UNIVERSITÀ DEGLI STUDI DEL MOLISE

Department of Agricultural, Environmental and Food Sciences

PhD Course in:

AGRICULTURE TECHNOLOGY AND BIOTECHNOLOGY

(CURRICULUM: ANIMAL WELFARE, BIOTECHNOLOGY AND QUALITY OF ANIMAL PRODUCTION)

(CYCLE XXXI)

Related disciplinary scientific section: AGR/17 (Zootecnica Generale e Miglioramento Genetico)

PhD thesis

Genomics, Transcriptomics and Computational Biology: new insights into bovine and swine breeding and genetics

Coordinator of the PhD Course: Prof. Giuseppe Maiorano

Supervisor: Prof. Mariasilvia D’Andrea

PhD Student: Valentino Palombo

155841

ACADEMIC YEAR 2017/2018

To my family:
we do share more than
just genetic material, after all.

Abstract

Enormi progressi sono stati fatti nella selezione degli animali per specifici caratteri di interesse zootecnico avvalendosi dei tradizionali approcci di genetica quantitativa. Tuttavia, una considerevole quantità di variabilità fenotipica resta ancora non completamente spiegata; in tal senso una migliore conoscenza delle sue basi molecolari e genetiche rappresenterebbe un ulteriore vantaggio. A tal proposito, il recente sviluppo di tecnologie high-throughput (HT), basate su metodi ad alta specificità di ibridazione e sulle ultime tecniche di sequenziamento (NGS), rappresenta una nuova opportunità per esplorare i più complessi meccanismi biologici. La rapida diffusione di queste tecnologie ha segnato l'inizio dell'era 'omica'. Gli approcci 'omici' si basano sull'analisi complessiva di una specifica classe di molecole contenute in una cellula, un tessuto o un organismo; ovvero sono primariamente indirizzati all'analisi di tutti i geni (genomica), di tutti i trascritti (trascrittomica), di tutte le proteine (proteomica) o di tutti i metaboliti (metabolomica) presenti in un campione biologico. La convizione è che un sistema complesso può essere compreso più a fondo, e più fedelmente, se considerato nella sua globalità. La grandissima mole di dati generata, tuttavia, ha senso soltanto se si è equipaggiati con opportuni strumenti per esplorala. Per questo motivo, di pari passo con tali progressi tecnologici, la bioinformatica, conosciuta anche come biologia computazionale, sta acquisendo progressiva importanza. Anche la zootecnia e il miglioramento genetico si stanno avvalendo delle opportunità offerte da questo nuovo scenario. In particolare, ci si sta spostando dagli approcci tradizionali a quelli che prevedono l'uso integrato di analisi omiche. Ciò permette di meglio investigare e decifrate l'architettura genetica alla base dei caratteri di interesse zootecnico ed utilizzare questa informazione per la selezione dei candidati destinati alla riproduzione. L'obiettivo di questa tesi è stato quello di utilizzare le più innovative analisi genomiche e trascrittomiche per (1) investigare le differenze genetiche alla base del profilo acidico del latte in due razze bovine italiane; (2) individuare i geni e i fattori di trascrizione coinvolti nel controllo della colostrogenesi/lattogenesi suina. A tal fine, sono stati effettuati rispettivamente uno studio di associazione lungo tutto il genoma (GWAS) considerando gli acidi grassi del latte in Frisona e Pezzata Rossa Italiana ed è stato sequenziato il trascrittoma (RNA-Sequencing) di ghiandola mammaria suina. In aggiunta (3) è stato sviluppato un nuovo strumento bioinformatico interamente in R, chiamato PIA (Pathways Interaction Analysis), che consente un'originale analisi delle pathway metaboliche utile ad agevolare l'interpretazione dei risultati genomici e trascrittomici.

GENERAL INDEX OF THESIS

GENOMICS, TRANSCRIPTOMICS AND COMPUTATIONAL BIOLOGY: NEW INSIGHTS INTO BOVINE AND SWINE ANIMAL BREEDING AND GENETICS

1. Abstract 1

2. Contents of the thesis 2

3. List of publications 2

Chapter I - GENERAL INTRODUCTION

Abstract 4

Index of chapter 5

Contents of chapter

I - 1. Animal breeding 5

I - 2. Omics sciences 8

I - 3. 'Omic' technologies in animal breeding and genetics 14

I - 4. Aim of the thesis 15

I - 5. References 17

Chapter II - GENOME-WIDE ASSOCIATION STUDY OF MILK FATTY ACID COMPOSITION IN ITALIAN SIMMENTAL AND ITALIAN HOLSTEIN COWS USING SNP ARRAYS

Abstract 24

Index of chapter 25

List of chapter figures and tables 26

Contents of chapter

II - 1. Introduction 28

II - 2. Aim of the study 35

II - 3. Materials and methods 36

II - 4. Results and discussion 40

II - 5. Conclusion 50

II - 6. Figures and tables 52

II - 7. References 68

Chapter III - TRANSCRIPTIONAL PROFILING OF SWINE MAMMARY GLAND DURING THE TRANSITION FROM COLOSTROGENESIS TO LACTOGENESIS USING RNA SEQUENCING

Abstract 86

Index of chapter 87

List of chapter figures and tables 88
Contents of chapter
III - 1. Introduction 90
III - 2. Aim of the study 94
III - 3. Materials and methods 94
III - 4. Results 96
III - 5. Discussion 99
III - 6. Conclusion 119
III - 7. Figures and tables 120
III - 8. References 129
Chapter IV - PIA (PATHWAYS INTERACTION ANALYSIS): AN R TOOL FOR ANALYSING AND INTERPRETING HIGH-THROUGHPUT DATA
Abstract 154
Index of chapter 155
List of chapter figures and tables 156
Contents of chapter
IV - 1. Introduction 158
IV - 2. Aim of the study 159
IV - 3. Methods 159
IV - 4. Validation 169
IV - 5. Conclusion 175
IV - 6. Figures and tables 176
IV - 7. References 184
Chapter V - GENERAL CONCLUSION 190
V - 1. References 193
SUPPLEMENTARY MATERIAL 196

1. Abstract

Enormous progress has been made in the selection of animals for specific traits using traditional quantitative genetic approaches. Nevertheless, a considerable amount of variation in phenotypes remains unexplained therefore a better knowledge of its genetic basis represents a potential additional gain for animal production. In this regard, the recently developed high-throughput (HT) technologies based on microarray and next-generation sequencing (NGS) methods are a powerful opportunity to prise open the 'black box' underlying complex biological processes. These technological advancements have marked the beginning of the 'omic era'.

Broadly, 'omic' approaches adopt a holistic view of the molecules that make up a cell, tissue or organism. They are aimed primarily at the universal detection of genes (genomics), RNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample. The basic aspect of these approaches is that a complex system can be understood more thoroughly if considered as a whole. At the same time, the large amount of data generated by these revolutionary approaches makes sense only if one is equipped with the necessary resources and tools to manage and explore it. For this reason, along with HT technical progresses, bioinformatics, often known as computational biology, is gaining immense importance.

Animal breeding is gaining new momentum from this renewed scenario. Particularly it pushed to move away from traditional approaches toward systems approaches using integrative analysis of 'omic' data to better elucidate the genetic architecture controlling the traits of interest and ultimately use this knowledge for selection of candidates.

The aim of this thesis is to (1) investigate the differences of genetic basis related to the milk fatty acids profiles in two Italian dairy cattle breeds and (2) delineate the genes and transcription regulators implicated in the control of the transition from colostrogenesis to lactogenesis in swine, using the state-of-art genomic and transcriptomic analyses. For these reasons, a genome-wide association study (GWAS) on milk fatty acids of Italian Holstein and Italian Simmental cattle breads and an RNASeq study on transcriptional profiles of swine mammary gland are conducted, respectively. In addition, (3) an in-house bioinformatics tool performing an original pathway analysis is presented. The tool, entirely built in R and named PIA (Pathways Interaction Analysis), is designed for post-genomic and transcriptomic data mining.

