptitude of the fattest animals towards a higher fat deposition and adiposity similar to obesity. Indeed, a decrease in the synthesis of the mRNAs of the heat shock proteins and an increase in the expression of many genes related to inflammatory status and immune response is a characteristic of the fattest pigs. For instance, an increase in the expression of INHBB and SPP1 denotes the augmented production of cytokines and the higher expression of ENPP1 and PIK3AP1 may indicate a status of insulin resistance, one of the typical signals connected with obesity.

Pig backfat deposition and impaired stress response may activate inflammation

Our results agree with recent studies showing that several immune system and anti-inflammatory processes are activated and play a critical role in the response to fat accumulation in porcine backfat tissue (Sodhi et al., 2014) and in visceral fat tissue (Toedebusch et al., 2014; Wang et al., 2013b). Wang et al. (2013b) and Zhou et al. (2013) used three female Landrace pigs to identify DEGs between subcutaneous, visceral and intramuscular fat, indicating that visceral and intramuscular adipose tissues are associated mainly with inflammatory features of the tissue and immune response. Our data suggest that in backfat a predominant role of immunity processes is related to increased adipose tissue deposition as well.

The results obtained seem to sustain the hypothesis that the high fat accumulation in adipose tissue of pigs can determine the development of an inflammatory process producing a cascade of defence and adaptive reactions in the tissue, such as activation of the immune system and mesenchymal cells differentiation in adipocytes. A deeper knowledge of the metabolic processes involved in fat deposition can be very important in developing the use of pig as a model species to study obesity and related disorders for humans because of their similar anatomy and physiology (Litten-Brown et al., 2010; Spurlock and Gabler, 2008; Varga et al., 2010) and considering the above-described similarities between pigs and humans.

To fully elucidate the complex gene network regulating backfat deposition in pigs, it is important to extend the basic knowledge by further coding and non-coding transcriptome characterisation. Additional information would probably come from studying interactions between the differentially expressed long RNAs identified in the present paper and the regulatory microRNAs expressed in porcine adipose tissue identified in some of the same animals (Gaffo et al., 2014). The results of the present work unlock the opportunity that some of the identified DEGs might be used as biomarkers (Ibáñez-Escriche et al., 2014) to improve carcass fat traits and to look for SNPs regulating their expression to be included in selection schemes to make the pig production chain more sustainable.

3.2.2 MIRNOME OF ITALIAN LARGE WHITE PIG SUBCUTANEOUS FAT TISSUE: NEW MIRNAS, ISOMIRS AND MORNAS

In this study, we used RNA-Seq to study the Italian Large White adult pig backfat miRNome. Sequencing data were analysed by means of several bioinformatic tools, including the *miR&moRe* computational pipeline described in Bortoluzzi et al. (2012) and *miRDeep2* (Friedländer et al., 2011). The miRNAs, isomiRs and moRNAs that we identified and characterized outline the complex nature of the porcine backfat tissue miRNome.

3.2.2 MATERIALS AND METHODS

Sample collection and sequencing

Small RNA sequencing data from two ILW pig backfat samples were used in this study (see chapter "Pig subject selection").

Bioinformatics analysis

Sequencing data were processed by the mir&more software pipeline. mir&more has been extended and adapted for its application to swine transcriptomic data. The pipeline work flow involves filters on raw sequencing data, assignment of reads to genomic loci by mapping to reference sequences, methods for known miRNAs expression quantification and isomiRs characterization as well as for miRNA and moRNA discovery. The pipeline is structured in three main branches: one is devoted to raw data pre-processing, the other two perform respectively the quantification and characterization of known miRNAs and the discovery of mature miRNAs in known precursor sequences. Further, we applied mirdeep2 to the data for the identification of miRNAs belonging to precursor hairpins that are still unknown in pig genome. Finally, we performed a prediction of the transcripts targeted by the miRNAs expressed in the tissue and a functional enrichment analysis on the target prediction outcome. More details about each step of the bioinformatic analysis follow.

Data pre-processing

A preliminary quality check of raw data was performed to clean out low quality and sample contamination reads to obtain a high-quality set of reads for downstream analysis. Raw reads were clipped from their adapter sequences using the FASTX-Toolkit software package (http://hannonlab.cshl.edu/fastx_toolkit/index.html), and unclipped reads were discarded. Reads with an overall sequence mean Phred quality lower than 30, reads with more than two nt (nucleotides) with quality lower than 20 and reads with uncalled bases were also discarded. Furthermore, reads with unique sequence and count <10 were considered as ground noise and were removed. Subsequently, reads between 15 and 30 nt long were selected by means of a Python script, in which we used the HTSeq package (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html). Read length thresholds were chosen according to a survey of the length distribution of all vertebrate mature miRNAs reported in miRBase v.19 (Griffiths-Jones et al., 2006, 2008; Griffiths-Jones, 2004; Kozomara and Griffiths-Jones, 2013). miRNA lengths resulted ranging from 15 to 28 nt, with average, median and mode equal to 22 nt. We kept reads up to 30 nt to allow identification of possible long isomiRs.

Mapping

The selected reads were mapped using Bowtie v.0.12.7 (Langmead et al., 2009). We used the Sscrofa9 genome as a reference to identify the reads that mapped to multiple genomic loci out of known miRNA precursor sequences. Reads mapping to more than five loci outside known miRNA genes were discarded. Moreover, we generated 271 known pig miRNA extended hairpin precursors (e-hairpins) as references for further mapping. E-hairpins are built as the nucleotide sequences of genomic regions including known hairpins from miRBase, plus the 30 nt upstream and the 30 nt downstream surrounding each annotated hairpin. Reads were mapped to e-hairpins allowing at most two mismatches within the 3' region or at most one mismatch in non-3' parts of the sequence. This approach accounts for the existence of SNPs; alternative miRNA processing; and post-processing modifications, such as nucleotide addition and editing and tolerates residual sequencing errors.

Known miRNA identification and quantification

To identify expressed miRNAs and to quantify their level of expression, the positions of the mapped reads were compared with the known mature miRNA coordinates in the e-hairpins. Matched reads were classified as perfectly matching to known miRNA positions ('exact'), perfectly matching to miRNA precursors and exceeding the known miRNA boundaries for at most three nt ('short-long'), '1-mismatch' and '2-3'-mismatch'. According to Bortoluzzi et al. (2012), we call 'expressed RNAs elements' (ERE) the blocks of alignments forming a group of subsequent reads, each one having a start position in the hairpin within four nt from the starting nucleotide of the previous read. The sum of read counts per ERE gives the quantification of the ERE expression level. The expression of known miRNAs is inferred from that of the corresponding EREs.

Sister miRNA discovery

Only 'exact' alignments were allowed for the identification of new miRNAs. At this stage, the search was performed exclusively among the precursor hairpins for which at most one mature miRNA is annotated. The prediction of a mature miRNA expressed from the hairpin strand that is opposite to the annotated miRNA (sister miRNA) was carried out by the joint analysis of the ERE position inferred from read mapping and the hairpin folding probability computed by the ViennaRNA RnaFold v1.8.5 package (Bompfünewerer et al., 2008; Hofacker and Stadler, 2006; Hofacker et al., 1994; McCaskill, 1990; Zuker and Stiegler, 1981). These two parameters were used to predict the most probable hairpin structure that defines a miRNA duplex.

MoRNA discovery

Reads that align with no mismatch to the e-hairpin sequence outside the reference miRNA positions identify expressed moRNAs, and the number of alignments quantifies the moRNA expression. ERE with central nucleotide localized upstream of the region covered by the 5' mature miRNA, or downstream of the region covered by the 3' mature miRNA, were named respectively 5' moRNAs and 3' moRNAs.

Novel miRNAs from new hairpins

We applied the *miRDeep2* program for the genome wide prediction of new hairpin precursors and mature miRNAs expressed in porcine fat tissue. As input to the *miRDeep2* software, we set the Sscrofa9 genome reference and miRNA annotations of *Homo sapiens* and *Mus musculus*, which are the two mammalian species with the largest sets of annotated miRNAs. To allow comparison between *miR&moRe* and *miRDeep2* expression estimates, the *miRDeep2* read counts were normalized using a rescaling factor computed as the proportion of the total amount of read mapped to known miRNAs respectively by *miRDeep2* and *miR&moRe*.

Target prediction and functional enrichment

To our knowledge, no miRNA-target predictions specific of pig data are available. Therefore, we performed a custom target prediction by the application of the *miRanda* method (v3.3a) (Enright et al., 2003) to the whole set of the detected miRNA sequences and to the 3'-UTRs annotated for the pig genome (other parameters were set to their default values). We retrieved the 3'-UTR sequences from the ENSEMBL repository using the biomaRt R package (Durinck et al., 2009) in a custom script. Then, to investigate the possible biological role in backfat adipose tissue of the highest expressed miRNAs, the human orthologs of the predicted miRNA targets were established using the ENSEMBL database and given as input to DAVID (Dennis et al., 2003).

qRT-PCR validation

qRT-PCR analysis was conducted on the two samples used for the RNA-Seq analysis and on 18 additional backfat samples from Italian Large White pig individuals farmed and slaughtered as indicated above. The 18 additional samples were used as biological validation for two new sRNAs to confirm the results obtained with the first two samples. The TaqMan® Micro RNA Assay kit (Applied Biosystems) was used, following the customer protocol. For each sRNA, analyses were performed in triplicate. As an internal reference, we used the U6 snRNA (Chen et al., 2012; Yu et al., 2012).

3.2.2 RESULTS

Sequencing and data pre-processing.

The sequencing of the two pig backfat samples produced respectively 68 126 656 and 64 250 776 reads. To attain a picture of the normal swine fat miRNome and to enhance our discovery power, we pooled together the two sequencing datasets, obtaining 132 377 432 raw reads. The sequences of the small RNAs reported in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE47748.

The adapter-clipping step discarded 19 787 039 reads (14.9%), the quality and length filtering further discarded 43 310 672 reads (32.7% of the initial raw reads set). Some 1 606 729 (1.2%) reads with sequence counted only once in the dataset were discarded, as were the 9 799 677 (7.4%) reads mapping to more than five loci outside known miRNAs. The filtering step resulted in 57 873 315 clean reads, corresponding to 43.7% of the initial raw reads (Figure 25a), which were used for subsequent analysis. All the clean reads had an average quality >30 (quantified with a Phred score), and most of them (86.4%) had average quality >36 (Figure 25b,c). With respect to the read length, the majority of the clean reads (84.6%) were between 20 and 23 nt long (Figure 25d). Reads with lengths between 15 and 19 nt and between 24 and 30 nt were less abundant, but still were present in a non-negligible amount (Figure 25d).

Known miRNAs expressed

From miRBase, we considered 343 known mature miRNAs annotated in 271 swine e-hairpins (see 'Materials and methods' for the e-hairpin definition). Because the processing of different hairpins may produce the same mature miRNA, the actual number of known

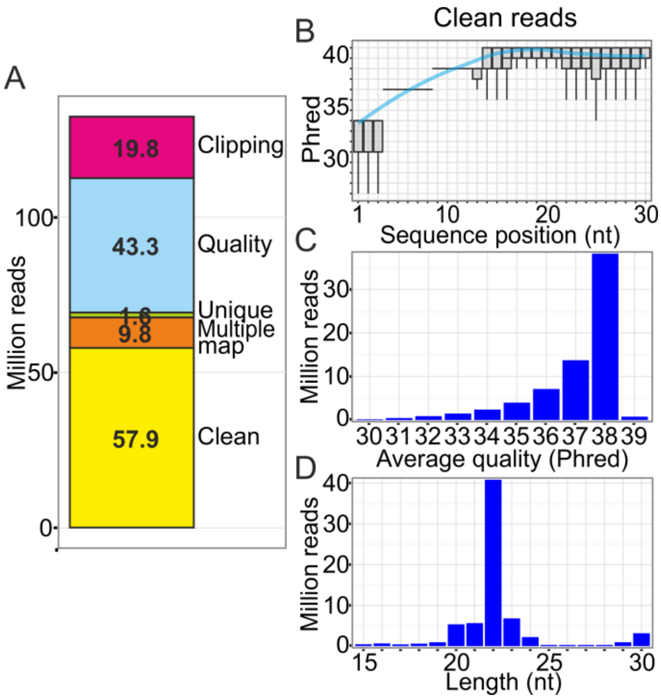


Figure 25. Raw data processing and filtering results. Panel (a) shows results of reads processing and filtering steps implemented to obtain a clean set of reads for all the following analyses. The overall height of the bar is proportional to the number of raw reads obtained from the sequencing. The four bars stacked on the top indicate the number and the proportion of discarded reads in the sequentially applied filtering steps. The average quality filter step (light blue bar) discarded the largest amount of reads. On the bottom, the yellow bar represents the set of reads that successfully passed all the filtering steps. The main properties of the clean reads set are described by the boxes on the right as (b) Phred base quality score distribution per position on the read, (c) distribution of the reads according to their average quality and (d) distribution of reads according to their length.

mature miRNA sequences is 306. Besides, 72 hairpins are associated with two known sister miRNAs, whereas in 199 hairpins, only one known miRNA is reported.

Known miRNAs identified with an expression level of at least 10 reads numbered 222 (Table S1). miRNA expression level distribution was considerably skewed, ranging from 10 to about 33 million reads. Few miRNAs were highly expressed (Figure 26a): only 47 known mature miRNAs (21% of the expressed known miRNAs) totaled up to 99% of the overall miRNA expression, and only nine of them totaled up to 90%. The most expressed miRNA, ssc-miR-10b, alone amounted to about 64% of the detected known miRNA expression. Other highly expressed miRNAs were ssc-miR-143-3p, amounting to 8.5%, ssc-miR-10a-5p and ssc-miR-191 (each one about 4.3%), ssc-miR-22-3p and ssc-miR-27b-3p, each one amounting to over 2% (Figure 26a).

We predicted with *miRanda* the possible targets of the nine most abundant miRNAs, using as input the 15 053 pig 3’-UTR sequences available in the ENSEMBL database. The predictions associated with the top 20% scores of max energy were selected for successive analysis. We obtained 1889 pig transcripts (corresponding to 1687 porcine genes) that were targets of the highly expressed miRNAs (Tables S2–S10). Some 348 genes were joint targets among up to five different miRNAs. Only 22% of target genes were functionally annotated in DAVID; thus, functional enrichment analysis was carried out using human orthologs of the pig genes. By this strategy, almost all the orthologous genes (1525 out of 1567) were recognized by DAVID. We focused on Gene Ontology biological process (GO

BP) terms and KEGG pathways. The significantly (P-value < 0.01) enriched 26 GO BP and nine KEGG pathways are listed in Table 7 and Table 8 respectively. Prominent processes were regulation of transcription and gene expression, intracellular transport, membrane organization, protein modification, intracellular signaling cascade, nuclear import, cell motility, regulation of actin cytoskeleton and cellular metabolism. Interesting KEGG pathways targeted by the most represented miRNAs were the Wnt signaling, insulin signaling and axon guidance pathways.

Table 7. GO biological process terms enriched in the group of predicted target genes of the top nine most expressed miRNAs in porcine backfat tissue. The functional enrichment was performed by DAVID on 1525 functionally annotated human orthologs of porcine genes.

GO biological processes	***n***	**%**	***P*-value**
Vesicle-mediated transport	79	5.2	4.40E-06
Membrane organization	56	3.7	1.90E-05
Intracellular signalling cascade	140	9.2	7.50E-05
Protein modification process	154	10.1	3.30E-04
Positive regulation of biosynthetic process	83	5.4	3.40E-04
Positive regulation of transcription	70	4.6	3.60E-04
Positive regulation of gene expression	71	4.7	5.10E-04
Protein kinase cascade	49	3.2	7.90E-04
Apoptosis	72	4.7	8.30E-04
Positive regulation of RNA metabolic process	60	3.9	8.80E-04
Positive regulation of nitrogen compound metabolic process	75	4.9	0.001
Positive regulation of transcription from RNA polymerase II promoter	48	3.1	0.002
Protein amino acid phosphorylation	77	5.0	0.002
Positive regulation of cellular metabolic process	97	6.4	0.002
Cell motility	41	2.7	0.002
Positive regulation of metabolic process	100	6.6	0.002
Positive regulation of macromolecule metabolic process	94	6.2	0.002
Cell death	81	5.3	0.002
Anatomical structure formation involved in morphogenesis	45	3.0	0.003
Regulation of transport	52	3.4	0.005
Cellular protein metabolic process	224	14.7	0.005
Regulation of transcription from RNA polymerase II promoter	79	5.2	0.006
Positive regulation of molecular function	65	4.3	0.009
Mitochondrial transport	13	0.9	0.009
Protein transport	81	5.3	0.010
Protein catabolic process	68	4.5	0.010

Table 8. KEGG pathways enriched in the group of predicted target genes of the top nine most expressed miRNAs. The group of 1525 functionally annotated human orthologs of porcine genes was considered for functional enrichment.

Term	*n*	%	*P*-value
T-cell receptor signalling pathway	19	1.2	5.80E-03
Phosphatidylinositol signalling system	14	0.9	1.20E-02
Pathways in cancer	42	2.8	1.20E-02
Axon guidance	20	1.3	1.70E-02
B-cell receptor signalling pathway	13	0.9	3.00E-02
Ubiquitin mediated proteolysis	20	1.3	3.00E-02
Hematopoietic cell lineage	14	0.9	3.70E-02
Wnt signalling pathway	21	1.4	4.10E-02
Insulin signalling pathway	19	1.2	4.80E-02

New sister miRNAs discovered

We define 'new sister miRNAs' as the mature miRNA sequences that are derived from precursors in which only one mature is known and annotated and that are produced from the hairpin arm opposite to that giving rise to the known miRNA. We identified the expression of 68 new mature sister miRNAs (Tables S1 & S11). The new miRNAs have been named from their precursor name and their position in the hairpin by attaching the suffix -3p or -5p. On average, the

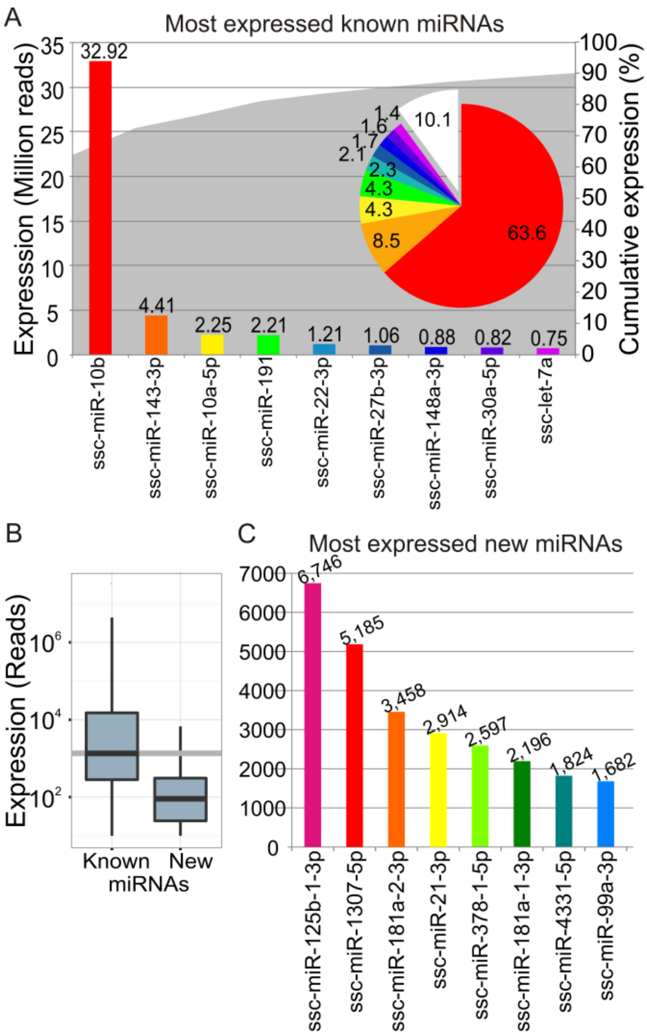


Figure 26. Most expressed known and new miRNAs. (a) Nine most expressed known miRNAs account for 90% of the pig backfat tissue expression (reads mapped to known hairpins). The bars and the grey shadow below show respectively the read count and the cumulative percentage of expression corresponding to the nine most expressed miRNAs. The pie chart highlights the contribution of each miRNA to the overall expression (colour notation of the pie corresponds to the bars). (b) The boxplot shows the distribution of expression, in logarithmic scale, of known and new miRNAs identified in known pig miRNA precursors. The median level of expression of the new miRNAs is lower than the known miRNA (grey horizontal line). (c) The barplot shows eight new miRNAs with an expression level higher than the median expression of known miRNAs.

new sister miRNAs were less expressed than were the group of known miRNAs (Figure 26b). As shown in Figure 26c, eight new miRNAs were expressed above the median expression level of the known miRNAs (1311 reads). Finally, we noticed that four new miRNAs, ssc-miR-2411-3p, ssc-miR-4331-5p, ssc-miR-4336-3p and ssc-miR-4333-3p, were the only small RNAs that were expressed by their precursor, whereas their known miRNA mate was not detected.

5'- and 3'-miRNA expression

Considering known and new miRNAs together, we detected 291 mature miRNAs expressed from 204 different hairpins. Among the precursors, 111 (54%) expressed both sister miRNAs in the pig fat tissue, whereas in the remaining cases only one miRNA per hairpin was expressed. This is consistent with previous findings of concurrent sister miRNAs expression in a cell (Biasiolo et al., 2011). The scatterplot of 5'- and 3'-miRNA expression values (Figure 27a) shows the expression ratio of miRNAs that derived from a same hairpin expressing both arms (3'- over 5'-

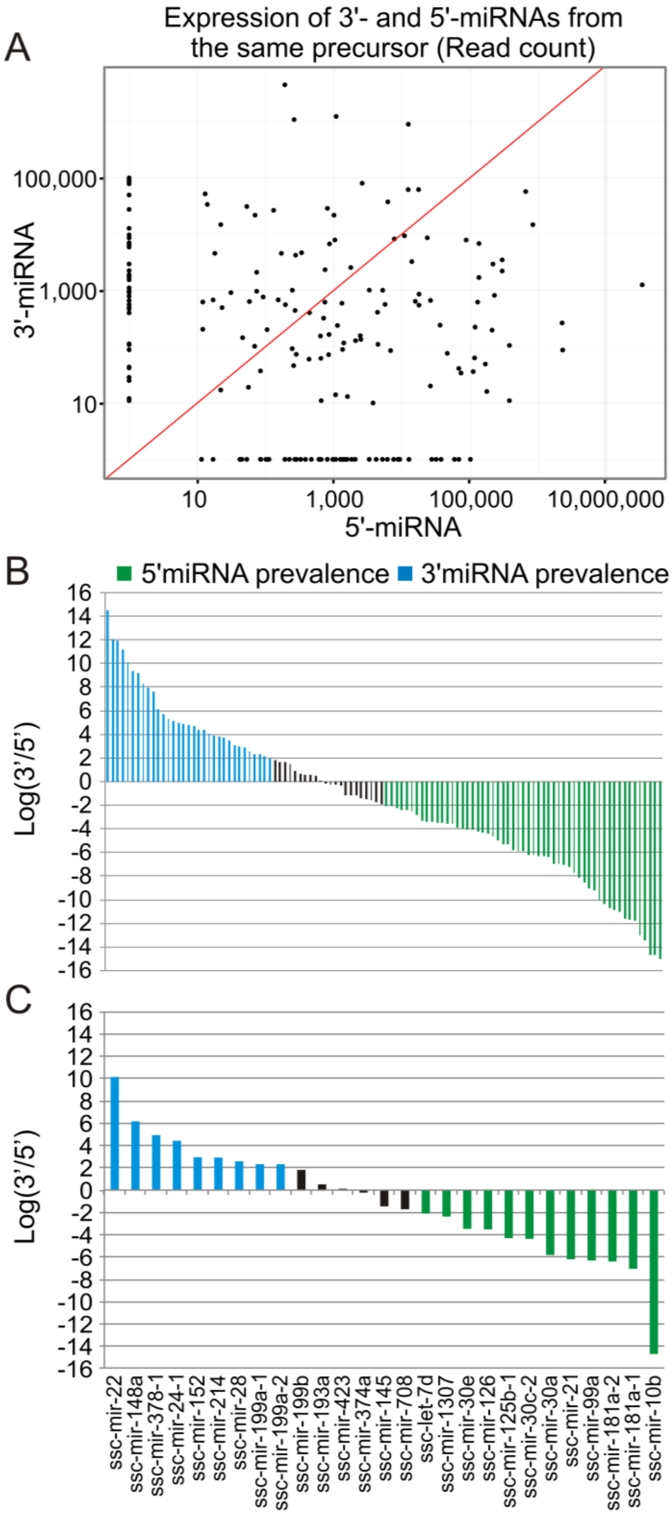


Figure 27. Expression levels of miRNAs derived from the 5' and 3' arms of the same hairpin precursor. The scatterplot in (a) compares the expression levels of miRNAs derived from the 5' and 3' arms of the same hairpin precursor, showing that there is not significant correlation between the values. Moreover, many points lie in the axes, indicating that in several cases only one of the two arms of the precursor is expressed. Panels (b) and (c) show the logarithmic fold change of expression levels calculated for the 5' and 3' miRNAs belonging to the same precursor respectively considering all miRNAs and only miRNAs expressed over the median expression.

miRNA expression). We noted that 22 hairpins (20%) expressed both miRNAs at comparable levels (absolute log fold change being at most two), whereas in 34 and 55 cases, the 3' or the 5' miRNA tended to be prevalent (Figure 27b). We hypothesized that sister miRNAs expressed concurrently at a moderate to high level might be biologically relevant. In this regard, we further assessed those 27 hairpins having both matures expressed over the median expression of all miRNAs: six hairpins (22%) expressed both matures at a comparable level, with a slight tendency towards 5' arm prevalence (Figure 27c).

Variability of miRNA sequences

As mentioned before, recent studies showed that miRNAs are represented in cells as mixtures of isomiRs contributing to the miRNA expression. We investigated sequence variability of 249 miRNAs coming from distinct e-hairpins. We considered first all isomiRs associated with known miRNAs; then, for each miRNA, we selected only isomiRs that each amounted to at least 10% of the miRNA expression ('relevant' isomiRs). Most of the known mature miRNAs (92.8%; 231 out of 249) were from mixes of two to 558 isomiRs. Considering only 'relevant' isomiRs, 81.1% of miRNAs (202 of 249) were represented by

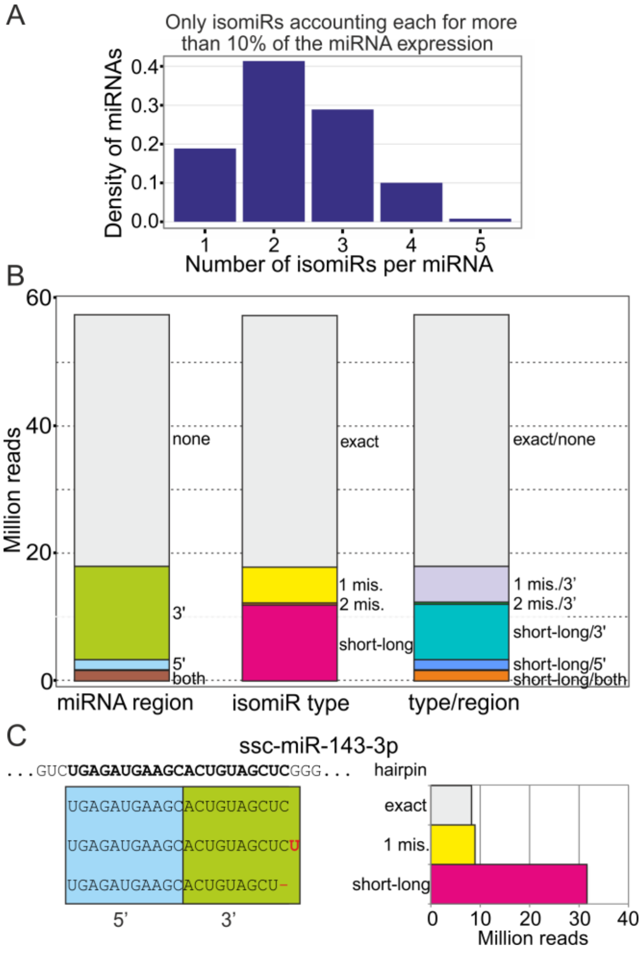


Figure 28. miRNA sequence variants (isomiRs). Most of the known mature miRNAs resulted in mixes of two or more isomiRs. Panel (a) shows the proportion of miRNAs grouped by the number of isomiRs (horizontal axis) composing the expression of each miRNA, considering, for each miRNA, only those isomiRs accounting individually for at least the 10% of the total miRNA expression. Panel (b) details the fraction of expression of isomiRs grouped in different categories. From the left to the right are shown: the miRNA region (5' or 3' half) affected by the variation, the type of sequence difference with the hairpin locus and the known mature miRNA position from miRBase (isomiR type), and the combinations of the previous categories. Results clearly indicate that the category of isomiRs differing from the annotated sequence only in length is prevalent (red portion in the left bar), that only a minority of variants involves the 5' seed region (azure part in the right bar) and that no isomiRs with mismatches in the 5' region are observed, ruling out a more than non-negligible contribution of residual sequencing error to observed variability. Panel (c) shows, as an example, the composition in isomiRs (accounting each for at least the 10% of the total miRNA expression) of the highly expressed *ssc-miR-143-3p*.

two to five isomiRs (Figure 28a). Regarding the isomiR expression fraction, 31.2% of known miRNA expression was composed of isomiRs that are different from the canonical mature miRNA reported in miRBase (the ‘exact’ isomiR; Figure 28b), even in sequence or in length. As shown in Figure 28b, the fraction of expression belonging to the ‘short-long’ isomiRs was prevalent. The miRNA 5’ region, which includes the seed region, is crucial for target recognition in canonical miRNA-target pairs. The minority of variants involved the 5’ region: isomiRs changing only the 5’ region and both the 5’ and 3’ regions amounted respectively to 8.8% and 9.2% of the non-canonical isomiR expression. No isomiR with mismatches in the 5’ region (Figure 28b, rightmost bar) were detected. Conversely, isomiRs with variation (length and/or sequence) in the 3’ region added up to 80% of non-canonical isomiR expression. As an example, in Figure 28c, we report ‘relevant’ isomiRs (three out of 328) of the highly expressed *ssc-miR-143-3p*.

MoRNAs discovered

We detected the expression of 17 moRNAs coming from 16 different precursors (Table 9).

Table 9. Predicted moRNAs with nucleotide sequences, positions in the e-hairpin, and expression levels

Name	**Sequence**	**Expression**	**Hairpin precursor**	**Start**	**End**
ssc-moR-21-5p	AUGGCUGUACCACCUUGUCGGG	685	ssc-mir-21	26	47
ssc-moR-125a-5p	CCACACUGCCGGCCUCUGAG	282	ssc-mir-125a	21	40
ssc-moR-126-3p	CUGUCGGCAGCCCAGCACCGAGA	207	ssc-mir-126	98	120
ssc-moR-16-2-5p	UAGCAAUGUCAGCAGUGCCU	143	ssc-mir-16-2	19	38
ssc-moR-874-5p	CGGCCCCACGCACCAGGGUAAGA	70	ssc-mir-874	45	67
ssc-moR-let-7e-5p	CCUGCCGCGCGCCCCGGGC	54	ssc-let-7e	21	39
ssc-moR-331-5p	UGGUUUGUUUGGGUUUGUU	52	ssc-mir-331	27	45
ssc-moR-24-2-5p	UGUCGAUUGGACCCGCCCUCCG	49	ssc-mir-24-2	22	43
ssc-moR-331-3p	CCAACCUAAACUCGCGCAUCAUUCC	33	ssc-mir-331	102	126
ssc-moR-24-1-5p	CCUCCCUGGGCUCUGCCUCCC	32	ssc-mir-24-1	21	41
ssc-moR-708-5p	GUGAUGUGGUAACUGCCCUC	29	ssc-mir-708	16	35
ssc-moR-99b-5p	CGGAUUCCUGGGUCCUGGCACC	17	ssc-mir-99b	15	36
ssc-moR-103-1-5p	AAGUUUGCUUACUGCCCUC	14	ssc-mir-103-1	24	42
ssc-moR-1343-5p	UGGGGAGCGGCCCCCGGGCGGG	12	ssc-mir-1343	41	62
ssc-moR-let-7i-5p	UCCCCGACACCAUGGCCCUGGC	12	ssc-let-7i	14	35
ssc-moR-324-5p	CUGAGCUGACUAUGCCUCCC	11	ssc-mir-324	22	41
ssc-moR-27b-5p	CGACGACCUCUCUGACGAGGUGC	10	ssc-mir-27b	17	39

The moRNA expression levels were lower than that of the known and new miRNAs. The most abundant moRNA, ssc-moR-21-5p, was less expressed than the known miRNA median expression level. Besides, four moRNAs (Table 10) yielded significant expression between the first and second quartile of total miRNA expression. A dominance of 5’ moRNAs was observed: 15 moRNAs were 5’, whereas only two belonged to a 3’ region. In one case, ssc-mir-331 precursor, we found both 3’ and 5’ moRNAs expressed, with a larger abundance of the 3’ moRNAs (Figure 29a). To understand whether expressed moRNAs can be by-products of miRNAs biogenesis, we considered moRNA expressions in relation with mature miRNAs expressed from the same precursor (Figure 29a). The 5' miRNA was less expressed than was the 3' miRNA in seven of 15 hairpins expressing a 5' moRNA. In 14 of 17 cases, the moRNA is expressed together with the mature miRNA from the same arm of the precursor. We also noticed three cases (ssc-mir-874, ssc-mir-103-1 and ssc-miR-1343) in which the 5' moRNA was expressed alone in the 5' arm and the 3'-miRNA was concurrently expressed.

Table 10. Four moRNAs expressed by extended hairpin precursors are associated with more than 100 reads. For each moRNA, the table indicates name, sequence and expression level as well as the position of the moRNA relatively to the extended hairpin precursor sequence. Bold text indicates the moRNA chosen for validation.

Name	Sequence	Expression (reads)	Hairpin precursor	Start	End
ssc-moR-21-5p	**AUGGCUGUACCACCUUGUCGGG**	**685**	**ssc-mir-21**	**26**	**47**
ssc-moR-125a-5p	CCACACUGCCGGCCUCUGAG	282	ssc-mir-125a	21	40
ssc-moR-126-3p	CUGUCGGCAGCCCAGCACCGAGA	207	ssc-mir-126	98	120
ssc-moR-16-2-5p	UAGCAAUGUCAGCAGUGCCU	143	ssc-mir-16-2	19	38

Furthermore, for the four most expressed moRNAs (ssc-moR-21-5p, ssc-moR-125a-5p, ssc-moR-16-2-5p and ssc-moR-126-3p), we compared moRNA, miRNA

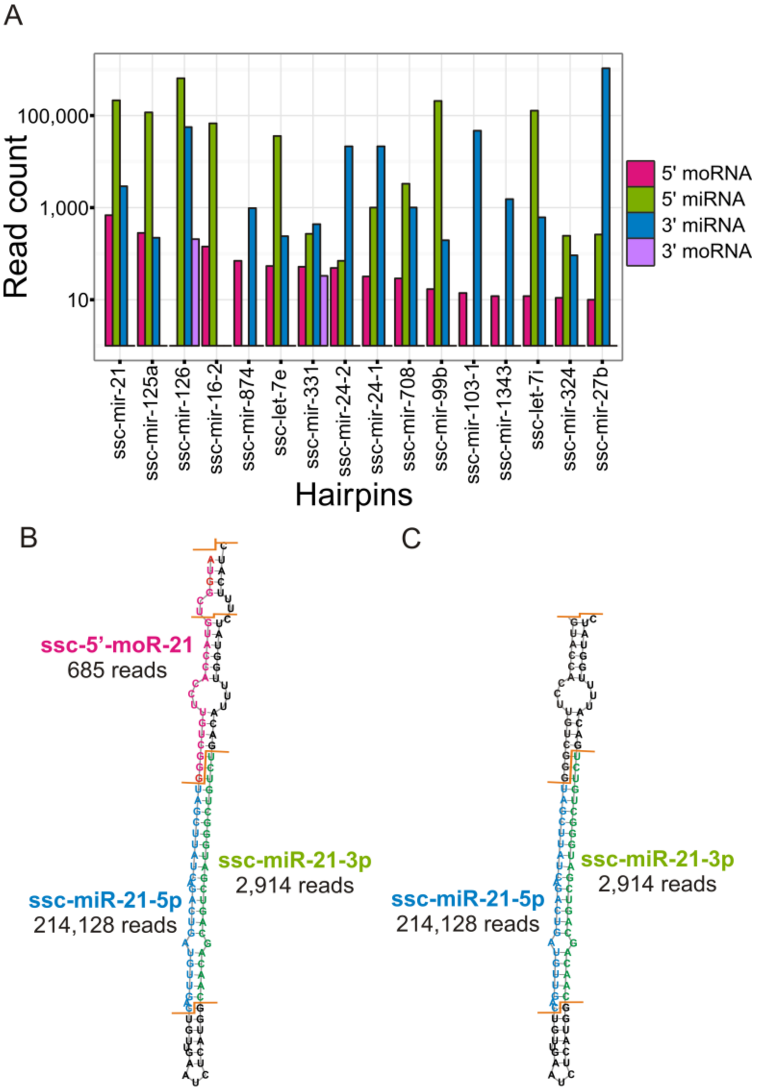


Figure 29 . Hairpin precursors expressing moRNAs. Panel (a) shows the expression in pig backfat tissue of the 16 moRNAs detected and the miRNAs expressed from the same hairpin precursors. The level of expression of moRNAs (crimson and lilac bars) is lower than that of known and newly predicted miRNAs. moRNAs are prevalently expressed by the 5' arm of the precursor, independently by the prevalence of the 5' or 3' miRNA (green and azure bars) from the same hairpin. Panel (b) shows the location of the moRNA sequence in relation to the miRNAs expressed by the ssc-miR-21 locus. Notably, the moRNA extends over the canonical hairpin (c), indicating that the 5' moRNA may be more probably produced from a non-canonical precursor.

and canonical hairpin end positions. All the moRNAs started exactly after the end of the miRNA with no overlap or gaps, as shown for ssc-moR-21 in Figure 29b,c. The considered moRNAs were located in a region extending from 5 to 17 nt outside the canonical hairpin and Drosha cutting site, extending over the end of the canonical hairpin. This evidence indicates that these moRNAs might be produced by a non-canonical processing of primary miRNA sequences by Drosha and that moRNAs and miRNAs are not mutually exclusive products of miRNA precursors processing by Dicer.

Novel miRNAs from new hairpins

Because the set of known swine miRNAs available in miRBase (271 precursors and 306 mature miRNAs) is considerably smaller than for human (1600 precursors and 2042 matures), we applied the mirdeep2 algorithm to our data for the discovery of unknown expressed miRNAs coming from non-annotated precursors. *MirDeep2* predicted 316 new hairpins, 253 of them with significant hairpin folding score (Table S13). As observed for the new miRNAs predicted from known hairpins, most of the new precursors were represented by few reads: 50% of the predicted hairpins had at most a read count of 155. In Table 11, we report six precursors having significant folding score and producing mature miRNAs with normalized expression estimate greater than the known miRNA median expression level. Only one of the predicted pig miRNA that are reported in Table 11 (chr16_37624) has a seed sequence corresponding to a human miRNA (hsa-let-7a-5p). Neither mouse nor human miRNAs with equivalent seed sequence were found for other mature miRNAs produced by the predicted hairpins.

Table 11. New miRNAs predicted by *mirRDeep2* with an expression above the median expression level of known miRNAs. Bold text indicates the miRNA chosen for validation..