2. Contents of the thesis

This thesis is structured into two general chapters (Introduction and Conclusion) and three core chapters containing two studies already published on international peer-reviewed journals and one study under review for publication.

In chapter 2, a GWAS analysis is run to investigate milk fatty acid profile traits in Italian Holstein and Italian Simmental breeds. Data are analysed with a well-established method implemented in the GenABEL R package and with the MUGBAS gene-based association method. Association is investigated between 120K SNPs and 85 fatty acids (as single, aggregated or index values).

In chapter 3, an RNASeq analysis is performed to investigate the mechanism of transition from colostrogenesis to lactogenesis in swine. For this reason, the mammary tissue is collected from three sows at five different time points close to parturition. Once the transcriptome profile is sequenced, gene set enrichment and gene network analyses are performed to uncover the most-impacted pathways and to identify the transcription regulators (TR) involved.

In chapter 4, Pathway Interaction Analysis (PIA) package is introduced. PIA is an in-house tool, entirely build in R, useful for data mining of genomic and transcriptomic outcomes. In particular, PIA helps to infer possible functional candidates among a list of significant genes, extending the concept of classical pathway analysis and taking into account the investigation of relations among multiple pathways.

3. List of publications

Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays.

Palombo V, Milanesi M, Sgorlon S, Capomaccio S, Mele M, Nicolazzi E, Ajmone-Marsan P, Pilla F, Stefanon B, D'Andrea M.

J Dairy Sci. 2018 Sep 19. pii: S0022-0302(18)30869-5. doi: 10.3168/jds.2018-14413.

Transcriptional profiling of swine mammary gland during the transition from colostrogenesis to lactogenesis using RNA sequencing.

Palombo V, Loor JJ, D'Andrea M, Vailati-Riboni M, Shahzad K, Krogh U, Theil PK.

BMC Genomics. 2018 May 3;19(1):322. doi: 10.1186/s12864-018-4719-5.

Chapter I - GENERAL INTRODUCTION

Abstract

Most genetic progress for quantitative traits in livestock has been made by selection on phenotype, i.e. on estimates of breeding values (EBV) derived from phenotype, through the application of best linear unbiased prediction (BLUP) methodology. In the second half of the 20th century, the advent of molecular biology provided new opportunities to enhance quantitative genetics and breeding programs. In particular, new advances in animal genotyping fostered the development of marker-assisted selection (MAS) and recently of genomic selection (GS), producing positive genetic trends in many productive traits and leading ultimately to more accurate selection results and a faster genetic improvement across generations. Despite the successful integration of DNA information into several breeding programs, much more it is expected. In fact, the main traits of interest in livestock production are much more complex than expected so a deeper understanding of genome organization and information would further increase the accuracy of breeding evaluation. In this regard, fascinating opportunities are offered by the revolutionary advent of high-throughput 'omics' (HTO) technologies. The trademark characteristic of omic approach is its holistic capability: the staple is that a complex system can be understood more thoroughly if considered as a whole. This may open up new avenues to illuminate the biological mechanisms of important livestock complex traits and to explore relationships between genetic variation and phenotypic variability with high resolution.

Index of chapter

I - 1. Animal breeding 5
I - 1.1. Breeding value 5
I - 1.2. From quantitative to molecular genetics 6
I - 2. Omics sciences..8
I - 2.1. Genesis 8
I - 2.2. The paradigm shift 9
I - 2.3. Genomics 10
I - 2.4. Transcriptomics..11
I - 2.5. Bioinformatics..13
I - 3. 'Omic' technologies in animal breeding and genetics..14
I - 4. Aim of the thesis..15
I - 5. References..17

I - 1. Animal breeding

I - 1.1. Breeding value

Animal breeding is the selective mating of animals to increase the possibility of obtaining desired traits in the offspring and to lead to potential and stable gain for animal production as quickly as possible. The objective of livestock breeding has constantly changed since human beings began to breed livestock. Nevertheless, one of the goals has always been the acceleration of genetic gains to satisfy the demands of production and consumers (Yang et al., 2017), in terms of yield and recently also quality of animal products (Hocquette et al., 2005).

These production-related traits (such as milk yield, fat yield, protein yield, longevity, growth rate, fatness, feed intake, etc.) usually include a combination of multiple characteristics most of which have a quantitative nature, i.e. controlled by many genes as well as environmental factors (Falconer and Mackay, 1996). For this reason, statistical models and selection theory used in animal breeding are traditionally based on the so-called infinitesimal genetic model (Falconer and Mackay, 1996) that assumes a large (infinite) number of unlinked genes with very small and additive effects influencing the trait.

In animal breeding, the main criterion to identify candidates for mating (known as candidates selection) is the evaluation of their genetic merit, i.e. estimation of breeding values (BV). Based on infinitesimal model, the BV of an individual is defined as the sum of the additive effects of all loci that contribute to the trait (quantitative trait loci or QTL), deviated from the population mean (Falconer and Mackay, 1996). In this regard, extensive databases of recorded phenotypes for traits of interest have been used and, along with pedigree information, they traditionally represent the main source to estimate the BV of candidates under selection. With this aim, sophisticated statistical methods mainly based on best linear unbiased prediction (BLUP) mixed linear model methodology have been implemented (Henderson, 1984). These methods capitalize on information contained in the recorded phenotypes of not only the individual itself but also of its relatives, in order to maximize the accuracy of the resulting estimated breeding value (EBV).

I - 1.2. From quantitative to molecular genetics

Although selection programs based on traditional BV estimated from phenotype have been very successful, they also face a number of limitations. These primarily relate to (1) the ability to routinely record phenotypes on selection candidates and/or their close relatives in a timely manner, such that accurate selection decisions can be made at an early age reducing ultimately the generation intervals. (2) The cost of phenotype recording and (3) the nature of phenotype itself. In fact, many traits of interest are only recorded late in life (e.g. longevity) or only on one sex (e.g. milk yield), require animals to be sacrificed (e.g. meat quality) or require animals to be exposed to conditions that would hamper the ability to market or export their germplasm (e.g. disease resistance).

In the 1970's the advent of molecular biology provided new opportunities to enhance quantitative genetics and breeding programs in livestock. Particularly, nucleic acid-based markers has had a great impact on gene mapping, allowing identification of the underlying genes that control part of the variability of traits. For this reason, with the rapid development of DNA marker genotyping technologies, animal breeding has moved from conventional breeding to molecular breeding and has leaded to marker-assisted selection (MAS) (Dekkers, 2004), i.e. selection on a combination of information derived from the traditional phenotypic information and genetic markers.

MAS refers to the use of DNA markers that are tightly linked to target loci (associated with quantitative traits loci - QTL) as a substitute for or to assist phenotypic screening. In other words, an individual may be identified based on its genotype rather than its phenotype. This may greatly increase the efficiency and accelerate the breeding programs. For example, time and labour savings may arise from the substitution of difficult or time-consuming on-farm trials (that are conducted in particular standardized condition or are technically complicated) with DNA marker tests. Furthermore, selection based on DNA markers may be more reliable due to the influence of environmental factors on farm tests.

Thanks to increasing of DNA marker genotyping devices, a large numbers of candidate gene and QTL mapping studies were conducted (Andersson, 2001; Dekkers and Hospital, 2002). This resulted in the discovery of substantial numbers of QTL, marker-phenotype associations, causative mutations or genomic regions that affect quantitative traits (Grisart et al., 2002; Dekkers, 2004) providing opportunities to enhance response to selection, in particular for traits that are difficult to improve by conventional selection.

In general, although molecular genetic information is used in industry programs and is growing, the extent of use has not lived up to initial expectations. The implementation of molecular information in breeding programs, was limited for various reasons, particularly because most QTL studies were conducted in experimental crosses to create extensive *linkage disequilibrium*, rather than in the populations that are used for genetic improvement (Dekkers, 2004). For more details on *linkage disequilibrium* see chapter 2 paragraph 1.4 (‘*The use of genetic information to enhance the response to selection*’). Furthermore, the use of molecular information requires a comprehensive integrated strategy that must be closely aligned with business on which the cost of routine genotyping of selection candidates still affects (Dekkers, 2004).