Predicted hairpin precursor ID	Hairpin genome position and strand	Predicted mature miRNA	Read count	Sequence	miRBase miRNA with corresponding seed sequence
chr5_14516	chr5:4391289..4391351:–	***ssc-new-miR-chr5-3p***	**9433**	**augcggaaccugcggauacgg**	–
		ssc-new-miR-chr5-5p	1342	auguccgcggguucccuaucc	–
chr16_37624	chr16:5539901..5539950:–	*ssc-new-miR-chr16-3p*	6735	ugagguaguaggcugugugg	*hsa-let-7a-5p*
ch2r_6672	chr2:6684288..6684348:–	*ssc-new-miR-chr2-3p*	3923	uuuguuggcuccucugaaguga	–
chr10_25960	chr10:43502498..43502555:+	*ssc-new-miR-chr10-5p*	3561	gcgggcccacgggggcccc	–
chr12_29405	chr12:40808308..40808378:–	*ssc-new-miR-chr12-5p*	2307	ucccuggucuagugguuaggauuug	–
chr3_9002	chr3:87379785..87379841:+	*ssc-new-miR-chr3-5p*	1601	uggcguauaucacagacacagc	–

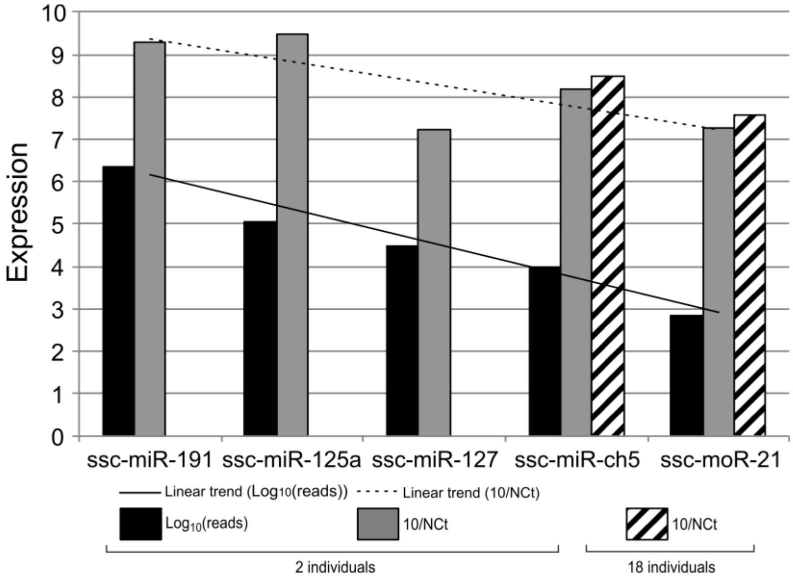


Figure 31 . qRT-PCR validation of five sRNAs, including three known miRNAs expressed at different levels in pig backfat tissue, a new miRNA predicted by mirdeep2 on chromosome 5 and the newly discovered ssc-moR-21 expressed from mir-21 locus. RNA-Seq expression levels are in base 10 logarithmic scale. qRT-PCR expression estimates are reported as the inverse of the normalized Ct (NCt = Ct(miR)/Ct(U6)), multiplied by 10, to better display the RNA-Seq and qRT-PCR results on the same scale. Black and grey bars represent respectively the RNA-Seq and qRT-PCR expression measures resulting from the experiments conducted on the same RNA (two individuals pooled). Solid and dashed lines highlight the common trend of RNA-Seq and qRT-PCR expression estimates. Textured bars show the average of qRT-PCR measures in 18 additional individuals.

qRT-PCR validation

A qRT-PCR validation was carried out for five small RNAs detected in backfat samples, including four miRNAs and ssc-moR-21-5p. Among miRNAs, we chose three known mature miRNAs (ssc-miR-191, ssc-miR-125-a and ssc-mir-137) that are expressed at different levels (2 207 436, 117 118 and 30 592 reads respectively), plus the most expressed (9433 reads) mature miRNA predicted by mirdeep2, which comes from chromosome 5, ssc-new-miR-chr5-3p. ssc-moR-21-5p was chosen because it yielded the largest expression (685 reads) among the detected moRNAs. All considered elements were validated by sequence-specific qRT-PCR carried out on the original RNA samples that were used for library preparation and deep sequencing. A good agreement observed between RNA-Seq and qRT-PCR expression estimates (Pearson correlation. −0.56; Figure 31) supports the robustness of the estimates we have reported for the small RNAs detected in this study. These results also confirm the expression of ssc-moR-21-5p, which to our knowledge, is the first validation of a moRNA

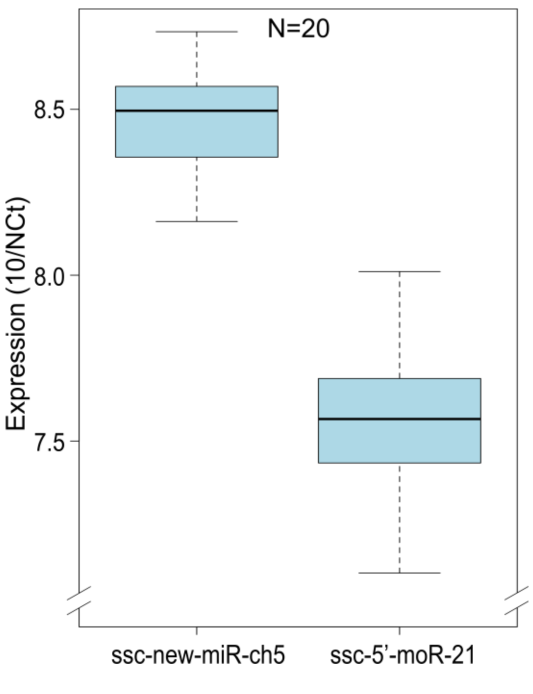


Figure 30. qRT-PCR expression measure variation, in a population of 20 normal adult individuals, regarding two newly discovered pig RNAs.

expressed in a normal tissue. Moreover, we obtained qRT-PCR data regarding two newly discovered small RNAs (ssc-new-miR-chr5-3p and ssc-moR-21-5p) in an independent group of 18 individuals providing a biological validation for these RNAs (Figure 31). Figure 30 shows as a boxplot the variation of the two considered RNAs in the considered group of 20 normal adult individuals.

3.2.2 DISCUSSION

Sus scrofa is an important species for comparative genomics, biomedical studies and also for the meat production industry. Pig adipose tissue, and backfat in particular, is one of the principal components determining the quality of dry cured hams. With the perspective of selection and genetic improvement, the understanding of molecular and genetic mechanisms regulating gene expression in porcine adipocytes is of great interest. The expression of many genes is regulated by miRNAs, which usually play a repressive role on gene expression.

In this study, we explored the miRNome of Large White pig backfat by sequencing the small RNA fraction with the Illumina technology, reaching high depth (around 65 million reads per sample). We aimed to identify known sRNAs and discover new ones that are expressed in fat cells and to obtain a quantification of sRNAs expression to facilitate further studies of gene expression regulation in pigs. Sequencing reads were analyzed with a computational pipeline that integrates several custom-developed and publicly available bioinformatics tools, which we adapted for this specific task.

As a first result, we provided the identification and quantification of 222 known pig mature miRNAs expressed in fat tissue, which account for 65% of the miRNAs annotated in miRBase. The expression level distribution is considerably skewed, with few highly expressed miRNAs: only 20% of expressed miRNAs comprise 99% of the reads. We compared our results with miRNAs reported in previous studies on porcine fat miRNome of diverse pig breeds (Chen et al., 2012; Cho et al., 2010; Li et al., 2011, 2012a). The overlap among the aforementioned studies is limited. This can be due to a specific expression profile that is peculiar and characteristic of each breed as well as to the adoption of different library preparation protocols and/or sequencing techniques. Moreover, the pattern of miRNA expression in fat tissue could depend on age and rearing conditions, including the diet supplied, as reported for pig (Cirera et al., 2010) and other mammals (Parra et al., 2010; Romao et al., 2012). However, most of the miRNAs that have been previously identified as highly expressed in adipose tissue are listed among the miRNAs we estimated as the most expressed. The four previous studies concordantly reported the high expression of let-7 family elements, and this is almost the only result they have in common. According to our data, ssc-let-7a is among the nine most expressed miRNAs, and four elements of the family are amid the 25 most expressed miRNAs. Considering the 35 miRNAs listed in at least two of the previous studies, we noticed that 21 miRNAs overlap with the 25 most expressed ones of our study and eight (ssc-miR-10b, ssc-miR-143-3p, ssc-miR-10a-5p, ssc-let-7a, ssc-miR-191, ssc-mir-148a-3p, ssc-miR-30a-5p, and ssc-miR-103) are ranked among the top nine. In addition, ssc-let-7f and ssc-miR-199b (-5p and -3p), listed by Li et al. (2011) as very abundant in swine adipose tissue and recurrently reported as highly expressed in previous studies, were among the 20% most expressed miRNAs according to our data. Similarly, ssc-miR-199a/b, ssc-miR-125a/b and ssc-miR-126-5p and -3p, which were reported by Cho et al. (2010) as more expressed in adipose than in muscle tissue, are among the most expressed miRNAs in our samples. Finally, ssc-miR-21-5p, which is the main miRNA promoting adipogenesis (Cho et al. 2010; Guo et al. 2012), was scored as highly expressed; in addition, we demonstrated that the mir-21 precursor is associated with two further sRNAs (ssc-miR-21-3p and ssc-moR-21-5p).

We reasoned that the nine most expressed miRNAs, which amount to 90% of the sequenced reads, could be considered important regulators in the tissue. All previous studies on pig miRNAs reported target prediction results obtained by a cross-species analysis (Li et al. 2011, 2012a; Chen et al. 2012). In these studies, orthologous human genes and miRNAs were used to infer the associated functional information under the assumptions that orthologous miRNAs can target orthologous genes in different species and that all the functions and interactions of gene products are conserved. Computational target prediction is affected by a relatively low specificity (Baek et al., 2008; Bisognin et al., 2012; Sales et al., 2010) due to the limited length of the sequences involved in the target recognition, to the fact that perfect complementarity between miRNA and target sequences is not needed for regulatory action, and to the context-specificity of miRNA-target interactions, which can be modulated by many additional factors. In this study, we thought it would be more accurate to obtain the predicted target set of pig, and not human, mRNAs, to provide a preliminary trace on which processes, functions and pathways are controlled by the highly expressed miRNAs. We achieved a custom target prediction by applying *miRanda* to the whole set of the 15 053 Ensembl 3'-UTR transcripts annotated for the pig genome. To our knowledge, this approach has never been performed in previous studies on pigs. A definitive assessment of the miRNA-target matches that we report would require subsequent biological validation, but this was out of the scope of our study. The species-specific target prediction reported 1889 pig transcripts, corresponding to 1687 porcine genes that are possibly regulated by at least one of the nine miRNAs. Moreover, 20% of the predicted target genes are in common between two or more of the abundantly expressed miRNAs. The functional enrichment performed on functionally annotated human orthologs of porcine genes highlighted relevant biological processes and pathways. The identified biological processes, such as positive regulation of transcription, gene expression and protein kinase cascade, indicate that many targets of these miRNAs are regulatory molecules, for instance transcription factors, which may affect the expression of a large number of genes. These results also suggest that the highly expressed miRNAs may take part in the regulation of biological processes and pathways that are associated with intracellular transport, membrane organization, protein modification processes, intracellular signaling cascade, nuclear import, cell motility, regulation of actin cytoskeleton and cellular metabolism. One interesting KEGG pathway targeted is the Wnt signaling path, which is involved in adipocyte differentiation and adipogenesis (Qin et al., 2010; Ross et al., 2000). The growth factors belonging to the family of the Wnt signaling pathway are known to regulate adult tissue maintenance and remodeling, mediating adipose cell communication. (Qin et al., 2010) proposed a role of miRNAs in suppressing or activating Wnt signaling during adipogenesis, and other authors (Chen et al., 2012; Li et al., 2011, 2012a) considered the possible interactions between the fat miRNome and the Wnt signaling at different ages and in different porcine breeds. The insulin signaling and axon guidance pathways were also enriched, as previously found by Li et al. (2011) and Guo et al. (2012b). The insulin signaling enrichment in predicted target genes of most expressed miRNAs suggests that the considered miRNAs may function as regulators in porcine adipogenesis. The concurrent enrichment of axon guidance pathways could indicate a potential relation between the nervous system and adipocyte metabolism.

The second result that arose from our analysis is the complexity of the pig backfat miRNome, both in terms of miRNA sequence variability and in terms of different small RNAs represented. We showed that most of the known miRNAs are found in fat tissue as mixtures of two to 558 isomiRs. However, considering only the isomiRs each accounting for at least the 10% of miRNA expression, we observed that over 80% of the miRNAs are represented by two to five major isomiRs, and only one-third of the overall miRNA expression is composed of the canonical sequence annotated in miRBase. As previously reported (Neilsen et al., 2012), sequence variation involving the 3’ region of the miRNA is prevalent. This is concordant also with reported observations of higher 5’ fidelity of Dicer cleavage sites and of non-template 3’ addition prevalence (Zhou et al., 2012). Nevertheless, in our study, isomiRs altered in the 5’ miRNA region, which includes the seed sequence, are present in non-negligible amounts, accounting for <20% of the expression. The biological relevance of isomiRs is supported by several observations. As indicated in the

'Introduction', isomiR mixtures can change in different cells and/or conditions. IsomiRs and canonical miRNAs are equally associated with translational machinery, and isomiRs can act as functionally cooperative partners of canonical miRNAs to co-ordinate pathways of functionally related genes (Cloonan et al., 2011). Indeed, a recent study that focused on miR-101 (Llorens et al., 2013) showed that specific functions for miR-101 and 5'-isomiR-101 are suggested by a correlation analysis on the expression profiles of miR-101 variants and predicted mRNA targets in human brains at different ages, even if the canonical miRNA and its isomiRs may target sets of genes that are highly overlapping. In this view, we think that quantitative and also qualitative changes in isomiR mixtures may play specific roles in adipose tissue biology.

Our work also reported 68 new sister miRNAs and 17 moRNAs expressed from known hairpins. Moreover, from a genome wide analysis, we predicted 253 new hairpins expressing 312 putative new miRNAs. We used miRBase as a reference in this study and assigned to newly identified miRNA sequences provisional names consistent with the public data and nomenclature. However, it is worth noting that some pig miRNAs discovered by other studies were not submitted to public databases. It can be noted that ssc-miR-4336-3p and ssc-miR-4333-3p have sequences overlapping with SNORA53 and SNORA18 of the H/ACA family of small nucleolar RNAs. Nevertheless, because of the strict filtering steps of the pipeline used, we think that is fair to include them, as well as potential similar cases, as actual miRNAs. These small RNAs could be classified in the group of miRNA-like RNAs produced by the processing of larger housekeeping RNAs, such as snoRNAs, rRNAs and tRNAs (Li et al., 2012c). More specifically, they might be sdRNAs (snoRNA-derived RNAs) that reportedly can act as canonical miRNAs, associate with argonaute proteins and influence translation (Ender et al., 2008; Falaleeva and Stamm, 2013). Similarly, clear evidence of biological activity of tRNA derived fragments was also provided, showing that they possess the functional characteristics of micro-RNAs and are able to repress mRNA transcripts in a sequence-specific manner (Maute et al., 2013).

Considering, on the whole, known and new miRNAs expressed by known hairpins, more than 50% of the expressed hairpins produce both mature miRNAs in the tissue and, in some cases, at comparable levels. About 22% of the 27 hairpins with both matures expressed over the median level do not exhibit prevalence of expression of one arm over the other. Similarly, about 25% of the newly predicted hairpins are associated with two miRNAs expressed. The detected moRNAs are in general less expressed than are miRNAs. Yet, we think they are of interest because they indicate detectable non-canonical processing of miRNA primary transcripts and/or hairpin precursors. The biological function of moRNAs is still not entirely understood. The sister miRNA pair and moRNAs produced from the same precursor have different sequences, so in principle they may play different biological roles. There are hypotheses that moRNAs may repress target transcripts like miRNAs (Umbach et al., 2010). Nevertheless, nuclear moRNA enrichment might indicate that some moRNAs play a different role specifically related to nuclear processes (Taft et al., 2010), as was shown for tiny RNAs and specific miRNAs (Zardo et al., 2012).

Finally, to assess our results, selected known and new sRNAs expressed at different levels were validated with qRT-PCR. The assay confirmed the expression of all five sRNAs considered, including three known miRNAs (ssc-miR-191, ssc-miR-125-a and ssc-miR-137), the new moRNA ssc-moR-21-5p and the new mature miRNA produced from a putative hairpin precursor in chromosome 5. The agreement observed between RNA-Seq and qRT-PCR expression measurements for five sRNAs and the biological validation in an independent group of 18 individuals support the strength of the sRNA expression estimates reported in this study.

This study permitted the identification of known and new miRNAs, isomiRs and moRNAs in pig backfat by bioinformatics analysis of RNA-seq data. We examined sequence variation and relations among reported sRNAs. Unlike previous studies, miRNA-target prediction was performed on swine transcript sequences. By means of in silico analysis, we identified processes and pathways in which the miRNAs might be involved. Finally, we assayed some of the identified sequences with qRT-PCR, also in samples from other animals. We think that this work might be a valuable source of information for future studies on gene expression regulation of pig adipose tissue.

3.2.3 DIFFERENTIALLY EXPRESSED SMALL RNAS IN ITALIAN LARGE WHITE PIG ADIPOSE TISSUE

3.2.3 MATERIALS AND METHODS

Sample collection and sequencing

Small RNA sequencing data from 18 ILW pig backfat samples were used in this study (see chapter "Pig subject selection"), which formed two equal size groups (LEAN and FAT) from pigs with extreme and divergent backfat thickness.

Computational processing of small RNA sequencing data

The study workflow is depicted in Figure 32. RNA-seq data have been processed with the *miR&moRe* pipeline (Bortoluzzi et al., 2012; Gaffo et al., 2014) and *miRDeep2* v0.0.5 (Friedländer et al., 2011). *MiR&moRe* quantifies small RNAs from RNA-seq experiments. Moreover, it detects and quantifies miRNA isoforms (isomiRs), novel miRNAs expressed in precursor hairpins that have only one annotated miRNA, and miRNA offset RNAs (moRNAs). *MiRDeep2* was exploited

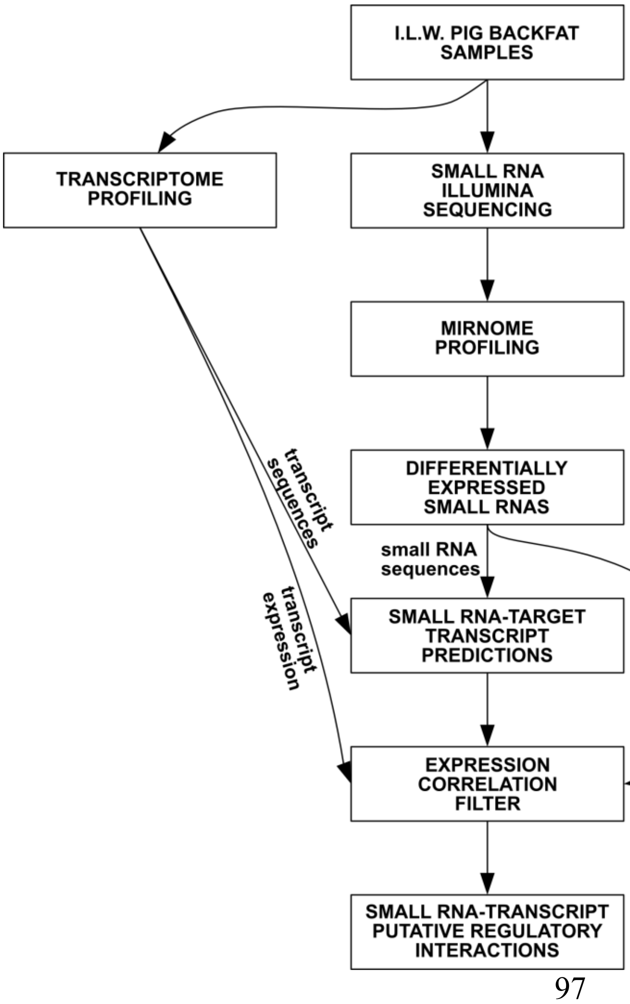

I.L.W. PIG BACKFAT SAMPLES
TRANSCRIPTOME PROFILING
SMALL RNA ILLUMINA SEQUENCING
MIRNOME PROFILING
DIFFERENTIALLY EXPRESSED SMALL RNAS
transcript sequences
small RNA sequences
transcript expression
SMALL RNA-TARGET TRANSCRIPT PREDICTIONS
small RNA expression
EXPRESSION CORRELATION FILTER
SMALL RNA-TRANSCRIPT PUTATIVE REGULATORY INTERACTIONS

Figure 32. Study workflow. The chart shows the main steps of the study. Going top to down, pig backfat samples were collected and the fraction of miRNA-like RNAs was sequenced with Illumina technology. The same set of samples was used for the sequencing of the long RNA fraction and transcriptome characterization, as described in chapter "Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness" and Zambonelli, Gaffo et al. (in press). Raw sequence data was processed with the enhanced miR&moRe pipeline to retrieve small RNAs' expression and nucleotide sequences. SRNAs differentially expressed between the LEAN and FAT groups (DEMs) were estimated with *DESeq2*. By means of *miRanda*, DEMs were further inspected by predicting their target transcripts among ILW pig backfat long RNA sequences. Expression correlation between DEMs and their targets were computed and used to refine the target predictions. Only target transcripts anti-correlated with sRNAs and passing a correlation test threshold were kept and used to reconstruct the putative regulatory relations between small RNAs and their target transcripts in Italian Large White (ILW) pig backfat tissue.

to detect precursor hairpins and related miRNAs that are not annotated in pig genome according to miRBase v.21 (Kozomara and Griffiths-Jones, 2013). *MiRDeep2* novel predictions combined with miRBase swine data were used as annotation for *miR&moRe*. Each sample raw sequenced reads were processed by *miR&moRe*. Briefly, the *miR&moRe* pipeline performs a preliminary sanitation and quality preprocessing of the input raw sequences. Reads passing the quality filter are aligned to the reference miRNA precursors and to the reference genome for expression quantification. Identification and expression quantification of isomiRs and moRNAs follow from the alignments and sequence folding predictions. Small RNAs expression levels are measured as read alignment counts in each sample. Read counts were normalized across all the samples according to the *DESeq2* (v1.4.5) (Love et al., 2014) approach. Small RNAs represented by less than ten normalized reads were excluded from further analysis. SRNA differential expression between the FAT and LEAN groups was assessed by *DESeq2*, considering FDR<0.05 (Benjamini-Hochberg adjusted P-values) as significant scores.