In this regard, new opportunities are offered by the development of high-throughput (HT) commercial platforms for genotyping, which have marked the beginning of ‘omic’ analyses.

This new scenario allows exploring the genome looking for QTLs and associations between molecular markers and phenotypes with high resolution, providing information to estimate genomic breeding value (GEBV), and allowed ultimately the implementation of genomic or whole-genome selection (GS) that can be considered as a MAS on a genome-wide scale (Meuwissen, 2007). For more details see paragraph 3 (‘*Omic’ technologies in animal breeding and genetics*’) and chapter 2 (‘*Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using SNP arrays*’).

I - 2. Omics sciences

I - 2.1. Genesis

In 1977, Sanger and colleagues published the dideoxynucleotide method for DNA sequencing (Sanger et al., 1977). This method revolutionized the biology and culminated in 2001 in a milestone for human history: the completion of the human genome sequence. The Human Genome Sequence Project (HGSP) was the result of an extensive international effort and cost about 2.7$ billion for over 13 years of intensive work (https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome). Although 'Sanger sequencing' resulted in many technical improvements in throughput, accuracy, safety, robustness and sensitivity over the HGSP years, it remained a high-cost and time-consuming method. Sequencing a complex genome with Sanger technology is estimated to cost about €25 million for several years of work (https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome). For this reason, between 2005 and 2007, driven by growing demands of research in human genetics, agriculture and environmental sciences, Roche, Illumina (ex Solexa) and Applied Biosystems developed innovative sequencing technologies, known as next-generation sequencing (NGS).

Staple of NGS technologies is their ability to sequence massive amounts of templates in parallel, producing millions of reads in one run and with a relative low cost: the era of the Sanger monopoly on sequencing was over. Along with the benefits of cost-effective DNA sequencing with a revolutionary depth, scale and throughput (the whole genome sequence is now available for most species), the new DNA sequence information encouraged the development of powerful molecular biology tools for genome-wide analysis. Those tools, known as high-throughput (HT) devices, provided further impulse to novel applications in analysis of (1) genome-wide genetic variation and (2) gene expression by transcriptomic profiling. This new scenario motivated the researchers to change their research perspective and address biological questions on a genome-wide scale. They progressively shifted from a traditional candidate gene approach to a more new and holist approach that considers thousands of genes together or even whole genome. Ultimately the massive capacity of HT technologies has led to a paradigm shift from gene by gene analyses to 'omic' analyses covering the whole genome, exome or transcriptome (Goldman and Domschke, 2014). The 'omic era' has definitely begun.

I - 2.2. The paradigm shift

The addition of 'omic' to a molecular term implies a comprehensive, or global, detection of a set of molecules which contain part of the information related to the biological system under study (Hasin et al., 2017 Vailati-Riboni et al., 2017): genes (genomics), RNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics). The trademark characteristic of omics sciences is their holistic capability. The staple is that a complex system can be understood more thoroughly if considered as a whole (Vailati-Riboni et al., 2017). In this sense, the omics approach has twin advantages: (1) is applicable to well-known scenarios to study the deep connections and interrelationships among the many faces of a complex physiological state and to discover missing pieces in the current knowledge (Vailati-Riboni et al., 2017). (2) Is suitable for hypothesis-free experiments, i.e. situations when no hypothesis is known or prescribed due to lack of data. In this case, the holistic approach, based on acquisition and analysis of all available data, helps to define a preliminary hypothesis, which can be further tested (Vailati-Riboni et al., 2017).

In this regard, it must be emphasized that investigating biological phenomena at the 'omics scale' come with the need for implementing a novel *modus operandi*. The challenge is to address data generation, analysis and sharing from a 'larger perspective'. In fact, it is worth noting that the revolutionary gain of omic data relies on being analysed and interpreted as a whole through effective and integrative pipelines (integrated (multi)omics approaches). This clearly requires the cooperation of multidisciplinary teams as well as the fundamental support of bioinformatics and biostatistics (Manzoni et al., 2018), which have a key and prominent role in omic research (Yadav, 2015). This new scenario lead to the birth of the 'Systems Biology' which is an inter-disciplinary study based on 'omic' technologies and is concerned with understanding the dynamic outcome of molecular interactions among biomolecules at pathway, cellular network, cell, tissue and organismal levels (Berry et al., 2011). Instead of analysing individual components or aspects of the organism, such as the response of a single cell type to a specific disease, systems biologists focus on all the components and the interactions among them, all as part of one system (Berry et al., 2011). While genomics, transcriptomics and proteomics, coupled with bioinformatics, are gaining momentum, they are still, for the most part, assessed individually with distinct approaches generating monothematic rather than integrated knowledge (Manzoni et al., 2018).

I - 2.3. Genomics

The first omics discipline to appear was 'genomics', which is the study organisms' whole genome. Nowadays, high quality reference genome sequences are available for model species and economically important agricultural animals; a revolution built on the shoulders of the HGSP that has provided an invaluable resource for genomic studies in several fields (Dekkers, 2012; Van Emon, 2016). In particular, the completion of many species genome provided a very useful framework for mapping and studying specific genetic variants contributing to phenotypic variation of complex traits of interest (Manzoni et al., 2018). In this regard, many genetic variants exist in genomes and can be broadly categorized into two groups: simple nucleotide variations (SNVs) and structural variations (SVs). The former comprises single nucleotide variations and small insertion/deletions (indels) (known as single nucleotide polymorphism - SNP); the latter includes large indels, copy number variants (CNVs) and inversions (Metzker, 2010).

Several techniques to capture genetic variants are available and include (1) 'Sanger sequencing', the base-by-base sequencing of a locus of interest. (2) 'DNA-microarrays', based on hybridization of the DNA sample with a set of pre-defined oligonucleotide probes distributed across the entire genome or enriched around regions of interest. (3) 'Next-generation sequencing' (NGS) methods based on the fragmentation of the genomic DNA into pieces that are subsequently sequenced and aligned to a reference sequence (Manzoni et al., 2018). In this context, genome-wide association studies (GWAS) represent the gold standard method to exploit the biological information associated with DNA variants. Broadly, GWAS analyse DNA variants focusing on their inheritance and ultimately identify candidate loci associated with quantitative traits (QTL) in a hypothesis-free discovery study.

More specifically, a typical GWAS design involves using a SNP-based microarray to genotype a cohort of interest. GWAS arrays are probe-based chips with a large number of SNP markers spread across the genome (encompassing 10,000 up to 2,000,000 SNPs), having the capability to perform high-throughput genotyping for large scale samples (Fan et al., 2010). Such coverage ensures that any QTL will be closely linked with at least one marker. The expected result is a list of significant SNPs, evaluated for their *linkage disequilibrium*, i.e. the correlation structure that exists among DNA variants in a genome, in relation to the trait under study using the most appropriate statistical model in a given population. For more details see chapter 2 ('*Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using SNP arrays*'). The ultimate goal is to link genotypic variations to corresponding

differences in phenotype and ultimately to identify locations in the genome (i.e. genes) that harbour variability with possible influence on traits of interest (Visscher et al., 2017). It is widely reported in literature that significant SNPs in GWAS mostly fall inside intronic or intergenic regions (Manolio et al., 2009), i.e. the association typically identifies variants supposed to affect DNA structure and gene expression rather than protein sequence. For this reason, it is clear that balancing the statistical evidence of SNP-based genotype/phenotype correlation with prior evidence of biological relevance is the challenging problem (Saccone et al., 2008). This means that the statistically significant associations require prioritization with follow-up studies. In this regard, many methods are available for the prioritizing GWAS results and for handling the GWAS statistical challenges (Cantor et al., 2010; Sobota et al., 2015). Among others, some methods have recently been proposed using a priori knowledge, as the 'candidate pathway' analysis (Raven et al., 2013), and the 'gene-based' association strategies (Cantor et al., 2010; Liu et al., 2010; Akula et al., 2011). For more details see chapter 4 and 2 ('*PIA (Pathways Interaction Analysis): an R tool for analysing and interpreting high-throughput data'* and *'Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using SNP arrays*').