Small RNA target transcript prediction using backfat long RNA data

To obtain miRNA-transcript target relation predictions, we applied *miRanda* v3.3a (Enright et al., 2003). We used RNA sequences reconstructed in Zambonelli et al. (in press) as a custom target sequence database. For miRNAs and moRNAs we considered expressed isomiR sequences.

Then, we calculated Spearman correlations among the transcript profiles estimated as FPKM in chapter "Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness" and Zambonelli et al. (in press) and expression profiles of the differentially expressed small RNA in this study. According to miRNA prevalently repressive action, we focused on negative correlations giving a possible indication of direct regulation. Further, test for the significance pairwise Spearman correlation were computed. P-values were adjusted according to the Benjamini-Hochberg (BH) procedure.

3.2.3 RESULTS AND DISCUSSION

Raw data preprocessing, small RNAs identified and expression level estimates

RNA-seq resulted in 97 million reads per sample on average. The trimming and filtering pipeline steps resulted in on average 37 million reads (Table 12 and Figure 33), which were further processed for the sRNAs characterization.

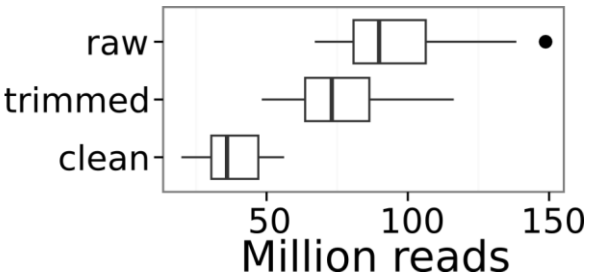


Figure 33. Total number of reads for each main data filtering step.

Table 12. Total number of reads for each main data filtering step, for each sample

Sample ID	Raw reads	Trimmed reads	Clean reads
S473	78.538.121	57.537.862	32.756.840
S474	67.076.458	48.302.409	22.902.760
S475	93.130.433	63.482.435	34.256.027
S476	105.628.078	74.020.736	40.358.547
S477	148.774.413	116.264.009	37.734.714
S478	78.686.470	67.649.827	30.064.360
S479	88.683.293	72.143.266	34.034.251
S480	106.585.223	82.930.354	50.618.318
S481	138.414.161	97.359.514	56.203.477
S482	103.315.208	87.526.775	38.340.337
S483	90.969.362	79.591.890	45.220.470
S484	85.181.674	63.205.498	47.681.714
S485	114.251.255	102.793.579	55.799.228
S486	125.087.146	97.464.630	49.205.100
S487	88.464.121	77.525.963	19.836.034
S488	77.927.270	64.168.164	31.575.982
S489	79.298.639	61.850.652	20.778.522
S490	87.328.141	67.389.681	22.256.476

MiRDeep2 v0.0.5 (Friedländer et al., 2011) was run with default parameters, input susScr3 genome from the UCSC database, miRBase v21 miRNA annotation, and as input sequences the pool of the 18 samples' read sets and the two samples from chapter “miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs”and (Gaffo et al., 2014). *MiRDeep2* predicted 1,340 new miRNA precursors (NP) in pig genome, for a total of 2,680 novel mature miRNAs from new precursors (NPmiRNAs) that do not overlap to and have different sequence from pig miRNA precursors reported in miRBase. Minimum *miRDeep2* scores were zero, first quartile 1.4, median 3.0, mean 1,569.2 and maximum 1,205,860.9. New precursors with alerts for rRNA/tRNA (Rfam sequences were

provided in the miRDeep2 package) were only five. *MiRDeep2* calculated probability of being a real miRNA was 85% at maximum, with mean 40% and median 5%. Novel precursors' genomic coordinates and relative miRNAs positions in the precursors were retrieved from the results providing annotation for the newly predicted precursor hairpins and miRNAs. To obtain homogenous data for known and new miRNAs, annotation of predicted new small RNAs, as long as their sequences, were combined with the miRBase pig annotation used for the *miRDeep2* run and used as input for the *miR&moRe* pipeline. *MiR&moRe* received as input the same reference genome used with *miRDeep2*. *MiR&moRe* detected the expression of 442 NPmiRNAs (16.5%). We applied conservative filter for new precursors discarding hairpins with miRDeep2 score smaller than 1, with probability of miRNA prediction smaller than 60%, with Rfam alert and less than 50 nt long (as done by Friedländer et al. (2014) and Londin et al. (2015) (Friedländer et al., 2014; Londin et al., 2015). With these settings, 224 NPmiRNAs were discarded, resulting in 218 NPmiRNAs that we considered more reliable and that were taken in to account for further analysis. All the detected miRNAs were required to be represented by at least ten normalized reads on average within either the LEAN or FAT groups. Additional 121 NPmiRNAs were removed, resulting in 103 NPmiRNAs (3.8% of the initial NPmiRNAs) considered as expressed. After read count quantification and normalization, sRNAs with group mean expression lower than ten normalized reads were discarded. Overall, we detected 426 sRNAs expressed, including 231 known miRs, 69 new miRs from known precursors, 103 NPmiRNAs, and 23 moRNAs. *Ssc-miR-10b*, *ssc-miR-143-3p* and *ssc-148a-3p* together account 52% of the total expression, with the first two composing respectively 33% and 13% of the expression. Only 24 known miRNAs (5.6% of the expressed sRNAs) account for the 90% of the total expression. Notably, only 21% of expressed sRNAs account the 99% of the total expression (Figure 34). These findings are consistent with our previous work carried out on a smaller sample size of ILW pig backfat (Gaffo et al., 2014 and chapter "miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs", with 214 known miRNAs commonly identified by the two works.

Differentially expressed sRNAs

Comparing the expression profiles between the LEAN and FAT samples, we found 31 differentially expressed sRNAs (DEMs) (Table 13): 18 known miRNAs, 6 new sister miRNAs, 6 miRNAs from new precursors and ssc-moR-21-5p that was validated in previous work. DEM absolute fold changes ranged between -0.97 (ssc-miR-362-3p) and 0.97 (ssc-miR-146b), with expression levels ranging from 52.8 (ssc-miR-362-3p) to 243,835.8 (ssc-miR-199a-3p) mean read count.

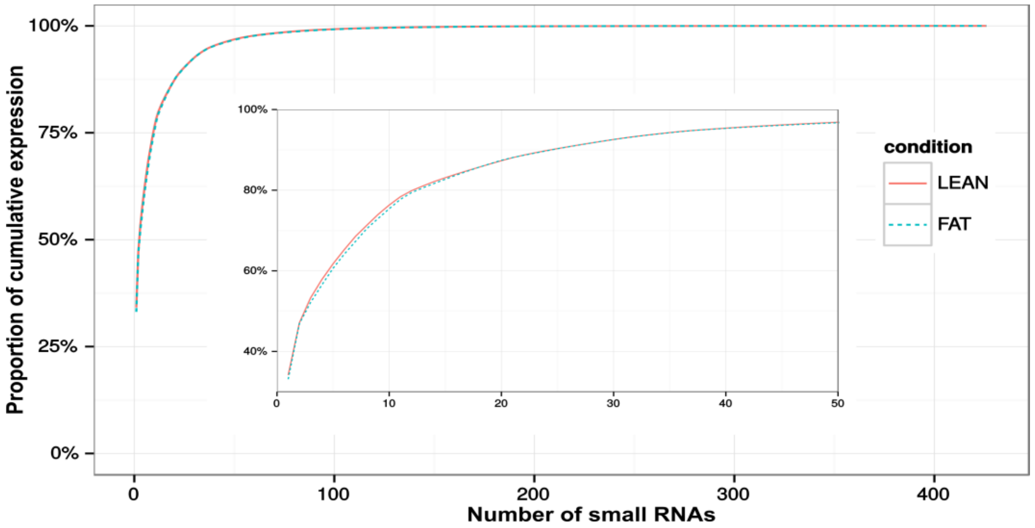


Figure 34. Cumulative sRNA average expression in LEAN and FAT groups

Table 13. List of differentially expressed small RNAs. For each small RNA the table indicates expression estimate, log2 fold change in sample comparison, small RNA group and predicted sequence. The table is ordered according to fold change expression.

sRNA	Mean reads	Log$_2$ (FAT / LEAN)	Expr. in FAT	FDR	sRNA group	sRNA mature sequence
ssc-miR-146b	20131,80	0,97	up	0,01	Known	ugagaacugaauuccauaggc
ssc-miR-365-5p	799,48	0,89	up	0,00	Known	gagggacuuucaggggcagcugu
chr9_40038-5p	21,81	0,83	up	0,01	NPmiRNA	cuccuggcuggcucgcca
ssc-miR-221-5p	2916,44	0,77	up	0,01	Known	accuggcauacaauguagauuucugu
ssc-moR-21-5p	9,63	0,76	up	0,01	moRNA	ctccatggctgtaccaccttgtcgg
ssc-miR-222	2863,37	0,75	up	0,01	Known	agcuacaucuggcuacugggucuc
chr5_31322-3p	27,77	0,72	up	0,01	NPmiRNA	ucugagaugugaccugggcau
JH118494-1_44794-3p	11,40	0,72	up	0,01	NPmiRNA	ucugagaugugaccugggcau
chr7_37486-3p	24,49	0,69	up	0,02	NPmiRNA	aagucccaucugggucgcc
chr3_26283-5p	30,91	0,69	up	0,00	NPmiRNA	uuggcucugcgaggucggcuca
ssc-miR-128-1-5p	138,40	0,53	up	0,02	New sister	cggggccgtagcactgtctgag
ssc-miR-132	55877,75	0,50	up	0,03	Known	uaacagucuacagccauggucg
ssc-miR-1306-5p	70505,29	0,47	up	0,02	Known	ccaccuccccugcaaacgucca
GL896302-1_48155-5p	12,42	0,39	up	0,02	NPmiRNA	ucucugggccugugucuuaggcu
ssc-let-7c	2435309,05	-0,32	down	0,01	Known	ugagguaguagguuguaugguu
ssc-miR-130a	56716,38	-0,38	down	0,02	Known	cagugcaauguuaaaagggcau
ssc-miR-181c	9716,73	-0,39	down	0,01	Known	aacauucaaccugucggugagu
ssc-miR-214-5p	93,15	-0,42	down	0,00	New sister	tgcctgtctacacttgctgtgc
ssc-miR-199a-5p	5221,81	-0,44	down	0,00	Known	cccaguguucagacuaccuguuc
ssc-miR-199a-3p	5352,48	-0,45	down	0,01	Known	acaguagucugcacauugguua
ssc-miR-199b-3p	5118,61	-0,45	down	0,01	Known	uacaguagucugcacauugguu
ssc-miR-99a-3p	56,14	-0,47	down	0,00	New sister	caagctcgcttctatgggtctg
ssc-miR-195	6255,99	-0,50	down	0,02	Known	uagcagcacagaaauauuggc
ssc-miR-101-2-5p	152,31	-0,61	down	0,02	New sister	tcagttatcacagtgctgatgct
ssc-miR-193a-3p	6485,75	-0,62	down	0,02	Known	aacuggccuacaaagucccagu
ssc-miR-136-3p	133,91	-0,63	down	0,02	New sister	catcatcgtctcaaatgagtct
ssc-miR-335	1118,26	-0,75	down	0,02	Known	ucaagagcaauaacgaaaaaug
ssc-miR-133a-3p	52301,03	-0,79	down	0,03	Known	uuggucccccuucaaccagcug
ssc-miR-545-3p	381,19	-0,81	down	0,04	Known	aucaacaaacauuuauugugug
ssc-miR-136	32337,35	-0,94	down	0,01	Known	acuccauuuguuuugaugaugga
ssc-miR-362-3p	76,01	-0,97	down	0,01	New sister	aacacacctattcaaggattc

Small RNAs up- and down-regulated in FAT were 14 and 17 respectively. SRNAs Up-regulated in FAT were mainly new elements (6 from new precursor, plus ssc-moR-21-5p and ssc-miR-128-1-5p), while sRNAs down-regulated in FAT were all from already annotated precursors: 12 known miRNAs and 5 new sister miRNAs. From Figure 35 we see how this set of DEMs discriminates the LEAN and FAT samples using unsupervised clustering. For twelve DEMs among the 25 deriving from known hairpins (let-7c, miR-99a-3p, miR-130a, miR-132, miR-146b, miR-181c, miR-193a-3p, miR-199a-5p, miR-221-5p, miR-222, miR-335, and moR-21-5p) have orthologs and member of the same miRNA family associated to adipose tissue, adipogenesis and obesity from studies in humans and mice (Arner and Kulyté, 2015; Chen et al., 2013; Hilton et al., 2013; McGregor and Choi, 2011).

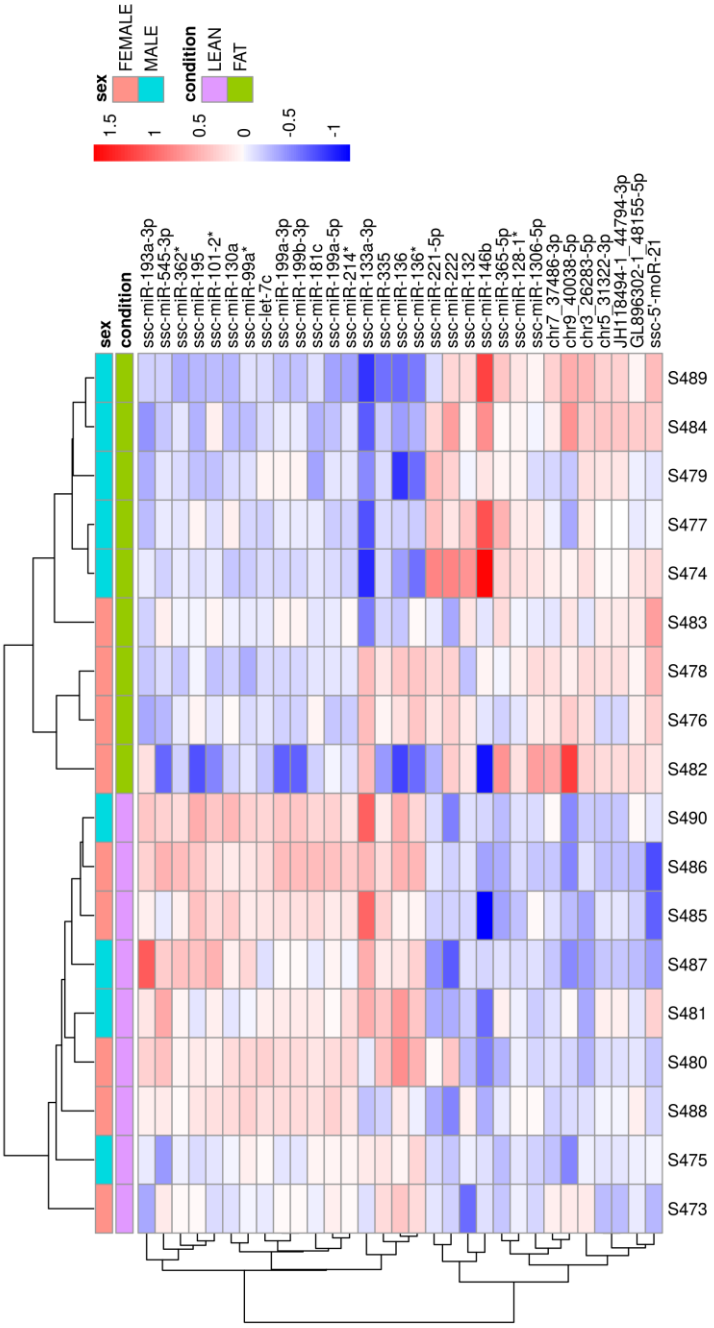


Figure 35. Heatmap of DEMs. Rows of the heatmap represent the 31 DEMs, columns correspond to samples. The heatmap cells are coloured according to the deviance of the sRNA expression in the sample from the average expression of the sRNA. Red cells represent sRNAs expressed more than their mean expression across all samples (white); blue cells represent sRNAs expressed less than their mean expression. Color intensity is proportional to the difference from the mean, in the regularized logaritmic scale. Notably, the clustering of samples, using Pearson correlation as distance measure and complete linkage, shows a perfect separation of the LEAN and FAT samples. Sex of the samples seems not to have influence on the clustering, instead.

We validated by qRT-PCR the differential expression of 9 DEMs, including 4 known, 4 new sisters and ssc-moR-21-5p. As shown in Figure 36 $\log_2$ fold changes agree between RNA-seq and qRT-PCR expression estimates (0.88 Pearson correlation; for technical details on validation

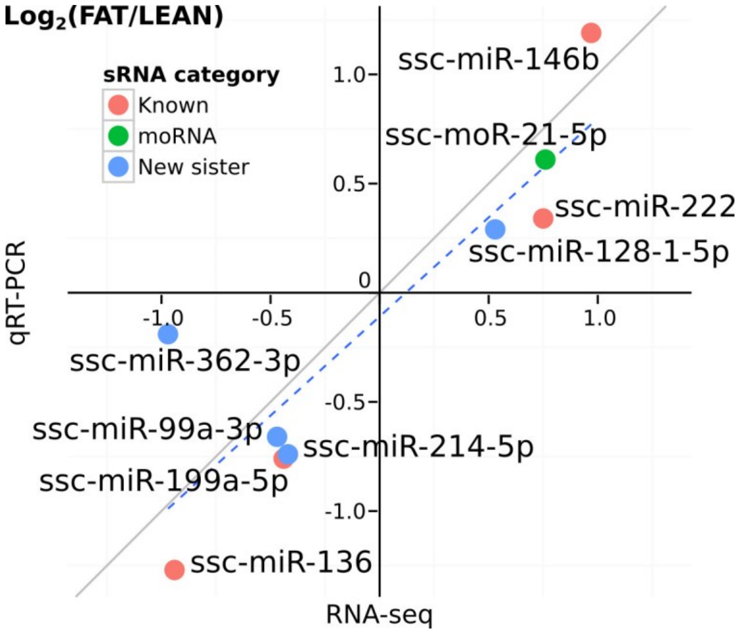


Figure 36. Correlation of $\log_2$ fold change values between RNA-seq and qRT-PCR measures in nine small RNAs. Gray line is the diagonal, blue dashed line represent the fit line calculated according to the points. From the plot we can observe good correlation (Pearson 0.88) between the two method measures.

procedure referer to Gaffo et al. (2014)). From we observe that five sRNAs have a significant (P-value <0.05) statistical test, including ssc-miR-99a-3p and ssc-miR-214-5p new sister miRNAs. According to the validation assay, ssc-moR-21-5p has a less stringent significance (<0.1).

Table 14. $\log_2$ fold changes of FAT/LEAN average expression values (in 18 backfat samples) measured with RNA-seq and qRT-PCR (Δ ΔCt). T-test is reported for qRT-PCR assays.

Mature sRNA	RNA-seq	qRT-PCR	T-test P-value	sRNA category
ssc-miR-136	-0,94	-1,27	0,00	Known
ssc-miR-199a-5p	-0,44	-0,76	0,00	Known
ssc-miR-99a-3p	-0,47	-0,66	0,02	New sister
ssc-miR-214-5p	-0,42	-0,74	0,02	New sister
ssc-miR-146b	0,97	1,19	0,04	Known
ssc-moR-21-5p	0,76	0,61	0,07	moRNA
ssc-miR-222	0,75	0,34	0,21	Known
ssc-miR-362-3p	-0,97	-0,19	0,46	New sister
ssc-miR-128-1-5p	0,53	0,29	0,80	New sister

DEM isomiR composition

IsomiRs were investigated for 24 miRNAs (18 known and 6 novel-precursor miRNAs). We considered isomiRs composing at least 10% of the corresponding miRNA expression and discarding very rare isoforms. We detected 59 distinct isomiRs, specifically 58 in LEAN and 55 in FAT (54 in common). Four isomiRs were specific for LEAN samples and one was specific for FAT (Table 15). Three miRNAs (*chr7_37486-3p*, *ssc-miR-136* and *ssc-miR-193a-3p*) are represented by only one major isoform accounting for more than half of the miR expression. The number of isomiRs composing each DEM expression is at maximum four, detected in three cases (*chr5_31322-3p*, *JH118494-1_44794-3p* in both FAT and LEAN samples, *chr9_40038-5p* only in FAT samples), and 2.5 on average (Figure 37 and Figure 38). According to *miR&moRe* classification, we distinguished four types of sRNA isoforms: “exact”, “shorter or longer”, “one-mismatch”, “two-mismatches”.