I - 2.4. Transcriptomics

The science of the genomes, or 'genomics', initially dedicated to the determination of DNA sequences and its significant variants has promptly expanded toward a more functional level: study of the expression profiles, i.e. 'transcriptome'. The transcriptome is the total RNA (i.e., mRNA, noncoding RNA, rRNA, and tRNA) expressed by a cell or tissue, and represents a snapshot of cellular metabolism (Vailati-Riboni et al., 2017). In particular, the study of messenger RNA (mRNA) dynamics in a given sample, facilitates a global understanding of the molecular changes in gene activation/suppression levels (gene expression differences) that controls the synthesis of proteins within the cell, which ultimately affect function and phenotype of an individual (Berry et al., 2011).

The first transcriptomic studies were started along with the development of the 'microarray' technology in the late '90s (Schena et al., 1995). The recent introduction of high-throughput next-generation DNA sequencing (NGS) technology has revolutionized transcriptomics by allowing RNA analysis through cDNA sequencing on a massive scale (Voelkerding et al.,

2009). RNASeq has now displaced microarrays as the preferred method for gene expression profiling (Kukurba and Montgomery, 2015; Costa-Silva et al., 2017).
The most popular technology for RNASeq has been the Illumina Genome Analyzer and Hi-Seq (http://www.illumina.com). Since its introduction in 2007, sequencing technology has steadily increased read length and overall number of reads generated per run, enabling higher levels of mappability (McGettigan, 2013). The latest and innovative devices also enable better identification and mapping of spliced reads as well as enabling the assembly of transcriptomes in the absence of a reference genome using *de novo* assembly approaches (McGettigan, 2013). The principal goal of RNASeq analysis is to identify key differentially expressed genes (DEGs) involved in specific physiological conditions (Costa-Silva et al., 2017). To detect differential expression, a variety of statistical methods have been designed specifically for RNASeq data. Nowadays, well-structured bioinformatics ready to use tools are freely available (Anders and Huber, 2010; Hardcastle and Kelly, 2010; Robinson et al., 2010; Wang et al., 2010; Trapnell et al., 2013). For more details see chapter 3 ('*Transcriptional profiling of swine mammary gland during the transition from colostrogenesis to lactogenesis using RNA sequencing*').
Gene expression studies can also be used for the detection or validation of potential QTL (Berry et al., 2011). Simplistically, a GWA study suggests potential biological processes associated with a trait to be subsequently investigated in functional work (Pearson and Manolio, 2008; Edwards et al., 2013). A more sophisticated approach is seen in the 'genetical-genomics' study (Jansen and Nap, 2001). 'Genetical genomics' is a term coined by Jansen and Nap (2001) to describe the marrying of genetic mapping methodology with gene expression data by combining a genome-wide study of gene expression with a genome wide scan of loci controlling variation in gene expression (Berry et al., 2011). This approach can be used to dissect gene expression differences among individuals into genetic and non-genetic components using populations and methods similar to those used for QTL mapping in GWAS, but instead of actual phenotypes, the dependent variable is the expression levels of multiple transcripts (Berry et al., 2011). The ultimate goal is to identify significant loci explaining a fraction of the genetic variance of a gene expression phenotype, i.e. expression QTLs (eQTLS) (Nica and Dermitzakis, 2013). One current limitation of genetical-genomics approaches for QTL detection is the cost of undertaking genome wide expression profiles on a large number of individuals to gain sufficient statistical power. However, this limitation will diminish as the cost of expression profiling reduces.

I - 2.5. Bioinformatics

The science of bioinformatics has developed in the wake of methods to determine the sequences of the informational macromolecules (mainly DNA and RNA). With the advent of whole genome sequencing and the related HT platforms that offer genome-wide information, bioinformatics has grown into the scientific field of management and analysis of biological information (Eisenberg et al., 2006). Bioinformaticians use computers and statistics to perform 'omic-research' in order to gain useful information from such comprehensive sets of data (Schneider and Orchard, 2011). In fact, it is obvious that without robust computational methods it is impossible to make sense of the huge data produced. This is so true that now, bioinformatics-based applications have been tightly incorporated in all omics research. Bioinformatics analyses include a huge number of applications continuously developed and updated, with the goal of (1) data processing and molecule identification, (2) statistical data analysis, (3) pathway analysis, and (4) data modelling in a system wide context (Schneider and Orchard, 2011).

Among the others 'pathway analysis' deserve particular mention which represents the first choice to simplify the analysis of omic data, extracting meaning from the list of key outcomes and providing insights into the underlying biology of the state being studied (Khatri et al., 2012). In fact, to reduce the complexity of data mining challenges, one common approach is to simplify the analysis by grouping long lists of individual genes into smaller sets of related genes sharing the same biological processes and/or molecular functions (i.e. pathways). This method of analysing high-throughput data has become popular during the last few years and it is known as 'functional enrichment' or 'pathway analysis' (Curtis et al., 2005) This approach is driven by increasing availability of public pathway knowledge based on hierarchical classification of terms (i.e. gene ontology - GO). The GO contains standardized annotation of gene products and has become the *de facto* standard for the secondary analysis of high throughput experiments (Khatri and Drăghici, 2005). Many sources of pathway and functional information, which can be either generic or species-specific, are now available and, at the same time, a large number of tools for pathway analysis have been developed, based on increasing availability of gene annotations databases (Berg et al., 2009a). For more details see chapter 4 ('*PIA (Pathways Interaction Analysis): an R tool for analysing and interpreting high-throughput data).*

The rise of a high number of bioinformatics tools has fostered initiatives aimed at generating web portals to list them and support their effective use. For example, EMBL-EBI has a bioinformatics service portal listing a variety of databases and tools tailored for specific quests

or topics (McWilliam et al., 2013); OMICtools is a library of software, databases and platforms for big-data processing and analysis (Henry et al., 2014); ExPASy is a library particularly renowned for proteomics tools (Gasteiger et al., 2003). In this context, it is worth noting that R-environment programming is gaining immense importance. R is an open-source programming language created in 1995 (Ihaka and Gentleman, 1996a) and is the *de facto* standard for the development of tools and *ad hoc* scripts useful for a variety of analyses and bioinformatics solutions, often shared on free repositories such as CRAN (https://www.R-project.org/), Bioconductor (http://www.bioconductor.org/) and GitHUB (https://github.com/). The flourishing of all of those analytic tools and software is remarkable, and increases the speed at which data can be processed and analysed. However, with this abundance of possibilities, caution is warranted, as no single tool is comprehensive and none is infallible. It is imperative to understand the principles behind bioinformatics tools and to choose the most suitable ones for the purposes of the end user's projects (Manzoni et al., 2018).

I - 3. 'Omic' technologies in animal breeding and genetics

To date most applications of 'omic' technologies in animal breeding have been through genomics (Berry et al., 2011). Several genotyping solutions on a wide-scale are currently available for most livestock species at a reasonable cost. Those platforms allow the analysis of an individual for tens of thousands of SNP across the genome in one single analysis. The first such high-density SNP genotyping platform available in livestock was the 50K Bovine Illumina SNP panel (Matukumalli et al., 2009). Similar SNP panels are now available for other livestock species, including pigs, poultry, sheep and horse. Recently, panels with over 700K SNP have become available in cattle and such higher density panels are also under development in other species.

Particularly in dairy cattle, the main use of high-density SNP genotyping has been to implement genomic or whole-genome selection (GS). In 2001, Meuwissen et al. (2001) proposed the GS, in which an estimated genomic breeding value (GEBV) was used to select a suitable breeding strategy. GEBV considers the effect of each SNP on the high-density (HD) array using models that fit all SNP simultaneously (Dekkers, 2012), i.e. potentially capturing all the quantitative trait loci (QTL) that contribute to the variation in a phenotype (Hayes et al., 2009). At the same time, HD SNP genotypes can be useful to construct so-called genomic relationship matrix among individuals, as an alternative to traditional pedigree-based relationship matrix for BLUP

model used to estimate BVs in livestock (Clark and van der Werf, 2013). Estimation of BV using HD SNP data has been implemented in dairy cattle breeding programs in several countries. Research is also underway with the goal of implementing genomic selection in other livestock species (Dekkers, 2010). It is estimated that by using GS strategy, the rates of genetic improvement in sheep and dairy cows could be increased by 20–100% (Yang et al., 2017) and that the generation interval can be reduced to 1.5 years compared to current 5-6 years in dairy cattle (Pryce and Daetwyler, 2012). This because GS strategy allows selection of young candidates for breeding prior to the availability of extensive progeny data (Dekkers, 2012).
Thanks to increasing of availability of HD genotyping solutions, nowadays, MAS and GS have become mainstream practices in molecular breeding of livestock (Yang et al., 2017), nevertheless the near future is represented by the use of next generation sequencing (NGS) for animal genotyping. This may open up new avenues to explore relationships between genetic and phenotypic diversity with high resolution. In fact, NGS data analysis provides a clear advantage over HD arrays, as it is not bound by the extent of *linkage disequilibrium* between SNP markers and the causal mutation but the causal mutation is in the data itself (Sharma et al., 2017). Although NGS technologies open a promising and interesting perspective and their use in animal breeding is an active field of research, its advantages and drawbacks are yet to be seen in practical situations before becoming a standard practice in livestock breeding (Sharma et al., 2017). Overall, genomic prediction of production in crossbreeding and across-breed schemes, costs and choice of individuals for genotyping are reasons for a reluctance to fully rely on genomic information for selection decisions (Jonas and Koning, 2015). Nevertheless, what is certain is that omics sciences, built on possibilities offered by HT and NGS technologies, represent the near future of animal breeding and genetics; the best method to illuminate the biological mechanisms of complex economic traits and to accurately predict their phenotypic variations (Yang et al., 2017). It is expected that population-personalized multi-omics livestock breeding will be realized in the future (Yang et al., 2017) and this will allow breeders to achieve rapid, accurate, and selective breeding of livestock according to their breeding objectives.

I - 4. Aim of the thesis

The general aim of this thesis is to explore the biology of livestock complex traits, such as lipid metabolism and colostrogenesis/lactogenesis transition respectively in bovine and pig species.

In particular, the specific objectives are to (1) investigate the differences of genetic basis related to the milk fatty acids profiles in two Italian dairy cattle breeds and (2) delineate the genes and transcription regulators implicated in the control of the transition from colostrogenesis to lactogenesis in swine, using the state-of-art genomic and transcriptomic analyses.

For these reasons, a GWAS analysis is run to investigate milk fatty acid profile traits in Italian Holstein and Italian Simmental breeds. Data are analysed with a well-established method implemented in the GenABEL R package and with the MUGBAS gene-based association method. Association is investigated between 120K SNPs and 85 fatty acids (as single, aggregated or index values). In addition, an RNASeq analysis is performed to investigate the mechanism of transition from colostrogenesis to lactogenesis in swine. For this reason, the mammary tissue is collected from three sows in five different time points close to parturition. Once the transcriptome profile is sequenced, gene set enrichment and gene network analyses are performed to uncover the most-impacted pathways and to identify the transcription regulators involved.

Along with objectives and related studies introduced above, an in-house bioinformatics tool performing an original pathway analysis is presented. The tool, entirely built in R and named PIA (Pathways Interaction Analysis), is designed for post-genomic and transcriptomic data mining. In particular, PIA helps to infer possible functional candidates among a list of significant genes, extending the concept of classical pathway analysis and taking into account the investigation of relations among multiple pathways.

I - 5. References

Akula, N., A. Baranova, D. Seto, J. Solka, M.A. Nalls, A. Singleton, L. Ferrucci, T. Tanaka, S. Bandinelli, Y.S. Cho, Y.J. Kim, J.-Y. Lee, B.-G. Han, B.D.G.S. (BiGS) Consortium, T.W.T.C.-C. Consortium, and F.J. McMahon. 2011. A network-based approach to prioritize results from genome-wide association studies. PLOS ONE 6:e24220. doi:10.1371/journal.pone.0024220.

Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106. doi:10.1186/gb-2010-11-10-r106.

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. 2:130–138. doi:10.1038/35052563.

Berg, B.H. van den, C. Thanthiriwatte, P. Manda, and S.M. Bridges. 2009. Comparing gene annotation enrichment tools for functional modeling of agricultural microarray data. BMC Bioinformatics 10:S9. doi:10.1186/1471-2105-10-S11-S9.

Berry, D.P., K.G. Meade, M.P. Mullen, S. Butler, M.G. Diskin, D. Morris, and C.J. Creevey. 2011. The integration of 'omic' disciplines and systems biology in cattle breeding. animal 5:493–505. doi:10.1017/S1751731110002120.

Cantor, R.M., K. Lange, and J.S. Sinsheimer. 2010. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. Am. J. Hum. Genet. 86:6–22. doi:10.1016/j.ajhg.2009.11.017.

Clark, S.A., and J. van der Werf. 2013. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. Methods Mol. Biol. Clifton NJ 1019:321–330. doi:10.1007/978-1-62703-447-0_13.

Costa-Silva, J., D. Domingues, and F.M. Lopes. 2017. RNA-Seq differential expression analysis: An extended review and a software tool. PLOS ONE 12:e0190152. doi:10.1371/journal.pone.0190152.

Curtis, R.K., M. Oresic, and A. Vidal-Puig. 2005. Pathways to the analysis of microarray data. Trends Biotechnol. 23:429–435. doi:10.1016/j.tibtech.2005.05.011.

Dekkers, J.C.M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3:22–32. doi:10.1038/nrg701.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82 E-Suppl:E313-328. doi:10.2527/2004.8213_supplE313x.

Dekkers, J.C.M. 2010. Use of high-density marker genotyping for genetic improvement of livestock by genomic selection.. CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour. 5:1–13.

Dekkers, J.C.M. 2012. Application of genomics tools to Animal Breeding. Curr. Genomics 13:207–212. doi:10.2174/138920212800543057.

Edwards, S.L., J. Beesley, J.D. French, and A.M. Dunning. 2013. Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. 93:779–797. doi:10.1016/j.ajhg.2013.10.012.

Eisenberg, D., E. Marcotte, A.D. McLachlan, and M. Pellegrini. 2006. Bioinformatic challenges for the next decade(s). Philos. Trans. R. Soc. B Biol. Sci. 361:525–527. doi:10.1098/rstb.2005.1797.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics, 4th Edition. Harlow UK Longman Xii 438.

Fan, B., Z.-Q. Du, D.M. Gorbach, and M.F. Rothschild. 2010. Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Australas. J. Anim. Sci. 23:833–847. doi:2010.23.7.833.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, and A. Bairoch. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31:3784–3788.

Goldman, D., and K. Domschke. 2014. Making sense of deep sequencing. Int. J. Neuropsychopharmacol. 17:1717–1725. doi:10.1017/S1461145714000789.

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

Hardcastle, T.J., and K.A. Kelly. 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422. doi:10.1186/1471-2105-11-422.

Hasin, Y., M. Seldin, and A. Lusis. 2017. Multi-omics approaches to disease. Genome Biol. 18:83. doi:10.1186/s13059-017-1215-1.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646.

Henderson, C.R. 1984. Applications of linear models in animal breeding. Appl. Linear Models Anim. Breed.

Henry, V.J., A.E. Bandrowski, A.-S. Pepin, B.J. Gonzalez, and A. Desfeux. 2014. OMICtools: an informative directory for multi-omic data analysis. Database J. Biol. Databases Curation 2014. doi:10.1093/database/bau069.