Table 15. LEAN and FAT group specific isomiRs

sRNA	Sequence	Variation type	Variation location	Group specificity
chr3_26283-5p	TTGGCTCTGCGAGGTCGGCTC	shorter_or_longer	3p	LEAN
ssc-miR-181c	AACATTCAACCTGTCGGTGAG	shorter_or_longer	3p	LEAN
ssc-miR-222	AGCTACATCTGGCTACTGGGTCTCT	shorter_or_longer	3p	LEAN
ssc-miR-365-5p	AGGGACTTTCAGGGGCAGCTGTGT	shorter_or_longer	both	LEAN
chr9_40038-5p	TCCTGGCTGGCTCGCC	shorter_or_longer	both	FAT

Most isoforms are variations on the length (“shorter_or_longer” category are 32 out of 58 in LEAN and 29 out of 55 in FAT), followed by the canonical (“exact”) variant (18 in both groups) and eight one-mismatch variation isomiRs. The two-mismatch variation is not found in DEM's isomiRs that we analyzed. Only less than half of DEMs (10 cases in LEAN and 11 in FAT) express the canonical isomiR as major form. Conversely, “shorter or longer” isoforms compose the largest part of the expression in 12 LEAN and 11 FAT cases. Notably, in 6 sRNAs do not express the exact isoform enough to reach 10% of the total miRNA expression (Figure 37 and Figure 38). These findings were consistent with previous results (Gaffo et al., 2014) showing that in pig backfat the canonical miRNA isoform is not always the most expressed for the miRNA. The variation observed could be either the result of genetic difference between the reference annotation and the population observed, or post-transcriptional editing of the miRNA sequencing. Nevertheless, analysis of isomiR expression composition revealed interesting patterns, particularly for the novel predicted precursor (Figure 38), that could be further studied. For instance, differential expression could be assessed at the isomiR level, investigating whether the contribution of non-canonical isomiRs is determinant of the observed expression variation.

Putative sRNA-transcript regulatory interactions and QTL enrichment

To predict miRNA target transcripts, we considered for each miRNA the isomiR sequences representing at least 10% of the expression of the 18 known miRNAs and the 6 NPmiRNAs that were differentially expressed. Regarding the 6 new sister miRNAs and ssc-moR-21-5p, we had only unique sequences because the miR&moRe pipeline does not perform isomiR analysis for new sister miRNAs and moRNAs. We selected the isomiRs that accounted at least 10% of the respective miR expression, resulting in 59 isomiRs. Thus, we obtained 66 sRNA sequences in total as input to target prediction.

MiRanda was run with default parameters on the full set of 63,418 transcript sequences expressed in the samples that are reported in chapter “Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness” and in Zambonelli et al. (in press).

The backfat transcript targets predicted for the 66 isomiRs were 50,161, for a total of 2,680,017 isomiR-mRNA target relations predicted by *miRanda*, corresponding to 699,493 sRNA-transcript putative relations. Spearman correlations between miRNA and transcript expression profiles ranged from -0.9 to +0.9. However, only predicted relations with correlations < -0.4 and FDR <

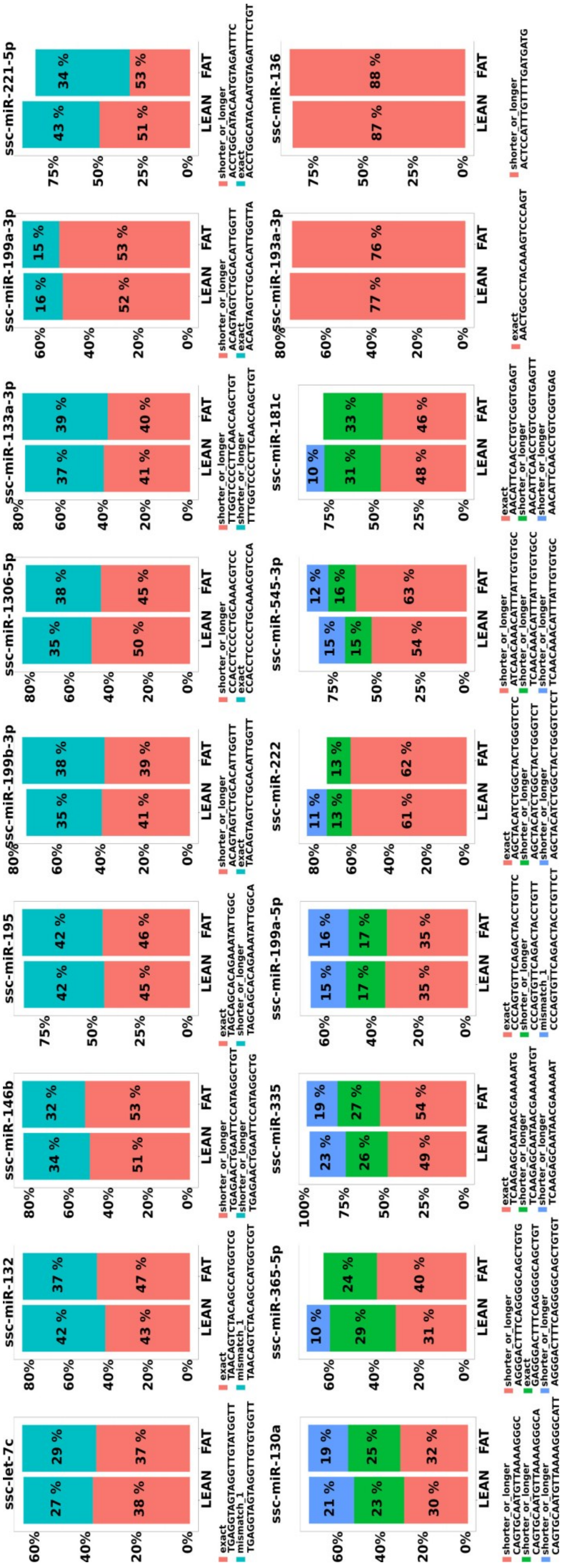


Figure 37. Known miRNAs’ expression isomiR composition. Each plot is associated to a specific legend, color code and maximum percentage shown. IsomiRs are ordered by expression proportion. Only isomiRs >10% are shown: the portion of the bars not reaching 100% represents mixture of weakly expressed isomiRs.

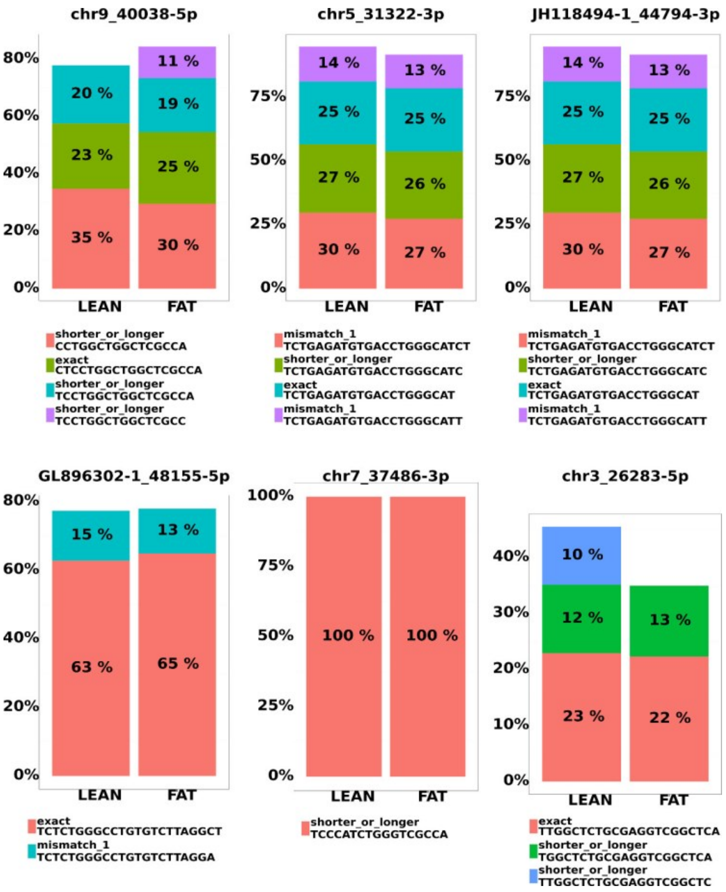


Figure 38. New precursor miRNAs' isomiR expression composition. Each plot is associated to a specific legend, color code and maximum percentage shown. IsomiRs are ordered by expression proportion. Only isomiRs >10% are shown: the portion of the bars not reaching 100% represents mixture of weakly expressed isomiRs.

10% were kept, obtaining 56,683 strongly negatively-correlated sRNA-transcript putative relations (48,259 sRNA-gene relations), which involved 22,362 transcripts (12,373 genes) and all the DEMs.

The largest number of targets among known miRs is given by *ssc-miR-365-5p* (3,095 different transcripts could be targeted, corresponding to 2,624 genes); while in absolute *chr3_26283-5p* has the largest number of target transcripts (3,383; 2,878 genes). *ssc-miR-136-3p* has the smallest number of targets (524; 465 genes), while among the new miRs, *chr7_37486-3p* has the minimum number of targets (1,011; 862 genes).

The group of the 86 differentially expressed transcripts (DETs) reported in Zambonelli et al. (in press), is represented by 40 transcripts (and genes) targeted by 30 DEMs, forming a total of 193 putative relations. The predicted relations are represented as a network in Figure 39.

Then, we considered only the transcripts showing a "sizeable" variation in expression between the LEAN and FAT groups and focused on the DETs within 30% FDR from differential expression analysis in Zambonelli et al. (in press) (from now on, this set will be referred to as extended-DETs; eDETs). SRNA-transcript target relations were 830, involving 197 transcripts (195 genes) and all 31 DEMs. *Chr3_26283-5p* targets 48 transcripts and *ssc-miR-365-5p* targets 39 transcripts, genes are the same number. *Ssc-miR-136-3p* targets 11 transcripts (and genes), *chr7_37486-3p* targets 15 transcripts from different genes.

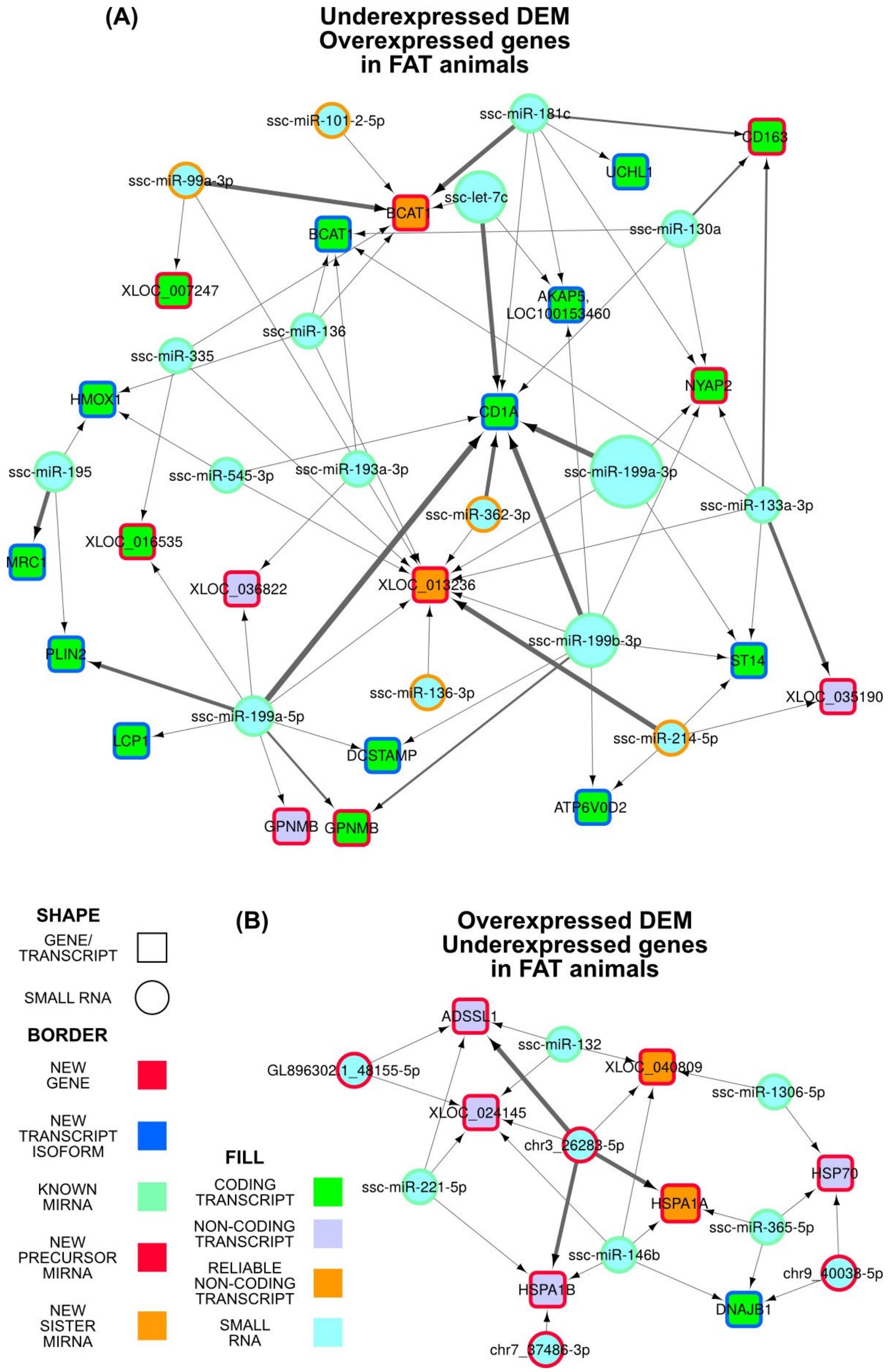


Figure 39. Common DEG DEMs' predicted regulatory relations with transcripts showing expression variation between LEAN and FAT groups.

In addition, we reasoned that DEMs and their putative target gene could be associated to relevant pig backfat quantitative trait loci (QTL). We tested the enrichment in pig QTL for the eDETs putatively targeted by the DEMs. For each QTL, Enrichment was computed with an upper-tailed hypergeometric test for over-representation, with P-values corresponding to *P[X > x-1]*. In other words, we give the probability to randomly select more than *x-1* genes falling in a QTL, when sampling 195 genes from the genome. To compute the test, we considered the whole pig genome as background, merging the genes annotated in Ensembl (v 10.2.80 in date 14 July 2015) with the new genes discovered from the RNA-seq experiment in Zambonelli et al. (in press). We obtained a total of 34,617 genes for the pig genome. We compared to the pig QTL annotation downloaded from the PigQTL database (http://www.animalgenome.org/cgi-bin/QTLdb/SS/download?file=gbpSS_10.2, in date 14 July 2015) to the gene annotation and counted the number of genes associated to each QTL, which were grouped by category. The same procedure was carried out to count the number of eDETs for each QTL and QT. By the same strategy, we retrieved the sRNAs and related target genes that come from a same QTL.

Table 16 shows the enrichment in QTL with P-values smaller than 0.05. Many traits directly associated with backfat are present (10 out of 52) among them most enriched, including “average backfat thickness”. Other interesting traits are enriched, some associated to meat products quality, such as “ham fat thickness” and “linoleic acid content”; and to appetite regulation, such as “leptin level”.

Table 16. DEM putative target genes enrichment in pig QTL. Only traits with P-value < 0.05 are shown. Rows are ordered according to decreasing fold enrichment.

Trait	Selected genes in QTL	total genes in QTL	P-value	FDR	Expected selected genes	Fold enrichment
Backfat at tenth rib (16 weeks)	2	6	0,008	0,140	0,0	59,2
Backfat linear at last rib	2	8	0,015	0,175	0,0	44,4
Phagocytic activity	2	8	0,015	0,175	0,0	44,4
Subjective abnormal flavor in fat	3	15	0,007	0,140	0,1	35,5
Backfat at last rib (13 weeks)	2	10	0,024	0,200	0,1	35,5
Skatole, sensory panel	2	10	0,024	0,200	0,1	35,5
Protein accretion rate	5	26	0,001	0,070	0,1	34,1
Immunoglobulin G level	4	21	0,003	0,081	0,1	33,8
Mycoplasmal pneumonia susceptibility	4	21	0,003	0,081	0,1	33,8
Total body fat tissue (22 weeks of age)	3	16	0,009	0,140	0,1	33,3
Half carcass weight	3	17	0,010	0,140	0,1	31,3
Backfat at tenth rib (13 weeks)	2	12	0,034	0,219	0,1	29,6
Leptin level	2	13	0,040	0,243	0,1	27,3
pH 40 minutes post mortem (ham)	3	20	0,016	0,180	0,1	26,6
Meat color-b	4	29	0,009	0,140	0,2	24,5
Carcass temperature (45 minutes post-mortem)	6	44	0,002	0,080	0,2	24,2
Chew score	8	61	0,000	0,070	0,3	23,3
Red cell distribution width	8	64	0,001	0,070	0,4	22,2
Feed intake (35-55 kg)	4	32	0,012	0,161	0,2	22,2

Trait	Selected genes in QTL	total genes in QTL	P-value	FDR	Expected selected genes	Fold enrichment
Backfat linear at tenth rib	3	24	0,027	0,200	0,1	22,2
Carcass temperature (24 hr post-mortem)	3	24	0,027	0,200	0,1	22,2
Diameter of type I muscle fibers	8	70	0,001	0,070	0,4	20,3
pH 96 hr post-mortem (loin)	4	37	0,021	0,200	0,2	19,2
Backfat (22 weeks of age)	5	48	0,013	0,161	0,3	18,5
Cooling loss	4	40	0,027	0,200	0,2	17,8
Toll-like receptor 2 level	5	52	0,018	0,185	0,3	17,1
NADP-malate dehydrogenase activity	4	45	0,040	0,243	0,3	15,8
CD4-positive leukocyte percentage	4	46	0,043	0,250	0,3	15,4
NADPH-generating enzyme activity	5	58	0,028	0,200	0,3	15,3
Total muscle fiber number	5	58	0,028	0,200	0,3	15,3
Average glycolytic potential	4	47	0,046	0,258	0,3	15,1
Left teat number	4	47	0,046	0,258	0,3	15,1
Ham percentage	6	72	0,020	0,200	0,4	14,8
Body weight (10 weeks)	9	120	0,010	0,140	0,7	13,3
Ear erectness	12	161	0,003	0,081	0,9	13,2
Ham fat thickness	12	162	0,003	0,081	0,9	13,1
Backfat (17 weeks of age)	6	82	0,037	0,230	0,5	13,0
Off-Flavor Score	7	96	0,026	0,200	0,5	12,9
Backfat (40 kg live weight)	7	97	0,027	0,200	0,5	12,8
Backfat (60 kg live weight)	7	97	0,027	0,200	0,5	12,8
Shoulder weight	11	157	0,008	0,140	0,9	12,4
Mean corpuscular hemoglobin content	7	100	0,032	0,212	0,6	12,4
Hemoglobin	13	196	0,006	0,135	1,1	11,8
Trimmed wholesale product / live weight	7	106	0,042	0,250	0,6	11,7
Linoleic acid content	9	144	0,032	0,212	0,8	11,1
Meat color-L	13	208	0,010	0,140	1,2	11,1
Ham weight	21	366	0,001	0,070	2,1	10,2
pH 24 hr post-mortem (loin)	17	300	0,007	0,140	1,7	10,1
Loin muscle depth	13	237	0,029	0,200	1,3	9,7
Loin muscle area	24	475	0,003	0,081	2,7	9,0
Carcass length	19	382	0,018	0,185	2,2	8,8
Average backfat thickness	22	486	0,029	0,200	2,7	8,0

3.2.3 CONCLUSIONS

In this study, we characterized the miRNome of Italian Large White pig backfat on a larger set of samples and with updated annotation with respect to our previous work (chapter "miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs" and (Gaffo et al., 2014)). As expected, because the samples of the two studies were taken from the same population of pigs with similar criteria and in the same conditions, we observed large overlap between the results, in terms of detected sRNAs, expression level distribution and sequence variation patterns. In processing the data, we were conservative and used stringent filtering criteria both for input read quality and novel precursor selection that greatly reduced the initial amount of input reads and novel findings.

Comparing the expression profiles between FAT and LEAN samples, we identified a set of 31 significantly differentially expressed small RNAs (DEMs), which well separated the two groups in unsupervised cluster analysis. Overexpressed and underexpressed DEMs were approximately equally numerous (14 vs. 17). Overexpressed DEMs included many novel sequences, while underexpressed elements derived all from already annotated precursors. Nearly half of DEMs were reported in other studies on humans and mice to be involved in important pathways related to adipose tissue and to play important roles in adipogenesis, adipose tissue homeostasis, and obesity. In addition, by custom target predictions combined with transcriptome expression profile correlation, we obtained miRNA-transcript putative regulatory interactions occurring in the tissue. We considered the relation network that comprised a set of 85 genes differentially expressed between the FAT and LEAN groups, as reported in our previous work (chapter "Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness" and Zambonelli et al. (in press)), which we are currently studying in details. Despite a small set of selected interactions can provide a starting point for further investigation, we reasoned that miRNAs could impact on the expression variation also for transcripts with less significant difference of expression, yet inducing meaningful cumulative effects. For this reason, we considered a larger set of transcripts modulated between FAT and LEAN groups and targeted by DEMs; many of them were enriched in backfat- and meat quality- related pig QTL, further supporting their putative involvement in backfat deposition and fat traits. Additional analyses considering involvement of putative target transcripts in specific pathways important for adipose tissue functions will help us to study the impact to regulatory pathways of specific or groups of miRNAs.

4 CONCLUSION AND PERSPECTIVES

Within the last decade we have seen a dramatic increase in the use of RNA-seq for transcriptomic experiments. RNA-seq technologies speeded up the process of sequencing cell transcripts and promise future improvements in accuracy, running time, and requirements. With respect to previous sequencing technologies, there has been an inversion in the timings of experiment processes. High-throughput sequencing produces large amount of data in a relatively short time. Currently, bioinformatics analysis is the new bottleneck in the process toward informative results (Funari and Canosa, 2014). Difficulties emerged first in the development of methods and software able to process big amounts of data; and single software tools were improved to perform better both in accuracy and computational resources requirements. The use of integrative frameworks that automate the execution of sequential tasks and parallelize the execution of independent steps can reduce computation time and human errors, but also enhance data analysis reproducibility (Nekrutenko and Taylor, 2012). Computational genomic research is evolving on these aspects, like The Galaxy Project (Goecks et al., 2010), an open web-based platform, which aims at supporting accessibility of complex computational resources, reproducibility, and communication of the results within an unified environment. Nevertheless, the diversity of experimental design and the wide range of experiment type that are possible with RNA-seq technologies hindered the formulation of a commonly adopted framework, and standard procedures have not yet emerged. Thus, this thesis produced both methodological and applicative results.

I developed an automated, modular, and parallelized computational pipeline for the characterization of transcriptomes from RNA-seq data by selecting, implementing, and combining the various software tools that perform the pipeline steps. The application of the pipeline to sequencing data of 20 pig adipose tissue samples resulted in the characterization of the sequences and abundance estimation of the transcripts expressed in the samples, including more than 15,000 putative novel transcripts from non-annotated pig genomic regions and more than 34,000 novel isoforms of known genes, including important genes involved in adipose tissue functions, such as *PLIN2*. Novel transcripts were characterized by comparative sequence analysis and coding potential prediction, providing annotation for large part of these new findings that will be further investigated, especially to experimentally validate their expression and assess their structure. Moreover, the comparison of expression profiles between two groups of individuals that showed extreme and divergent phenotypes for backfat thickness allowed the identification of differentially expressed genes and transcripts. Functional analysis of gene and transcripts differentially expressed in fattest animals showed a decreased expression of heat shock proteins and an increased expression of many genes modulating fat physiological processes or related to inflammatory status and immune response. For instance, an increase in the expression of INHBB and SPP1 linked to cytokine production and the higher expression of ENPP1 and PIK3AP1 may indicate a status of insulin resistance, one of the typical signals connected with obesity. Our results agree with recent studies showing that increased adiposity and impaired stress response may activate inflammation. Several immune system and anti-inflammatory processes are activated and play a critical role in the response to fat accumulation in porcine backfat tissue. High fat accumulation in adipose tissue of pigs can determine the development of an inflammatory process producing a cascade of defense and adaptive reactions in the tissue, such as activation of the immune system and mesenchymal cells differentiation in adipocytes. A deeper knowledge of the metabolic processes involved in fat deposition can be very important in developing the use of pig as a model species to study obesity and related disorders for humans.

The methods presented here are currently undergoing further development and extensions, and have applications well over and above those presented in this thesis. Because of experimental design, the transcriptome described in this manuscript referred only to long polyadenylated transcripts and did not consider other RNA species that lack poly(A) tails, such as many lncRNAs and circRNAs. Differently from the traditional poly(A)+ enriched RNA-seq libraries, which was used in our experiment, appropriate library preparation strategies are required to investigate the poly(A)- fraction. The use of ribosomal RNA-depletion protocols in RNA-seq experiments can increase discovery power and provide data suitable for the expression profiling of both poly(A)+ and poly(A)- transcripts. The combined use of these protocols with new computational methods in downstream analysis, recently allowed the investigation of novel classes of RNAs, for instance circRNAs (Jeck and Sharpless, 2014). The pipeline for circRNAs detection presented in this thesis represents an initial approach to improve transcriptome characterization, and is currently under development with the perspective of its application to a study on circRNAs in hematopoiesis. In particular, we are combining the quantification of linear transcripts extending also to poly(A)- sequences; and focusing on the comparison of expression proportion between poly(A)+ and poly(A)- transcripts, and between the proportion of circular to linear expression estimates. This approach will allow us to have a more complete profiling of the transcriptome under study, and by the integration with miRNome profiling, could contribute to the study of gene expression regulation.

Regarding small RNAs, we improved the *miR&moRe* pipeline for the analysis of small RNAs from RNA-seq data. We provided the possibility of application of the pipeline also to non-human data and enhanced the discovery power by allowing the identification of novel miRNA precursors. The small RNA fraction of the same set of pig backfat samples used for transcriptome reconstruction was sequenced to characterize the miRNome of pig adipose tissue and to investigate its underlying regulatory mechanisms. We applied our computational method for small RNA data using used stringent quality parameters and identified more than 400 expressed elements, which corroborated preliminary results from the analysis on two samples. Nearly half of the detected sequences were unknown small RNAs, including new miRNAs from known precursor, new miRNA precursors, new isomiRs, and many moRNAs. Experimental validation confirmed our expression estimates, especially for some novel small RNAs such as *ssc-moR-21-5p*. Comparison between sample groups identified 31 significantly differentially expressed small RNAs for which further examination about sequence variations was carried out. This highlighted non-canonical expression patterns regarding miRNAs' isomiR composition, which deserve additional investigation. Computational prediction tools of miRNA targets are usually centered on few model organism, mainly human and mouse. To investigate the putative small RNA-transcript regulatory interactions in pig adipose tissue, we set up custom methods exploiting the matched comparison of the profiled transcriptome and miRNome. The predicted relations and the corresponding network representation reported already provide a starting point to explore the complex regulatory mechanisms underlying pig adipose tissue biology. Additional investigation will consider synergistic action of miRNA expression impacting on post-transcriptional expression regulation of genes involved in pathways related to adipose tissue development and maintenance, whose coordinated expression modulation might have a biological significance beyond the differential expression at level of single gene.

In summary, the computational pipelines developed in this thesis allowed effective analysis of a complex RNA-seq experiment; yet, our methods are undergoing further improvements, and could be used also for other studies. Finally, the applicative results of this thesis enlarged the knowledge of transcripts and small RNAs expressed in the pig adipose tissue, as well as small RNA-transcripts regulatory interactions, providing information helpful for a better understanding of ILW pig backfat and future studies on gene expression regulation in this tissue.

REFERENCES

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*, e05005.

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.

Amano, S.U., Cohen, J.L., Vangala, P., Tencerova, M., Nicoloro, S.M., Yawe, J.C., Shen, Y., Czech, M.P., and Aouadi, M. (2014). Local proliferation of macrophages contributes to obesity-associated adipose tissue inflammation. Cell Metab. *19*, 162–171.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arner, P., and Kulyté, A. (2015). MicroRNA regulatory networks in human adipose tissue and obesity. Nat. Rev. Endocrinol. *11*, 276–288.

Arner, P., and Langin, D. (2014). Lipolysis in lipid turnover, cancer cachexia, and obesity-induced insulin resistance. Trends Endocrinol. Metab. *25*, 255–262.

Ashkar, S., Weber, G.F., Panoutsakopoulou, V., Sanchirico, M.E., Jansson, M., Zawaideh, S., Rittling, S.R., Denhardt, D.T., Glimcher, M.J., and Cantor, H. (2000). Eta-1 (Osteopontin): An Early Component of Type-1 (Cell-Mediated) Immunity. Science *287*, 860–864.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA Biogenesis Competes with Pre-mRNA Splicing. Mol. Cell *56*, 55–66.

Asikainen, S., Heikkinen, L., Juhila, J., Holm, F., Weltner, J., Trokovic, R., Mikkola, M., Toivonen, S., Balboa, D., Lampela, R., et al. (2015). Selective MicroRNA-Offset RNA Expression in Human Embryonic Stem Cells. PLoS ONE *10*, e0116668.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. *22*, 2773–2785.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64–71.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. *41*, D991–D995.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. Nucleic Acids Res. *41*, D36–D42.

Beretta, M., Bauer, M., and Hirsch, E. (2015). PI3K signaling in the pathogenesis of obesity: The cause and the cure. Adv. Biol. Regul. *58*, 1–15.

Berry, R., and Rodeheffer, M.S. (2013). Characterization of the adipocyte cellular lineage in vivo. Nat. Cell Biol. *15*, 302–308.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. *11*, R90.

Biasiolo, M., Sales, G., Lionetti, M., Agnelli, L., Todoerti, K., Bisognin, A., Coppe, A., Romualdi, C., Neri, A., and Bortoluzzi, S. (2011). Impact of Host Genes and Strand Selection on miRNA and miRNA* Expression. PLoS ONE *6*, e23854.

Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., et al. (2009). De novo transcriptome assembly with ABySS. Bioinformatics *25*, 2872–2877.

Bisognin, A., Sales, G., Coppe, A., Bortoluzzi, S., and Romualdi, C. (2012). MAGIA2: from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). Nucleic Acids Res. *40*, W13–W21.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bompfünewerer, A.F., Backofen, R., Bernhart, S.H., Hertel, J., Hofacker, I.L., Stadler, P.F., and Will, S. (2008). Variations on RNA folding and alignment: lessons from Benasque. J. Math. Biol. *56*, 129–144.

Bork, S., Horn, P., Castoldi, M., Hellwig, I., Ho, A.D., and Wagner, W. (2011). Adipogenic differentiation of human mesenchymal stromal cells is down-regulated by microRNA-369-5p and up-regulated by microRNA-371. J. Cell. Physiol. *226*, 2226–2234.

Bortoluzzi, S., Bisognin, A., Biasiolo, M., Guglielmelli, P., Biamonte, F., Norfo, R., Manfredini, R., and Vannucchi, A.M. (2012). Characterization and discovery of novel miRNAs and moRNAs in JAK2V617F-mutated SET2 cells. Blood *119*, e120–e130.

Bosi, P., and Russo, V. (2010). The production of the heavy pig for high quality processed products. Ital. J. Anim. Sci. *3*, 309–321.

Buermans, H.P.J., and den Dunnen, J.T. (2014). Next generation sequencing technology: Advances and applications. Biochim. Biophys. Acta BBA - Mol. Basis Dis. *1842*, 1932–1941.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

Bustin, S.A., and Nolan, T. (2004). Pitfalls of Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction. J. Biomol. Tech. JBT *15*, 155–166.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Čandek-Potokar, M., and Škrlep, M. (2012). Factors in pig production that impact the quality of dry-cured ham: a review. Animal *6*, 327–338.

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell *73*, 1019–1030.

Cawthorn, W.P., Scheller, E.L., and MacDougald, O.A. (2012). Adipose tissue stem cells meet preadipocyte commitment: going back to the future. J. Lipid Res. *53*, 227–246.

Chakrabarti, P., and Kandror, K.V. (2009). FoxO1 Controls Insulin-dependent Adipose Triglyceride Lipase (ATGL) Expression and Lipolysis in Adipocytes. J. Biol. Chem. *284*, 13296–13300.

Chandalia, M., Davila, H., Pan, W., Szuszkiewicz, M., Tuvdendorj, D., Livingston, E.H., and Abate, N. (2012). Adipose Tissue Dysfunction in Humans: A Potential Role for the Transmembrane Protein ENPP1. J. Clin. Endocrinol. Metab. *97*, 4663–4672.

Chen, L.-L., and Yang, L. (2015). Regulation of circRNA biogenesis. RNA Biol. *12*, 381–388.

Chen, C., Ai, H., Ren, J., Li, W., Li, P., Qiao, R., Ouyang, J., Yang, M., Ma, J., and Huang, L. (2011). A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. BMC Genomics *12*, 448.

Chen, C., Deng, B., Qiao, M., Zheng, R., Chai, J., Ding, Y., Peng, J., and Jiang, S. (2012). Solexa sequencing identification of conserved and novel microRNAs in backfat of Large White and Chinese Meishan pigs. PloS One *7*, e31426.

Chen, C.H., Lin, E.C., Cheng, W.T.K., Sun, H.S., Mersmann, H.J., and Ding, S.T. (2006). Abundantly expressed genes in pig adipose tissue: an expressed sequence tag approach. J. Anim. Sci. *84*, 2673–2683.

Chen, L., Song, J., Cui, J., Hou, J., Zheng, X., Li, C., and Liu, L. (2013). microRNAs regulate adipocyte differentiation. Cell Biol. Int. *37*, 533–546.

Cho, I.S., Kim, J., Seo, H.Y., Lim, D.H., Hong, J.S., Park, Y.H., Park, D.C., Hong, K.-C., Whang, K.Y., and Lee, Y.S. (2010). Cloning and characterization of microRNAs from porcine skeletal muscle and adipose tissue. Mol. Biol. Rep. *37*, 3567–3574.

Choi, S.M., Tucker, D.F., Gross, D.N., Easton, R.M., DiPilato, L.M., Dean, A.S., Monks, B.R., and Birnbaum, M.J. (2010). Insulin Regulates Adipocyte Lipolysis via an Akt-Independent Signaling Pathway. Mol. Cell. Biol. *30*, 5009–5020.

Choy, L., Skillington, J., and Derynck, R. (2000). Roles of Autocrine TGF-β Receptor and Smad Signaling in Adipocyte Differentiation. J. Cell Biol. *149*, 667–682.

Cirera, S., Birck, M., Busk, P.K., and Fredholm, M. (2010). Expression Profiles of miRNA-122 and Its Target CAT1 in Minipigs (Sus scrofa) Fed a High-Cholesterol Diet. Comp. Med. *60*, 136–141.

Civelek, M., Hagopian, R., Pan, C., Che, N., Yang, W., Kayne, P.S., Saleem, N.K., Cederberg, H., Kuusisto, J., Gargalovic, P.S., et al. (2013). Genetic regulation of human adipose microRNA expression and its consequences for metabolic traits. Hum. Mol. Genet. *22*, 3023–3037.

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E., et al. (2011). MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. *12*, R126.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA Binding Protein Quaking Regulates Formation of circRNAs. Cell *160*, 1125–1134.

Consortium, I.H.G.S. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castelló, A., Alves, E., Pena, R.N., Ballester, M., and Folch, J.M. (2013). Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. BMC Genomics *14*, 843.

Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics *11*, 485.

Cristancho, A.G., and Lazar, M.A. (2011). Forming functional fat: a growing understanding of adipocyte differentiation. Nat. Rev. Mol. Cell Biol. *12*, 722–734.

Cristancho, A.G., Schupp, M., Lefterova, M.I., Cao, S., Cohen, D.M., Chen, C.S., Steger, D.J., and Lazar, M.A. (2011). Repressor transcription factor 7-like 1 promotes adipogenic competency in precursor cells. Proc. Natl. Acad. Sci. *108*, 16271–16276.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. *43*, D662–D669.

Dani, C. (2013). Activins in adipogenesis and obesity. Int. J. Obes. *37*, 163–166.

Degerman, E., Landström, T.R., Wijkander, J., Holst, L.S., Ahmad, F., Belfrage, P., and Manganiello, V. (1998). Phosphorylation and Activation of Hormone-Sensitive Adipocyte Phosphodiesterase Type 3B. Methods *14*, 43–53.

Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F.M. (2013). An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. PLoS ONE *8*, e85024.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. *4*.

Desvignes, T., Batzel, P., Berezikov, E., Eilbeck, K., Eppig, J.T., McAndrews, M.S., Singer, A., and Postlethwait, J.H. (2015). miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. Trends Genet. *31*, 613–626.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Eguchi, J., Wang, X., Yu, S., Kershaw, E.E., Chiu, P.C., Dushay, J., Estall, J.L., Klein, U., Maratos-Flier, E., and Rosen, E.D. (2011). Transcriptional Control of Adipose Lipid Handling by IRF4. Cell Metab. *13*, 249–259.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Mol. Cell *32*, 519–528.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol. *5*, R1.

Esau, C., Kang, X., Peralta, E., Hanson, E., Marcusson, E.G., Ravichandran, L.V., Sun, Y., Koo, S., Perera, R.J., Jain, R., et al. (2004). MicroRNA-143 Regulates Adipocyte Differentiation. J. Biol. Chem. *279*, 52361–52365.

Eto, H., Suga, H., Matsumoto, D., Inoue, K., Aoi, N., Kato, H., Araki, J., and Yoshimura, K. (2009). Characterization of Structure and Cellular Components of Aspirated and Excised Adipose Tissue: Plast. Reconstr. Surg. *124*, 1087–1097.

Falaleeva, M., and Stamm, S. (2013). Processing of snoRNAs as a new source of regulatory non-coding RNAs. BioEssays *35*, 46–54.

Ferragina, P., and Manzini, G. (2001). An Experimental Study of an Opportunistic Index. In Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms, (Philadelphia, PA, USA: Society for Industrial and Applied Mathematics), pp. 269–278.

Ferraz, A.L.J., Ojeda, A., López-Béjar, M., Fernandes, L.T., Castelló, A., Folch, J.M., and Pérez-Enciso, M. (2008). Transcriptome architecture across tissues in the pig. BMC Genomics *9*, 173.

Festa, E., Fretz, J., Berry, R., Schmidt, B., Rodeheffer, M., Horowitz, M., and Horsley, V. (2011). Adipocyte Lineage Cells Contribute to the Skin Stem Cell Niche to Drive Hair Cycling. Cell *146*, 761–771.

Fonseca, N.A., Rung, J., Brazma, A., and Marioni, J.C. (2012). Tools for mapping high-throughput sequencing data. Bioinformatics *28*, 3169–3177.

Fox, A.H. (2014). The noncoding RNA revolution. Int. J. Biochem. Cell Biol. *54*, 287.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2011). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res.

Friedländer, M.R., Lizano, E., Houben, A.J., Bezdan, D., Báñez-Coronel, M., Kudla, G., Mateu-Huertas, E., Kagerbauer, B., González, J., Chen, K.C., et al. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. Genome Biol. *15*, R57.

Funari, V., and Canosa, S.J. (2014). The Importance of Bioinformatics in NGS: Breaking the Bottleneck in Data Interpretation. Science *344*, 653–653.

Gaffo, E., Zambonelli, P., Bisognin, A., Bortoluzzi, S., and Davoli, R. (2014). miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs. Anim. Genet. *45*, 685–698.

Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol. *16*, 4.

Garten, A., Schuster, S., and Kiess, W. (2012). The Insulin-Like Growth Factors in Adipogenesis and Obesity. Endocrinol. Metab. Clin. North Am. *41*, 283–295.

Ge, Q., Brichard, S., Yi, X., and Li, Q. (2014). MicroRNAs as a new mechanism regulating adipose tissue inflammation in obesity and as a novel therapeutic strategy in the metabolic syndrome. J. Immunol. Res. *2014*.

Gennarino, V.A., D'Angelo, G., Dharmalingam, G., Fernandez, S., Russolillo, G., Sanges, R., Mutarelli, M., Belcastro, V., Ballabio, A., Verde, P., et al. (2012). Identification of microRNA-regulated gene networks by expression analysis of target genes. Genome Res. *22*, 1163–1172.

Gil, A., Olza, J., Gil-Campos, M., Gomez-Llorente, C., and Aguilera, C.M. (2011). Is adipose tissue metabolically different at different sites? Int. J. Pediatr. Obes. *6*, 13–20.

Goecks, J., Nekrutenko, A., Taylor, J., and $author.lastName, $author firstName (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Gorodkin, J., Cirera, S., Hedegaard, J., Gilchrist, M.J., Panitz, F., Jørgensen, C., Scheibye-Knudsen, K., Arvin, T., Lumholdt, S., Sawera, M., et al. (2007). Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. Genome Biol. *8*, R45.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Granneman, J.G., Moore, H.-P.H., Krishnamoorthy, R., and Rathod, M. (2009). Perilipin Controls Lipolysis by Regulating the Interactions of AB-hydrolase Containing 5 (Abhd5) and Adipose Triglyceride Lipase (Atgl). J. Biol. Chem. *284*, 34538–34544.

Greenleaf, W.J., and Sidow, A. (2014). The future of sequencing: convergence of intelligent design and market Darwinism. Genome Biol. *15*, 303.

Gregoire, F.M., Smas, C.M., and Sul, H.S. (1998). Understanding Adipocyte Differentiation. Physiol. Rev. *78*, 783–809.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res. *32*, D109–D111.

Griffiths-Jones, S., Grocock, R.J., Dongen, S. van, Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. *34*, D140–D144.

Griffiths-Jones, S., Saini, H.K., Dongen, S. van, and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res. *36*, D154–D158.

Guo, L., Niu, J., Yu, H., Gu, W., Li, R., Luo, X., Huang, M., Tian, Z., Feng, L., and Wang, Y. (2014). Modulation of CD163 Expression by Metalloprotease ADAM17 Regulates Porcine Reproductive and Respiratory Syndrome Virus Entry. J. Virol. *88*, 10448–10458.