Hocquette, J.-F., R.I. Richardson, S. Prache, F. Medale, G. Duffy, and N.D. Scollan. 2005. The future trends for research on quality and safety of animal products. Ital. J. Anim. Sci. 4:49–72. doi:10.4081/ijas.2005.3s.49.

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. 5:299–314. doi:10.1080/10618600.1996.10474713.

Jansen, R.C., and J.-P. Nap. 2001. Genetical genomics: the added value from segregation. Trends Genet. 17:388–391. doi:10.1016/S0168-9525(01)02310-1.

Jonas, E., and D.-J. de Koning. 2015. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front. Genet. 6. doi:10.3389/fgene.2015.00049.

Khatri, P., and S. Drăghici. 2005. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinforma. Oxf. Engl. 21:3587–3595. doi:10.1093/bioinformatics/bti565.

Khatri, P., M. Sirota, and A.J. Butte. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput. Biol. 8:e1002375. doi:10.1371/journal.pcbi.1002375.

Kukurba, K.R., and S.B. Montgomery. 2015. RNA Sequencing and Analysis. Cold Spring Harb. Protoc. 2015:951–969. doi:10.1101/pdb.top084970.

Liu, J.Z., A.F. McRae, D.R. Nyholt, S.E. Medland, N.R. Wray, K.M. Brown, AMFS Investigators, N.K. Hayward, G.W. Montgomery, P.M. Visscher, N.G. Martin, and S. Macgregor. 2010. A versatile gene-based test for genome-wide association studies. Am. J. Hum. Genet. 87:139–145. doi:10.1016/j.ajhg.2010.06.009.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F.C. Mackay, S.A. McCarroll, and P.M. Visscher. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753. doi:10.1038/nature08494.

Manzoni, C., D.A. Kia, J. Vandrovcova, J. Hardy, N.W. Wood, P.A. Lewis, and R. Ferrari. 2018. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief. Bioinform. 19:286–302. doi:10.1093/bib/bbw114.

Matukumalli, L.K., C.T. Lawley, R.D. Schnabel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O'Connell, S.S. Moore, T.P.L. Smith, T.S. Sonstegard, and C.P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS One 4:e5350. doi:10.1371/journal.pone.0005350.

McGettigan, P.A. 2013. Transcriptomics in the RNA-seq era. Curr. Opin. Chem. Biol. 17:4–11. doi:10.1016/j.cbpa.2012.12.008.

McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y.M. Park, N. Buso, A.P. Cowley, and R. Lopez. 2013. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 41:W597-600. doi:10.1093/nar/gkt376.

Metzker, M.L. 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11:31–46. doi:10.1038/nrg2626.

Meuwissen, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie 124:321–322. doi:10.1111/j.1439-0388.2007.00708.x.

Nica, A.C., and E.T. Dermitzakis. 2013. Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. B Biol. Sci. 368. doi:10.1098/rstb.2012.0362.

Pearson, T.A., and T.A. Manolio. 2008. How to interpret a genome-wide association study. JAMA 299:1335–1344. doi:10.1001/jama.299.11.1335.

Pryce, J.E., and H.D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim. Prod. Sci.

Raven, L.-A., B.G. Cocks, J.E. Pryce, J.J. Cottrell, and B.J. Hayes. 2013. Genes of the RNASE5 pathway contain SNP associated with milk production traits in dairy cattle. Genet. Sel. Evol. GSE 45:25. doi:10.1186/1297-9686-45-25.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. 26:139–140. doi:10.1093/bioinformatics/btp616.

Saccone, S.F., N.L. Saccone, G.E. Swan, P.A.F. Madden, A.M. Goate, J.P. Rice, and L.J. Bierut. 2008. Systematic biological prioritization after a genome-wide association study: an application to nicotine dependence. Bioinforma. Oxf. Engl. 24:1805–1811. doi:10.1093/bioinformatics/btn315.

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

Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470.

Schneider, M.V., and S. Orchard. 2011. Omics technologies, data and bioinformatics principles. Methods Mol. Biol. Clifton NJ 719:3–30. doi:10.1007/978-1-61779-027-0_1.

Sharma, A., J.-E. Park, H.-H. Chai, G.-W. Jang, S.-H. Lee, and D. Lim. 2017. Next generation sequencing in livestock species- A Review. J. Anim. Breed. Genomics 1. doi:10.12972/jabng.20170003.

Sobota, R.S., D. Shriner, N. Kodaman, R. Goodloe, W. Zheng, Y.-T. Gao, T.L. Edwards, C.I. Amos, and S.M. Williams. 2015. Addressing population-specific multiple testing burdens in genetic association studies. Ann. Hum. Genet. 79:136–147. doi:10.1111/ahg.12095.

Trapnell, C., D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, and L. Pachter. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31. doi:10.1038/nbt.2450.

Vailati-Riboni, M., Palombo V., and Loor J.J. 2017. What are omics sciences? Periparturient Diseases of Dairy Cows: A Systems Biology Approach pp. 1-7.

Van Emon, J.M. 2016. The omics revolution in agricultural research. J. Agric. Food Chem. 64:36–44. doi:10.1021/acs.jafc.5b04515.

Visscher, P.M., N.R. Wray, Q. Zhang, P. Sklar, M.I. McCarthy, M.A. Brown, and J. Yang. 2017. 10 years of GWAS discovery: biology, function, ad translation. Am. J. Hum. Genet. 101:5–22. doi:10.1016/j.ajhg.2017.06.005.

Voelkerding, K.V., S.A. Dames, and J.D. Durtschi. 2009. Next-Generation Sequencing: From Basic Research to Diagnostics. Clin. Chem. 55:641–658. doi:10.1373/clinchem.2008.112789.

Wang, L., Z. Feng, X. Wang, X. Wang, and X. Zhang. 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinforma. Oxf. Engl. 26:136–138. doi:10.1093/bioinformatics/btp612.

Yadav, D. 2015. Relevance of Bioinformatics in the Era of Omics Driven Research. J. Gener. Seq. Appl. 2:1–2. doi:10.4172/2469-9853.1000e102.

Yang, Y., R. Zhou, and K. Li. 2017. Future livestock breeding: Precision breeding based on multi-omics information and population personalization. J. Integr. Agric. 16:2784–2791. doi:10.1016/S2095-3119(17)61780-5.

Chapter II - GENOME-WIDE ASSOCIATION STUDY OF MILK FATTY ACID COMPOSITION IN ITALIAN SIMMENTAL AND ITALIAN HOLSTEIN COWS USING SNP ARRAYS

Abstract

Milk bovine is important for human nutrition but is often criticized due to its fat levels potential association with cardiovascular diseases. With regard to this risk, a selective breeding program could affect milk fatty acids (FA) composition, in order to improve the healthiness of human diets. In this thesis, a genome-wide association study (GWAS) on bovine milk was performed with the aim to identify genomic regions or genes associated with FA profile and investigate genetic differences between the Italian Simmental (IS) and Italian Holstein (IH) breeds. For this reason, milk from 416 IS and 436 IH cows were sampled and fat profile characterized by gas-chromatography. Subjects were genotyped with high and medium density SNP array and single-marker regression model to facilitate a genome-wide association study was performed.

The findings confirmed that several previously reported quantitative trait loci (QTL) are strongly associated with bovine milk fat composition. In particular, the GWAS resulted in 95 significant SNP associations with milk FA, with the strongest signals on BTA19 and BTA26. Further gene-centric approach and pathway meta-analysis identified significant candidate genes, and some well-known genes underlying QTL for milk FA components, such as *FASN*, *SCD* and *DGAT1*, but also other possible interesting genes, some of which had a functional role into pathways in interaction with ‘Lipid Metabolism’. The findings provide insights into the patterns of genes related to FA profile including *ECI2*, *PCYT2*, *DCXR*, *G6PC3*, *PYCR1* and *ALG12* in IS and *CYP17A1*, *ACO2*, *PI4K2A*, *GOT1*, *GPT*, *NT5C2*, *PDE6G*, *POLR3H* and *COX15* in IH. Overall, the breed-specific association outcomes reflected the differences of genetic background of Italian Simmental and Italian Holstein breeds and their selective breeding history.