Guo, Y., Chen, Y., Zhang, Y., Zhang, Y., Chen, L., and Mo, D. (2012a). Up-regulated miR-145 expression inhibits porcine preadipocytes differentiation by targeting IRS1. Int. J. Biol. Sci. *8*, 1408–1417.

Guo, Y., Mo, D., Zhang, Y., Zhang, Y., Cong, P., Xiao, S., He, Z., Liu, X., and Chen, Y. (2012b). MicroRNAome Comparison between Intramuscular and Subcutaneous Vascular Stem Cell Adipogenesis. PLoS ONE *7*.

Gupta, R.K., Arany, Z., Seale, P., Mepani, R.J., Ye, L., Conroe, H.M., Roby, Y.A., Kulaga, H., Reed, R.R., and Spiegelman, B.M. (2010). Transcriptional control of preadipocyte determination by Zfp423. Nature *464*, 619–623.

Haasken, S., Auger, J.L., Taylor, J.J., Hobday, P.M., Goudy, B.D., Titcombe, P.J., Mueller, D.L., and Binstadt, B.A. (2013). Macrophage scavenger receptor 1 (Msr1, SR-A) influences B cell autoimmunity by regulating soluble autoantigen concentration. J. Immunol. Baltim. Md 1950 *191*, 1055–1062.

Han, L., Vickers, K.C., Samuels, D.C., and Guo, Y. (2015). Alternative applications for distinct RNA sequencing strategies. Brief. Bioinform. *16*, 629–639.

Han, S.-H., Shin, K.-Y., Lee, S.-S., Ko, M.-S., Oh, H.-S., and Cho, I.-C. (2012). Porcine SPP1 gene polymorphism association with phenotypic traits in the Landrace × Jeju (Korea) black pig F2 population. Mol. Biol. Rep. *39*, 7705–7709.

Hansen, K.D., Irizarry, R.A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics *13*, 204–216.

Hansen, T.B., Kjems, J., and Damgaard, C.K. (2013). Circular RNA and miR-7 in Cancer. Cancer Res. *73*, 5609–5612.

Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M.B., Leboeuf, M., Becker, C.D., See, P., Price, J., Lucas, D., et al. (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. Immunity *38*, 792–804.

Hauner, H. (2005). Secretory factors from human adipose tissue and their functional role. Proc. Nutr. Soc. *64*, 163–169.

Hausser, J., Syed, A.P., Bilen, B., and Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. Genome Res.

Henderson, C.R., and Quaas, R.L. (1976). Multiple Trait Evaluation Using Relatives' Records. J. Anim. Sci. *43*.

Hilton, C., Neville, M.J., and Karpe, F. (2013). MicroRNAs in adipose tissue: their role in adipogenesis and obesity. Int. J. Obes. *37*, 325–332.

Hofacker, I.L., and Stadler, P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics *22*, 1172–1176.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte Für Chem. Chem. Mon. *125*, 167–188.

Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L.M., Teupser, D., Hackermüller, J., et al. (2014). A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol. *15*, R34.

Hoggard, N., Cruickshank, M., Moar, K.M., Barrett, P., Bashir, S., and Miller, J.D.B. (2009). Inhibin βB expression in murine adipose tissue and its regulation by leptin, insulin and dexamethasone. J. Mol. Endocrinol. *43*, 171–177.

Hornshøj, H., Conley, L.N., Hedegaard, J., Sørensen, P., Panitz, F., and Bendixen, C. (2007). Microarray expression profiles of 20.000 genes across 23 healthy porcine tissues. PloS One *2*, e1203.

Hotamisligil, G.S. (2006). Inflammation and metabolic disorders. Nature *444*, 860–867.

Hotamisligil, G.S., Shargill, N.S., and Spiegelman, B.M. (1993). Adipose Expression of Tumor Necrosis Factor-α: Direct Role in Obesity-Linked Insulin Resistance. Science *259*, 87–91.

Hu, E., Tontonoz, P., and Spiegelman, B.M. (1995). Transdifferentiation of myoblasts by the adipogenic transcription factors PPAR gamma and C/EBP alpha. Proc. Natl. Acad. Sci. *92*, 9856–9860.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, J.C., Babak, T., Corson, T.W., Chua, G., Khan, S., Gallie, B.L., Hughes, T.R., Blencowe, B.J., Frey, B.J., and Morris, Q.D. (2007). Using expression profiling data to identify human microRNA targets. Nat. Methods *4*, 1045–1049.

Huang, S., Wang, S., Bian, C., Yang, Z., Zhou, H., Zeng, Y., Li, H., Han, Q., and Zhao, R.C. (2012). Upregulation of miR-22 promotes osteogenic differentiation and inhibits adipogenic differentiation of human adipose tissue-derived mesenchymal stem cells by repressing HDAC6 protein expression. Stem Cells Dev. *21*, 2531–2540.

Hulsmans, M., De Keyzer, D., and Holvoet, P. (2011). MicroRNAs regulating oxidative stress and inflammation in relation to obesity and atherosclerosis. FASEB J. *25*, 2515–2527.

Hunkapiller, T., Kaiser, R.J., Koop, B.F., and Hood, L. (1991). Large-scale and automated DNA sequence determination. Science *254*, 59–67.

Ibáñez-Escriche, N., Forni, S., Noguera, J.L., and Varona, L. (2014). Genomic information in pig breeding: Science meets industry needs. Livest. Sci. *166*, 94–100.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. *47*, 199–208.

Jeck, W.R., and Sharpless, N.E. (2014). Detecting and characterizing circular RNAs. Nat. Biotechnol. *32*, 453–461.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*, 141–157.

Jenkins, S.J., Ruckerl, D., Cook, P.C., Jones, L.H., Finkelman, F.D., Rooijen, N. van, MacDonald, A.S., and Allen, J.E. (2011). Local Macrophage Proliferation, Rather than Recruitment from the Blood, Is a Signature of TH2 Inflammation. Science *332*, 1284–1288.

Jeong Kim, Y., Jin Hwang, S., Chan Bae, Y., and Sup Jung, J. (2009). MiR-21 regulates adipogenic differentiation through the modulation of TGF-β signaling in mesenchymal stem cells derived from human adipose tissue. Stem Cells *27*, 3093–3102.

Jiang, S., Wei, H., Song, T., Yang, Y., Peng, J., and Jiang, S. (2013). Transcriptome Comparison between Porcine Subcutaneous and Intramuscular Stromal Vascular Cells during Adipogenic Differentiation. PLoS ONE *8*, e77094.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA Targets. PLoS Biol *2*, e363.

Johnson, A.M.F., and Olefsky, J.M. (2013). The Origins and Drivers of Insulin Resistance. Cell *152*, 673–684.

Kabir, S.M., Lee, E.-S., and Son, D.-S. (2014). Chemokine network during adipogenesis in 3T3-L1 cells. Adipocyte *3*, 97–106.

Kanazawa, A., Tsukada, S., Kamiyama, M., Yanagimoto, T., Nakajima, M., and Maeda, S. (2005). Wnt5b partially inhibits canonical Wnt/β-catenin signaling pathway and promotes adipogenesis in 3T3-L1 preadipocytes. Biochem. Biophys. Res. Commun. *330*, 505–510.

Kang, S., Akerblad, P., Kiviranta, R., Gupta, R.K., Kajimura, S., Griffin, M.J., Min, J., Baron, R., and Rosen, E.D. (2012). Regulation of Early Adipose Commitment by Zfp521. PLoS Biol *10*, e1001433.

Kanneganti, T.-D., and Dixit, V.D. (2012). Immunological complications of obesity. Nat. Immunol. *13*, 707–712.

Karbiener, M., Fischer, C., Nowitsch, S., Opriessnig, P., Papak, C., Ailhaud, G., Dani, C., Amri, E.-Z., and Scheideler, M. (2009). microRNA miR-27b impairs human adipocyte differentiation and targets PPARγ. Biochem. Biophys. Res. Commun. *390*, 247–251.

Keller, P., Gburcik, V., Petrovic, N., Gallagher, I.J., Nedergaard, J., Cannon, B., and Timmons, J.A. (2011). Gene-chip studies of adipogenesis-regulated microRNAs in mouse primary adipocytes and human obesity. BMC Endocr. Disord. *11*.

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat. Cell Biol. *14*, 659–665.

Kershaw, E.E., and Flier, J.S. (2004). Adipose Tissue as an Endocrine Organ. J. Clin. Endocrinol. Metab. *89*, 2548–2556.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat. Genet. *39*, 1278.

Kim, D., and Salzberg, S.L. (2011). TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. *12*, R72.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, S.Y., Kim, A.Y., Lee, H.W., Son, Y.H., Lee, G.Y., Lee, J.-W., Lee, Y.S., and Kim, J.B. (2010). miR-27a is a negative regulator of adipocyte differentiation via suppressing PPARγ expression. Biochem. Biophys. Res. Commun. *392*, 323–328.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.

Kinoshita, M., Ono, K., Horie, T., Nagao, K., Nishi, H., Kuwabara, Y., Takanabe-Mori, R., Hasegawa, K., Kita, T., and Kimura, T. (2010). Regulation of adipocyte differentiation by activation of serotonin (5-HT) receptors 5-HT2AR and 5-HT2CR and involvement of microRNA-448-mediated repression of KLF5. Mol. Endocrinol. *24*, 1978–1987.

Kitamura, T., Kitamura, Y., Kuroda, S., Hino, Y., Ando, M., Kotani, K., Konishi, H., Matsuzaki, H., Kikkawa, U., Ogawa, W., et al. (1999). Insulin-Induced Phosphorylation and Activation of Cyclic Nucleotide Phosphodiesterase 3B by the Serine-Threonine Kinase Akt. Mol. Cell. Biol. *19*, 6286–6296.

Klöting, N., Berthold, S., Kovacs, P., Schön, M.R., Fasshauer, M., Ruschke, K., Stumvoll, M., and Blüher, M. (2009). MicroRNA expression in human omental and subcutaneous adipose tissue. PLoS ONE *4*.

Kogelman, L.J.A., Cirera, S., Zhernakova, D.V., Fredholm, M., Franke, L., and Kadarmideen, H.N. (2014). Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med. Genomics *7*, 57.

Koh, Y.J., Kang, S., Lee, H.J., Choi, T.-S., Lee, H.S., Cho, C.-H., and Koh, G.Y. (2007). Bone marrow–derived circulating progenitor cells fail to transdifferentiate into adipocytes in adult adipose tissues in mice. J. Clin. Invest. *117*, 3684–3695.

Kommadath, A., Bao, H., Arantes, A.S., Plastow, G.S., Tuggle, C.K., Bearson, S.M., Luo Guan, L., and Stothard, P. (2014). Gene co-expression network analysis identifies porcine genes associated with variation in Salmonella shedding. BMC Genomics *15*.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. *35*, W345–W349.

Kong, Y.W., Ferland-McCollough, D., Jackson, T.J., and Bushell, M. (2012). microRNAs in cancer management. Lancet Oncol. *13*, e249–e258.

Konige, M., Wang, H., and Sztalryd, C. (2014). Role of adipose specific lipid droplet proteins in maintaining whole body energy homeostasis. Biochim. Biophys. Acta BBA - Mol. Basis Dis. *1842*, 393–401.

Koopmans, S.J., and Schuurman, T. (2015). Considerations on pig models for appetite, metabolic syndrome and obese type 2 diabetes: From food intake to metabolic disease. Eur. J. Pharmacol. *759*, 231–239.

Kozomara, A., and Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68–D73.

Krotkiewski, M., Björntorp, P., Sjöström, L., and Smith, U. (1983). Impact of obesity on metabolism in men and women. Importance of regional adipose tissue distribution. J. Clin. Invest. *72*, 1150–1162.

Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. *34*, W451–W454.

Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, P., and Stadler, P.F. (2009). Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics *25*, 2298–2301.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lee, E.K., Lee, M.J., Abdelmohsen, K., Kim, W., Kim, M.M., Srikantan, S., Martindale, J.L., Hutchison, E.R., Kim, H.H., Marasa, B.S., et al. (2011). miR-130 suppresses adipogenesis by inhibiting peroxisome proliferator-activated receptor γ expression. Mol. Cell. Biol. *31*, 626–638.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Leinonen, R., Sugawara, H., and Shumway, M. (2011). The Sequence Read Archive. Nucleic Acids Res. *39*, D19–D21.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, G., Li, Y., Li, X., Ning, X., Li, M., and Yang, G. (2011). MicroRNA identity and abundance in developing swine adipose tissue as determined by solexa sequencing. J. Cell. Biochem. *112*, 1318–1328.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, H.-Y., Xi, Q.-Y., Xiong, Y.-Y., Liu, X.-L., Cheng, X., Shu, G., Wang, S.-B., Wang, L.-N., Gao, P., Zhu, X.-T., et al. (2012a). Identification and comparison of microRNAs from skeletal muscle and adipose tissues from two porcine breeds. Anim. Genet. *43*, 704–713.

Li, J., Yang, J., Zhou, P., Le, Y., Zhou, C., Wang, S., Xu, D., Lin, H.-K., and Gong, Z. (2015a). Circular RNAs in cancer: novel insights into origins, properties, functions and implications. Am. J. Cancer Res. *5*, 472–480.

Li, P., Chen, S., Chen, H., Mo, X., Li, T., Shao, Y., Xiao, B., and Guo, J. (2015b). Using circular RNA as a novel type of biomarker in the screening of gastric cancer. Clin. Chim. Acta *444*, 132–136.

Li, X.J., Yang, H., Li, G.X., Zhang, G.H., Cheng, J., Guan, H., and Yang, G.S. (2012b). Transcriptome profile analysis of porcine adipose tissue by high-throughput sequencing. Anim. Genet. *43*, 144–152.

Li, Z., Ender, C., Meister, G., Moore, P.S., Chang, Y., and John, B. (2012c). Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Res. *40*, 6787–6799.

Liang, J., Fu, M., Ciociola, E., Chandalia, M., and Abate, N. (2007). Role of ENPP1 on Adipocyte Maturation. PLoS ONE *2*, e882.

Lin, Q., Gao, Z., Alarcon, R.M., Ye, J., and Yun, Z. (2009). A role of miR-27 in the regulation of adipogenesis. FEBS J. *276*, 2348–2358.

Lionetti, M., Biasiolo, M., Agnelli, L., Todoerti, K., Mosca, L., Fabris, S., Sales, G., Deliliers, G.L., Bicciato, S., Lombardi, L., et al. (2009). Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma. Blood *114*, e20–e26.

Litten-Brown, J.C., Corson, A.M., and Clarke, L. (2010). Porcine models for the metabolic syndrome, digestive and bone disorders: a general overview. Animal *4*, 899–920.

Liu, H.-C., Hicks, J.A., Trakooljul, N., and Zhao, S.-H. (2010). Current knowledge of microRNA characterization in agricultural animals. Anim. Genet. *41*, 225–231.

Liu, S., Yang, Y., and Wu, J. (2011). TNFα-induced up-regulation of miR-155 inhibits adipogenesis by down-regulating early adipogenic transcription factors. Biochem. Biophys. Res. Commun. *414*, 618–624.

Llorens, F., Bañez-Coronel, M., Pantano, L., Río, J.A. del, Ferrer, I., Estivill, X., and Martí, E. (2013). A highly expressed miR-101 isomiR is a functional silencing small RNA. BMC Genomics *14*, 104.

Londin, E., Loher, P., Telonis, A.G., Quann, K., Clark, P., Jing, Y., Hatzimichael, E., Kirino, Y., Honda, S., Lally, M., et al. (2015). Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc. Natl. Acad. Sci. 201420955.

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.

Lumeng, C.N., and Saltiel, A.R. (2011). Inflammatory links between obesity and metabolic disease. J. Clin. Invest. *121*, 2111–2117.

Macotela, Y., Emanuelli, B., Mori, M.A., Gesta, S., Schulz, T.J., Tseng, Y.-H., and Kahn, C.R. (2012). Intrinsic Differences in Adipocyte Precursor Cells From Different White Fat Depots. Diabetes *61*, 1691–1699.

Mardis, E.R. (2013). Next-Generation Sequencing Platforms. Annu. Rev. Anal. Chem. *6*, 287–303.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. Nat. Rev. Genet. *12*, 671–682.

Maury, E., and Brichard, S.M. (2010). Adipokine dysregulation, adipose tissue inflammation and metabolic syndrome. Mol. Cell. Endocrinol. *314*, 1–16.

Maute, R.L., Schneider, C., Sumazin, P., Holmes, A., Califano, A., Basso, K., and Dalla-Favera, R. (2013). tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. *110*, 1404–1409.

McCaskill, J.S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers *29*, 1105–1119.

McCullough, A.W. (1944). Evidence of the macrophagal origin of adipose cells in the white rat as shown by studies on starved animals. J. Morphol. *75*, 193–201.

McCurdy, C.E., and Klemm, D.J. (2013). Adipose tissue insulin sensitivity and macrophage recruitment. Adipocyte *2*, 135–142.

McDaneld, T.G., Smith, T.P., Doumit, M.E., Miles, J.R., Coutinho, L.L., Sonstegard, T.S., Matukumalli, L.K., Nonneman, D.J., and Wiedmann, R.T. (2009). MicroRNA transcriptome profiles during swine skeletal muscle development. BMC Genomics *10*, 77–2164 – 10–77.

McGregor, R.., and Choi, M.. (2011). microRNAs in the Regulation of Adipogenesis and Obesity. Curr. Mol. Med. *11*, 304–316.

Meerson, A., Traurig, M., Ossowski, V., Fleming, J.M., Mullins, M., and Baier, L.J. (2013). Human adipose microRNA-221 is upregulated in obesity and affects fat metabolism downstream of leptin and TNF-α. Diabetologia *56*, 1971–1979.

Mehta, J.L., and Li, D. (2002). Identification, regulation and function of a novel lectin-like oxidized low-density lipoprotein receptor. J. Am. Coll. Cardiol. *39*, 1429–1435.

Meiri, E., Levy, A., Benjamin, H., Ben-David, M., Cohen, L., Dov, A., Dromi, N., Elyakim, E., Yerushalmi, N., Zion, O., et al. (2010). Discovery of microRNAs and other small RNAs in solid tumors. Nucleic Acids Res. *38*, 6234–6246.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.

Meyre, D., Bouatia-Naji, N., Tounian, A., Samson, C., Lecoeur, C., Vatin, V., Ghoussaini, M., Wachter, C., Hercberg, S., Charpentier, G., et al. (2005). Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. Nat. Genet. *37*, 863–867.

Mikawa, A., Suzuki, H., Suzuki, K., Toki, D., Uenishi, H., Awata, T., and Hamasima, N. (2004). Characterization of 298 ESTs from porcine back fat tissue and their assignment to the SSRH radiation hybrid map. Mamm. Genome *15*, 315–322.

Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A.M., Lim, B., and Rigoutsos, I. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. Cell *126*, 1203–1217.

Moon, J.-K., Kim, K.-S., Kim, J.-J., Choi, B.-H., Cho, B.-W., Kim, T.-H., and Lee, C.-K. (2009). Differentially expressed transcripts in adipose tissue between Korean native pig and Yorkshire breeds. Anim. Genet. *40*, 115–118.

Moreno-Navarrete, J.M., and Fernández-Real, J.M. (2012). Adipocyte differentiation. Adipose Tissue Biol. 17–38.

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., and Schäffer, A.A. (2008). Database indexing for production MegaBLAST searches. Bioinformatics *24*, 1757–1764.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. *18*, 610–621.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Mudhasani, R., Imbalzano, A.N., and Jones, S.N. (2010). An essential role for dicer in adipocyte differentiation. J. Cell. Biochem. *110*, 812–816.

Neilsen, C.T., Goodall, G.J., and Bracken, C.P. (2012). IsomiRs – the overlooked repertoire in the dynamic microRNAome. Trends Genet. *28*, 544–549.

Nekrutenko, A., and Taylor, J. (2012). Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat. Rev. Genet. *13*, 667–672.

Newsholme, P., and de Bittencourt, P.I.H. (2014). The fat cell senescence hypothesis: a mechanism responsible for abrogating the resolution of inflammation in chronic disease. Curr. Opin. Clin. Nutr. Metab. Care *17*, 295–305.

Nomiyama, T., Perez-Tilve, D., Ogawa, D., Gizard, F., Zhao, Y., Heywood, E.B., Jones, K.L., Kawamori, R., Cassis, L.A., Tschöp, M.H., et al. (2007). Osteopontin mediates obesity-induced adipose tissue macrophage infiltration and insulin resistance in mice. J. Clin. Invest. *117*, 2877–2888.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. Cell *130*, 89–100.

Ørom, U.A., Nielsen, F.C., and Lund, A.H. (2008). MicroRNA-10a Binds the 5′UTR of Ribosomal Protein mRNAs and Enhances Their Translation. Mol. Cell *30*, 460–471.

Ortega, F.J., Moreno-Navarrete, J.M., Pardo, G., Sabater, M., Hummel, M., Ferrer, A., Rodriguez-Hermosa, J.I., Ruiz, B., Ricart, W., Peral, B., et al. (2010). MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. PLoS ONE *5*.

Oskowitz, A.Z., Lu, J., Penfornis, P., Ylostalo, J., McBride, J., Flemington, E.K., Prockop, D.J., and Pochampally, R. (2008). Human multipotent stromal cells from bone marrow and microRNA: Regulation of differentiation and leukemia inhibitory factor expression. Proc. Natl. Acad. Sci. *105*, 18372–18377.

Ottaviani, E., Malagoli, D., and Franceschi, C. (2011). The evolution of the adipose tissue: A neglected enigma. Gen. Comp. Endocrinol. *174*, 1–4.

Padilla, J., Jenkins, N.T., Lee, S., Zhang, H., Cui, J., Zuidema, M.Y., Zhang, C., Hill, M.A., Perfield, J.W., Ibdah, J.A., et al. (2013). Vascular transcriptional alterations produced by juvenile obesity in Ossabaw swine. Physiol. Genomics *45*, 434–446.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. *40*, 1413–1415.

Pan, W., Ciociola, E., Saraf, M., Tumurbaatar, B., Tuvdendorj, D., Prasad, S., Chandalia, M., and Abate, N. (2011). Metabolic consequences of ENPP1 overexpression in adipose tissue. Am. J. Physiol. - Endocrinol. Metab. *301*, E901–E911.

Park, B.O., Ahrends, R., and Teruel, M.N. (2012). Consecutive Positive Feedback Loops Create a Bistable Switch that Controls Preadipocyte-to-Adipocyte Conversion. Cell Rep. *2*, 976–990.

Parra, P., Serra, F., and Palou, A. (2010). Expression of Adipose MicroRNAs Is Sensitive to Dietary Conjugated Linoleic Acid Treatment in Mice. PLoS ONE *5*, e13005.

Peng, Y., Xiang, H., Chen, C., Zheng, R., Chai, J., Peng, J., and Jiang, S. (2013). MiR-224 impairs adipocyte early differentiation and regulates fatty acid metabolism. Int. J. Biochem. Cell Biol. *45*, 1585–1593.

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.

Peterson, S.M., Thompson, J.A., Ufkin, M.L., Sathyanarayana, P., Liaw, L., and Congdon, C.B. (2014). Common features of microRNA target prediction tools. Bioinforma. Comput. Biol. *5*, 23.

Petrovic, N., Walden, T.B., Shabalina, I.G., Timmons, J.A., Cannon, B., and Nedergaard, J. (2010). Chronic Peroxisome Proliferator-activated Receptor γ (PPARγ) Activation of Epididymally Derived White Adipocyte Cultures Reveals a Population of Thermogenically Competent, UCP1-containing Adipocytes Molecularly Distinct from Classic Brown Adipocytes. J. Biol. Chem. *285*, 7153–7164.

Qin, L., Chen, Y., Niu, Y., Chen, W., Wang, Q., Xiao, S., Li, A., Xie, Y., Li, J., Zhao, X., et al. (2010). A deep investigation into the adipogenesis mechanism: Profile of microRNAs regulating adipogenesis by modulating the canonical Wnt/β-catenin signaling pathway. BMC Genomics *11*, 320.

Qiu, Y., Nguyen, K.D., Odegaard, J.I., Cui, X., Tian, X., Locksley, R.M., Palmiter, R.D., and Chawla, A. (2014). Eosinophils and type 2 cytokine signaling in macrophages orchestrate development of functional beige fat. Cell *157*, 1292–1308.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. *17*, 47–62.

Raz, T., Kapranov, P., Lipson, D., Letovsky, S., Milos, P.M., and Thompson, J.F. (2011). Protocol Dependence of Sequencing-Based Gene Expression Measurements. PLoS ONE *6*, e19287.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-Content Normalization for RNA-Seq Data. BMC Bioinformatics *12*, 480.

Ritchie, W., Flamant, S., and Rasko, J.E.J. (2009). Predicting microRNA targets and functions: traps for the unwary. Nat. Methods *6*, 397–398.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011a). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics *27*, 2325–2329.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011b). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. *12*, R22.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Romao, J.M., Jin, W., He, M., McAllister, T., and Guan, L.L. (2012). Altered MicroRNA Expression in Bovine Subcutaneous and Visceral Adipose Tissues from Cattle under Different Diet. PLoS ONE *7*, e40605.

Ropka-Molik, K., Żukowski, K., Eckert, R., Gurgul, A., Piórkowska, K., and Oczkowicz, M. (2014). Comprehensive analysis of the whole transcriptomes from two different pig breeds using RNA-Seq method. Anim. Genet. *45*, 674–684.

Rosen, E.D., and MacDougald, O.A. (2006). Adipocyte differentiation from the inside out. Nat. Rev. Mol. Cell Biol. *7*, 885–896.

Rosen, E.D., Hsu, C.-H., Wang, X., Sakai, S., Freeman, M.W., Gonzalez, F.J., and Spiegelman, B.M. (2002). C/EBPα induces adipogenesis through PPARγ: a unified pathway. Genes Dev. *16*, 22–26.

Ross, S.E., Hemati, N., Longo, K.A., Bennett, C.N., Lucas, P.C., Erickson, R.L., and MacDougald, O.A. (2000). Inhibition of Adipogenesis by Wnt Signaling. Science *289*, 950–953.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86.

Russell, T.D., Palmer, C.A., Orlicky, D.J., Bales, E.S., Chang, B.H.-J., Chan, L., and McManaman, J.L. (2008). Mammary glands of adipophilin-null mice produce an amino-terminally truncated form of adipophilin that mediates milk lipid droplet formation and secretion. J. Lipid Res. *49*, 206–216.

Russo, V., and Nanni Costa, L. (1995). Suitability of pig meat for salting and the production of quality processed products. Pig News Inf. *16*, 17–26.

Russo, V., Buttazzoni, L., Baiocco, C., Davoli, M.R., Nanni Costa, N.L., Schivazappa, O.C., and Virgili, P.C. (2000). Heritability of muscular cathepsin B activity in Italian Large White pigs. J. Anim. Breed. Genet. *117*, 37–42.

Russo, V., Fontanesi, L., Scotti, E., Beretti, F., Davoli, R., Nanni Costa, L., Virgili, R., and Buttazzoni, L. (2008). Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs. J. Anim. Sci. *86*, 3300–3314.

Rydén, M., and Arner, P. (2007). Tumour necrosis factor-α in human adipose tissue – from signalling mechanisms to clinical implications. J. Intern. Med. *262*, 431–438.

Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S., and Romualdi, C. (2010). MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. Nucleic Acids Res. *38*, W352–W359.

Salmanidis, M., Pillman, K., Goodall, G., and Bracken, C. (2014). Direct transcriptional regulation by nuclear microRNAs. Int. J. Biochem. Cell Biol. *54*, 304–311.

Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-Type Specific Features of Circular RNA Expression. PLoS Genet *9*, e1003777.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. *74*, 5463–5467.

Saxena, S., Jónsson, Z.O., and Dutta, A. (2003). Small RNAs with Imperfect Match to Endogenous mRNA Repress Translation IMPLICATIONS FOR OFF-TARGET ACTIVITY OF SMALL INHIBITORY RNA IN MAMMALIAN CELLS. J. Biol. Chem. *278*, 44312–44319.

Schachtschneider, K.M., Madsen, O., Park, C., Rund, L.A., Groenen, M.A., and Schook, L.B. (2015). Adult porcine genome-wide DNA methylation patterns support pigs as a biomedical model. BMC Genomics *16*, 743.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*, 1086–1092.

Sellier, P. (1998). Genetics of meat and carcass traits. 463–510.

Serlachius, M., and Andersson, L.C. (2004). Upregulated expression of stanniocalcin-1 during adipogenesis. Exp. Cell Res. *296*, 256–264.

Sethi, J.K. (2010). Activatin' Human Adipose Progenitors in Obesity. Diabetes *59*, 2354–2357.

Sharp, P.A. (2009). The Centrality of RNA. Cell *136*, 577–580.

Shi, W., Hendrix, D., Levine, M., and Haley, B. (2009). A distinct class of small RNAs arises from pre-miRNA–proximal regions in a simple chordate. Nat. Struct. Mol. Biol. *16*, 183–189.

Siersbæk, R., Nielsen, R., and Mandrup, S. (2012). Transcriptional networks and chromatin remodeling controlling adipogenesis. Trends Endocrinol. Metab. *23*, 56–64.

Smith, S.H., Wilson, A.D., Van Ettinger, I., MacIntyre, N., Archibald, A.L., and Ait-Ali, T. (2014). Down-regulation of mechanisms involved in cell transport and maintenance of mucosal integrity in pigs infected with Lawsonia intracellularis. Vet. Res. *45*, 55.

Sodhi, S.S., Park, W.C., Ghosh, M., Kim, J.N., Sharma, N., Shin, K.Y., Cho, I.C., Ryu, Y.C., Oh, S.J., Kim, S.H., et al. (2014). Comparative transcriptomic analysis to identify differentially expressed genes in fat tissue of adult Berkshire and Jeju Native Pig using RNA-seq. Mol. Biol. Rep.

Sonkoly, E., and Pivarcsi, A. (2009). MicroRNAs in inflammation. Int. Rev. Immunol. *28*, 535–561.

Spalding, K.L., Arner, E., Westermark, P.O., Bernard, S., Buchholz, B.A., Bergmann, O., Blomqvist, L., Hoffstedt, J., Näslund, E., Britton, T., et al. (2008). Dynamics of fat cell turnover in humans. Nature *453*, 783–787.

Spurlock, M.E., and Gabler, N.K. (2008). The Development of Porcine Models of Obesity and the Metabolic Syndrome. J. Nutr. *138*, 397–402.

Steijger, T., Abril, J.F., Engström, P.G., Kokocinski, F., The RGASP Consortium, Hubbard, T.J., Guigó, R., Harrow, J., and Bertone, P. (2013). Assessment of transcript reconstruction methods for RNA-seq. Nat. Methods *10*, 1177–1184.

Sultan, M., Dökel, S., Amstislavskiy, V., Wuttig, D., Sültmann, H., Lehrach, H., and Yaspo, M.-L. (2012). A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. Biochem. Biophys. Res. Commun. *422*, 643–646.

Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H., and Yaspo, M.-L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics *15*, 675.

Sun, F., Wang, J., Pan, Q., Yu, Y., Zhang, Y., Wan, Y., Wang, J., Li, X., and Hong, A. (2009a). Characterization of function and regulation of miR-24-1 and miR-31. Biochem. Biophys. Res. Commun. *380*, 660–665.

Sun, T., Fu, M., Bookout, A.L., Kliewer, S.A., and Mangelsdorf, D.J. (2009b). MicroRNA let-7 regulates 3T3-L1 adipogenesis. Mol. Endocrinol. *23*, 925–931.

Taft, R.J., Simons, C., Nahkuri, S., Oey, H., Korbie, D.J., Mercer, T.R., Holst, J., Ritchie, W., Wong, J.J.-L., Rasko, J.E., et al. (2010). Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. Nat. Struct. Mol. Biol. *17*, 1030–1034.

Tang, Y.-F., Zhang, Y., Li, X.-Y., Li, C., Tian, W., and Liu, L. (2009). Expression of miR-31, miR-125b-5p, and miR-326 in the adipogenic differentiation process of adipose-derived stem cells. OMICS J. Integr. Biol. *13*, 331–336.

Tay, F.C., Lim, J.K., Zhu, H., Hin, L.C., and Wang, S. (2015). Using artificial microRNA sponges to achieve microRNA loss-of-function in cancer cells. Adv. Drug Deliv. Rev. *81*, 117–127.

Tchkonia, T., Thomou, T., Zhu, Y., Karagiannides, I., Pothoulakis, C., Jensen, M.D., and Kirkland, J.L. (2013). Mechanisms and Metabolic Implications of Regional Differences among Fat Depots. Cell Metab. *17*, 644–656.

Toedebusch, R.G., Roberts, M.D., Wells, K.D., Company, J.M., Kanosky, K.M., Padilla, J., Jenkins, N.T., Perfield, J.W., Ibdah, J.A., Booth, F.W., et al. (2014). Unique transcriptomic signature of omental adipose tissue in Ossabaw swine: a model of childhood obesity. Physiol. Genomics *46*, 362–375.

Tontonoz, P., Hu, E., and Spiegelman, B.M. (1994). Stimulation of adipogenesis in fibroblasts by PPARγ2, a lipid-activated transcription factor. Cell *79*, 1147–1156.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M.J. van, Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Trayhurn, P. (2005). Endocrine and signalling role of adipose tissue: new perspectives on fat. Acta Physiol. Scand. *184*, 285–293.

Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. Mol. Cell *26*, 753–767.

Tsang, J.S., Ebert, M.S., and van Oudenaarden, A. (2010). Genome-wide Dissection of MicroRNA Functions and Cotargeting Networks Using Gene Set Signatures. Mol. Cell *38*, 140–153.

Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H., Okumura, N., Hamasima, N., and Awata, T. (2004). PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. Nucleic Acids Res. *32*, D484–D488.

Uenishi, H., Eguchi-Ogawa, T., Shinkai, H., Okumura, N., Suzuki, K., Toki, D., Hamasima, N., and Awata, T. (2007). PEDE (Pig EST Data Explorer) has been expanded into Pig Expression Data Explorer, including 10 147 porcine full-length cDNA sequences. Nucleic Acids Res. *35*, D650–D653.

Umbach, J.L., Strelow, L.I., Wong, S.W., and Cullen, B.R. (2010). Analysis of rhesus rhadinovirus microRNAs expressed in virus-induced tumors from infected rhesus macaques. Virology *405*, 592–599.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*, research0034.1–research0034.11.

Varga, O., Harangi, M., Olsson, I. a. S., and Hansen, A.K. (2010). Contribution of animal models to the understanding of the metabolic syndrome: a systematic overview. Obes. Rev. *11*, 792–807.

Vicario, M., Skaper, S., and Negro, A. (2014). The Small Heat Shock Protein HspB8: Role in Nervous System Physiology and Pathology. CNS Neurol. Disord. - Drug Targets *13*, 885–895.

Walters, E.M., Wolf, E., Whyte, J.J., Mao, J., Renner, S., Nagashima, H., Kobayashi, E., Zhao, J., Wells, K.D., Critser, J.K., et al. (2012). Completion of the swine genome will simplify the production of swine as a large animal biomedical model. BMC Med. Genomics *5*, 55.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008a). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, J., Chen, J., and Sen, S. (2016). MicroRNA as Biomarkers and Diagnostics. J. Cell. Physiol. *231*, 25–30.

Wang, Q., Yan, C.L., Wang, J., Kong, J., Qi, Y., Quigg, R.J., and Li, X. (2008b). miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130. Proc. Natl. Acad. Sci. U. S. A. *105*, 2889–2894.

Wang, T., Jiang, A., Guo, Y., Tan, Y., Tang, G., Mai, M., Liu, H., Xiao, J., Li, M., and Li, X. (2013a). Deep Sequencing of the Transcriptome Reveals Inflammatory Features of Porcine Visceral Adipose Tissue. Int. J. Biol. Sci. *9*, 550–556.

Wang, T., Jiang, A., Guo, Y., Tan, Y., Tang, G., Mai, M., Liu, H., Xiao, J., Li, M., and Li, X. (2013b). Deep Sequencing of the Transcriptome Reveals Inflammatory Features of Porcine Visceral Adipose Tissue. Int. J. Biol. Sci. *9*, 550–556.

Wang, X., Gu, Z., and Jiang, H. (2013c). MicroRNAs in farm animals. Animal *7*, 1567–1575.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Weisberg, S.P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R.L., and Ferrante, A.W. (2003). Obesity is associated with macrophage accumulation in adipose tissue. J. Clin. Invest. *112*, 1796–1808.

Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat. Cell Biol. *11*, 228–234.

Wood, J.D., Enser, M., Fisher, A.V., Nute, G.R., Sheard, P.R., Richardson, R.I., Hughes, S.I., and Whittington, F.M. (2008). Fat deposition, fatty acid composition and meat quality: A review. Meat Sci. *78*, 343–358.

Wronska, A., and Kmiec, Z. (2012). Structural and biochemical characteristics of various white adipose tissue depots. Acta Physiol. *205*, 194–208.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.

Wu, Z., Rosen, E.D., Brun, R., Hauser, S., Adelmant, G., Troy, A.E., McKeon, C., Darlington, G.J., and Spiegelman, B.M. (1999). Cross-Regulation of C/EBPα and PPARγ Controls the Transcriptional Pathway of Adipogenesis and Insulin Sensitivity. Mol. Cell *3*, 151–158.

Wuchty, S., Fontana, W., Hofacker, I.L., and Schuster, P. (1999). Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers *49*, 145–165.

Xie, H., Lim, B., and Lodish, H.F. (2009). MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. Diabetes *58*, 1050–1057.

Xu, H., Barnes, G.T., Yang, Q., Tan, G., Yang, D., Chou, C.J., Sole, J., Nichols, A., Ross, J.S., Tartaglia, L.A., et al. (2003a). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J. Clin. Invest. *112*, 1821–1830.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003b). The Drosophila MicroRNA Mir-14 Suppresses Cell Death and Is Required for Normal Fat Metabolism. Curr. Biol. *13*, 790–795.

Xu, Z., Yu, S., Hsu, C.-H., Eguchi, J., and Rosen, E.D. (2008). The orphan nuclear receptor chicken ovalbumin upstream promoter-transcription factor II is a critical regulator of adipogenesis. Proc. Natl. Acad. Sci. *105*, 2421–2426.

Yadav, H., Quijano, C., Kamaraju, A.K., Gavrilova, O., Malek, R., Chen, W., Zerfas, P., Zhigang, D., Wright, E.C., Stuelten, C., et al. (2011). Protection from Obesity and Diabetes by Blockade of TGF-β/Smad3 Signaling. Cell Metab. *14*, 67–79.

Yamada, K., Santo-Yamada, Y., and Wada, K. (2002). Restraint stress impaired maternal behavior in female mice lacking the neuromedin B receptor (NMB-R) gene. Neurosci. Lett. *330*, 163–166.

Yang, X., Lu, X., Lombès, M., Rha, G.B., Chi, Y.-I., Guerin, T.M., Smart, E.J., and Liu, J. (2010). The G0/G1 Switch Gene 2 Regulates Adipose Lipolysis through Association with Adipose Triglyceride Lipase. Cell Metab. *11*, 194–205.

Yang, Z., Bian, C., Zhou, H., Huang, S., Wang, S., Liao, L., and Zhao, R.C. (2011). MicroRNA hsa-miR-138 inhibits adipogenic differentiation of human adipose tissue-derived mesenchymal stem cells through adenovirus EID-1. Stem Cells Dev. *20*, 259–267.

Yona, S., Kim, K.-W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A., et al. (2013). Fate Mapping Reveals Origins and Dynamics of Monocytes and Tissue Macrophages under Homeostasis. Immunity *38*, 79–91.

Yong, S.L., and Dutta, A. (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev. *21*, 1025–1030.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., et al. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nat. Neurosci. *18*, 603–610.

Young, S.G., and Zechner, R. (2013). Biochemistry and pathophysiology of intravascular and intracellular lipolysis. Genes Dev. *27*, 459–484.

Yu, X., Zhang, X., Dhakal, I.B., Beggs, M., Kadlubar, S., and Luo, D. (2012). Induction of cell proliferation and survival genes by estradiol-repressed microRNAs in breast cancer cells. BMC Cancer *12*.

Zamani, N., and Brown, C.W. (2010). Emerging Roles for the Transforming Growth Factor-β Superfamily in Regulating Adiposity and Energy Expenditure. Endocr. Rev. *32*, 387–403.

Zambonelli, P., Gaffo, E., Zappaterra, M., Bortoluzzi, S.,and Davoli, R. (2016). Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness. An. Gen. In press.

Zardo, G., Ciolfi, A., Vian, L., Starnes, L.M., Billi, M., Racanicchi, S., Maresca, C., Fazi, F., Travaglini, L., Noguera, N., et al. (2012). Polycombs and microRNA-223 regulate human granulopoiesis by transcriptional control of target gene expression. Blood *119*, 4034–4046.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829.

Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary Sequence-Mediated Exon Circularization. Cell *159*, 134–147.

Zhou, C., Zhang, J., Ma, J., Jiang, A., Tang, G., Mai, M., Zhu, L., Bai, L., Li, M., and Li, X. (2013). Gene expression profiling reveals distinct features of various porcine adipose tissues. Lipids Health Dis. *12*, 75.

Zhou, H., Arcila, M.L., Li, Z., Lee, E.J., Henzler, C., Liu, J., Rana, T.M., and Kosik, K.S. (2012). Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res. *40*, 5864–5875.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. *9*, 133–148.