Index of chapter

II - 1. Introduction 28
II - 1.1. Bovine milk fatty acids 28
II - 1.2. Factors influencing bovine milk fat 29
II - 1.3. Genes involved in bovine milk-fat composition 30
II - 1.4. The use of genetic information to enhance the response to selection 30
II - 1.5. Genome wide association studies (GWAS) 32
II - 1.6. Overcome the limitation of GWAS 33
II - 1.7. Italian Simmental and Italian Holstein dairy breeds 34
II - 2. Aim of the study 35
II - 3. Materials and methods 36
II - 3.1. Experimental population 36
II - 3.2. Fatty acid analysis 36
II - 3.3. Statistical analysis of phenotypes 37
II - 3.4. Genotyping and quality control 38
II - 3.5. Genome-wide association analysis 38
II - 3.6. Gene-based association analysis 39
II - 3.7. Pathways interaction analysis 40
II - 4. Results and discussion 40
II - 4.1. Phenotype and genotype statistics 40
II - 4.2. Single-population GWAS results 42
II - 4.3. Gene-based association analysis results 43
II - 4.4. Pathway analysis results 45
II - 4.5. Other significant genes 48
II - 5. Conclusion 50
II - 6. Figures and tables 52
II - 7. References 68

List of chapter figures and tables

Figure II-1. Manhattan plots of GWAS results showing the significance of SNP associations for C14:0 fatty acid.

Figure II-2. Manhattan plots of GWAS results showing the significance of SNP associations for C14:1c9 fatty acid.

Figure II-3. Manhattan plots of GWAS results showing the significance of SNP associations for ID 10-1/(10+10-1) fatty acid.

Figure II-4. Manhattan plots of GWAS results showing the significance of SNP associations for ID 14-1/(14+14-1) fatty acid.

Figure II-5. Manhattan plots of GWAS results showing the significance of SNP associations for C14:1c9 fatty acid.

Figure II-6. Manhattan plots of GWAS results showing the significance of SNP associations for ID 10-1/(10+10-1) fatty acid.

Figure II-7. Manhattan plots of GWAS results showing the significance of SNP associations for ID 14-1/(14+14-1) fatty acid.

Figure II-8. Manhattan plots of GWAS results showing the significance of SNP associations for milk fat percentage content (FP) in Italian Holstein (IH) breed.

Figure II-9. Diagram showing the results obtained with Pathway Interaction Analysis (PIA) on MUGBAS significant genes for the Italian Simmental (IS) breed.

Figure II-10. Diagram showing the results obtained with Pathway Interaction Analysis (PIA) on MUGBAS significant genes for the Italian Holstein (IH) breed.

Table II-1. Mean and standard deviation (SD) of single fatty acids.

Table II-2. Heritability (h^2) and standard errors (e) of single fatty acids.

Table II-3. Most significant SNPs associated with milk fatty acid traits in Italian Simmental (IS) and Italian Holstein (IH) breeds.

Table II-4. Most significant SNPs associated with milk fat percentage (FP) in Italian Holstein (IH) cows.

Table II-5. Genes obtained with PIA analysis considering three degrees of interaction with KEGG Lipid Metabolism pathways.

In supplementary material section

Supplementary Table II-S1. Significant genes obtained with MUGBAS analysis and associated with milk fatty acid (FA) traits in Italian Simmental (IS) and Italian Holstein (IH).

Supplementary Table II-S2. Significant genes obtained with MUGBAS analysis divided for each chromosome (Chr), breed and fatty acid (FA).

II - 1. Introduction

II - 1.1. Bovine milk fatty acids

Milk is a fluid secreted by the mammary glands of female mammals and provides a primary source of nutrition for the neonate. Raw milk contains fat, protein, lactose, vitamins, minerals and water. In addition to a natural source of nutrition for infant mammals, milk and dairy products are major components of the human diet in many parts of the world. In this regard, bovine milk plays a prominent role, being by far the principal type of milk used throughout the world and having long traditions in human nutrition (Fox, 2003). Among its components, fat is the main source of energy and influences the taste of milk and dairy products (Fenelon and Guinee, 1999; Esposito et al., 2014; Martini et al., 2016). About 98% of the total fat in milk is present in the form of triacylglycerols (Jensen, 2002), which are a single molecules of glycerol combined with three fatty acids (FA). In addition to the triacylglycerols, milk contains small amounts of diacylglycerols, monoacylglycerols, free (unesterified) FA and phospholipids. In this regard, bovine milk-fat is considered as one of the most complex fats with about 400 FA identified, although only few are abundant (Jensen, 2002). FA are named and categorized according to their length (i.e. number of carbon atoms) and to their degree of saturation (i.e. number of double bonds). Generically, FA without double bonds are called saturated (SFA), with one double bond are called monounsaturated (MUFA), and with two or more double bonds are called polyunsaturated (PUFA). Each single FA has a different structure and proprieties. The impact of dietary fat on human diseases has been investigated for decades (Hu and Willett, 2002). In this regard, milk fat is often criticized due to high percentage of its SFA content, claimed to increase blood cholesterol, heart disease, weight gain and obesity (Shingfield et al., 2013; Tullo et al., 2014; Pulina et al., 2017). Conversely, MUFA are considered to have a favourable effect on human health, because of their cholesterol-declining properties (Schwingshackl and Hoffmann, 2012). PUFA, especially the n-6 and n-3 series, are considered beneficial to human health and influence plasma lipids serving cardiac and endothelial functions for the prevention and the treatment of coronary heart diseases (FAO, 2010; Li et al., 2014a; Zhang et al., 2016). Furthermore, special attention is paid to Conjugated Linoleic Acids (CLA, long chain fatty acids) due to their supposed role on the modulation of plasma lipid concentration, also having anticarcinogenic and anti-inflammatory activity as shown *in vitro* and animal model studies (Parodi, 1999; Haug et al., 2007).

In dairy cows, typical milk fat content is about 70% SFA, 25% MUFA, and 5% PUFA, which considerably differs from the ideal fatty acid profile for human health (8% SFA, 82% MUFA and 10% PUFA) (Bilal et al., 2014). In order to improve the healthiness of human diets and to satisfy the consumer's demands, one of the animal breeding objectives is acting on milk FA profile, so that it is expected that a premium price for milk quality would include in the near future also the lipid specific composition (Tullo et al., 2014).

II - 1.2. Factors influencing bovine milk fat

The composition of milk fat is affected mainly by feeding factors (type, quantity and quality of forage) (Palmquist et al., 1993; Jensen, 2002; Palmquist, 2006a) nevertheless other sources of variability are well-known (Syrstad et al., 1982; German and Dillard, 2006; Stoop et al., 2009a). It is also documented that a significant part of the variability in FA composition is determined genetically (Soyeurt et al., 2007; Stoop et al., 2008). This has opened the possibility to optimize bovine milk-fat composition through selective breeding and not only by nutritional strategies. This possibility would represent a more permanent and reliable solution than other livestock production systems changes, offering a more consistent result to consumers (Mele et al., 2007; Schennink et al., 2009b; Bilal et al., 2014).

In general, the major prerequisite for selective breeding is the existence of genetic variation. In this regard, phenotypic variation in milk-fat composition has been documented over the years, both between and within breeds (Stull and Brown, 1964; Syrstad et al., 1982) and more recently many authors discussed cow's breed effect and genetic variability on milk composition (Soyeurt et al., 2006; Arnould and Soyeurt, 2009; Adamska et al., 2016). Estimating heritability values on FA traits is not an easy task, with studies often varying and discordant results (Soyeurt et al., 2007; Petrini et al., 2016). However, generally it is possible to conclude that the estimates reflect the common origin of groups of FA. In fact, the FA in milk arise from two sources: *de novo* synthesis in the mammary gland and plasma lipids. Short- and medium-chain (C4 to C16) SFA and MUFA are largely synthesized *de novo* in the mammary gland and show a moderate to high heritability. Whereas long-chain FA (C18 or more carbon atoms) are derived from circulating plasma lipids, which originate from diet and from body fat metabolism, and obviously show low to moderate heritability (Schennink et al., 2009b; Stoop et al., 2009b; Buitenhuis et al., 2014).

II - 1.3. Genes involved in bovine milk-fat composition

There are several genes known to be involved in FA synthesis and that explain part of the genetic variation of milk fat (Pegolo et al., 2016). In particular, causative variants in *diacylglycerol O-acyltransferase-1* (*DGAT1*) and *stearoyl-CoA desaturase-1* (*SCD*) genes have been documented in numerous studies (Grisart et al., 2002b; Moioli et al., 2007; Schennink et al., 2007, 2008). Specifically, a quantitative trait locus (QTL) mapping study in cattle (Grisart et al., 2002b) identified a polymorphism into the gene coding for DGAT1 (K232A) with a strong effect on milk-fat percentage and other milk-production characteristics. In *SCD1*, a non-synonymous SNP in exon 5 (A293V), has been associated with milk fatty acid composition in Italian Holstein, Piedmontese and Valdostana cattle breeds (Mele et al., 2007; Moioli et al., 2007). More recently, because of its known role in fat synthesis, *fatty acid synthase* (*FASN*) gene has been extensively studied as a candidate gene for fat content in milk (Schennink et al., 2009a; Bouwman et al., 2012). In particular it was the focus of studies designed to identify SNPs linked with the causative mutations for differences in milk FA composition (Li et al., 2016; Knutsen et al., 2018). Several SNPs in *FASN* resulted in association with milk FA traits, one of them was predicted to result in an amino acid substitution from threonine (ACC) to alanine (GCC) (Li et al., 2016). Although the large effect of these major genes on milk fat in dairy cattle is well-established, other genes are expected to explain fat composition variability because of the complexity of biological mechanism of FA synthesis.

II - 1.4. The use of genetic information to enhance the response to selection

Most of the traits with an economic interest in livestock have a complex quantitative expression affected by environmental factors and simultaneously coded by a large number of genes. Starting from this paradigm, the estimation of genetic merit (breeding value) of animal candidate for selection has been historically based on Fischer's infinitesimal model, according to which an infinite number of loci, each with an infinitesimal additive effect, affects observed phenotypes (Fisher, 1919). Traditionally, best linear unbiased predictors (BLUP) methodology has been used to estimate a breeding value for all the animals in the population starting from all sources of phenotypes and pedigree information. This approach has dominated quantitative genetics over the years and allowed us to reach high rates of genetic improvement in many livestock species. More recently, due to the application of advanced techniques in molecular

genetics, limited chromosomal regions influencing quantitative traits have been discovered (Lander and Botstein, 1989). This has suggested that the control of quantitative traits (quantitative trait loci – QTL) expression is under a more limited number of loci, i.e. few genes with large effect and many of small effect. Hayes and Goddard (Hayes and Goddard, 2001) enforced this hypothesis studying the distribution of QTLs effect in dairy cattle and swine. Starting from this novel evidence, several approaches have been indicated to integrate molecular information in current breeding programs using genetic markers (Dekkers, 2004). Three types of genetic markers useful to this purpose have been defined: (1) direct, (2) linkage disequilibrium (LD) and (3) linkage equilibrium (LE) markers (Dekkers, 2004). The direct markers code for a functional causative mutation, whereas the LD and LE markers are loci in population-wide LD or LE with the functional mutation, respectively. The classical definition of LD refers to the non-random association of alleles between two loci. If two (A and B) bi-allelic markers are considered, four haplotypes of markers are possible (A1_B1, A1_B2, A2_B1 and A2_B2). If the frequencies of alleles A1, A2, B1 and B2 in the population are all 0.5, then the frequencies of each of the four haplotypes in the population are expected to be 0.25. Any deviation of the haplotype frequencies from 0.25 is LD, i.e. the genes are not in random association. Several measures of LD are available (Lewontin, 1964; Hill and Robertson, 1968; Hill, 1981; Zhao et al., 2005), nevertheless statistic r^2 (Hill and Robertson, 1968) is preferred over the others as a measure of the extent of LD for bi-allelic markers:

$$r^2 = \frac{D^2}{freq(A1) \times freq(A2) \times freq(B1) \times freq(B2)}$$

where D is calculated as described by Hill (Hill, 1981).

In current breeding programs, the use of markers (particularly in LD with QTL) has opened the possibility to develop the so-called ‘marker assisted selection’ (MAS). MAS represents an opportunity to enhance the response to selection especially for low-heritability traits, or those whose phenotype is difficult and expensive to measure or is expressed later in age. Overall, MAS could help to increase the accuracy of breeding values estimated for young animals and reduce the generation interval, nevertheless its use has not yet deliver its expected benefits in commercial breeding programmes. In fact, although advances in molecular genetics have been able to explain part of the genetic variances due to QTL, the use of MAS have been limited by several reasons. Firstly, only a restricted number of causative gene mutations have been

identified. Secondly, it is obvious that marker effect needs to be re-estimated frequently, because of recombination affecting on decreasing of LD across generations. This limitation is particularly important when low-density marker maps are used, as frequently happened in the past. In this case, the QTL is mapped within very large confidence intervals and there is risk the selection is only on the marker and not directly on the QTL or gene.

II - 1.5. Genome wide association studies (GWAS)

Nowadays, the availability of high-throughput sequencing techniques allows us to discover thousands of single nucleotide polymorphism (SNP) spread across the whole genome in several livestock species. Currently, arrays for genotyping animals at about 1,000,000 marker loci are commercially available. These new techniques give rise to new opportunities for genetic evaluation of farm animals with a so called genome-wide approach (Meuwissen et al., 2001). This new advance firstly allows us to explore the genome looking for QTLs and associations between SNP and phenotypes with high resolution and ultimately provides information to estimate genomic breeding value (GEBV). In the former case we talk about genome-wide association studies (GWAS), while in the latter we refer to genomic selection (GS) that can be considered as a MAS on a genome-wide scale (Meuwissen, 2007). Currently, GWAS represents the most promising method for dissecting the biology that underlies complex traits (McCarthy et al., 2008). The rationale of GWAS is simple: find marker-trait associations exploiting the LD that exists between the causative mutation or QTL (which is ignored) and a very large number of markers spread across all the genome (whom positions are known). In this way, it is possible to pinpoint genomic regions carrying causal variants for any trait, with a high probability as well as high resolution. The association between markers and QTL arise because there are small segments of chromosome in the population descending from the same common ancestor. These chromosome segments, which trace back to the same common ancestor without intervening recombination, will carry identical marker alleles or marker haplotypes. If there is a QTL somewhere within the chromosome segment, they will also carry identical QTL alleles. Several statistical methodologies, which exploit these associations, are available (Gondro et al., 2013); single marker regression GWAS is the simplest. In fact, in a random mating population with no population structure, the association can be tested as:

$$\mathbf{y} = \mathbf{Wb} + \mathbf{X}g + \mathbf{e}$$

Where **y** is a vector of phenotypes, **W** is a design matrix assigning phenotype records to fixed effects, **b** is a vector of fixed effects, **X** is a design matrix allocating records to the marker effect, **g** is the effect of the marker and **e** is a vector of random deviates $e_{ij} \sim N(0, \sigma_e^2)$, where σ_e^2 is the error variance (Gondro et al., 2013). In this model the effect of the marker is treated as a fixed effect, and the model is additive, such that two copies of the second allele has twice as much effect as one copy, and no copies has zero effect. The underlying assumption here is that the marker will only affect the trait if it is in LD with an unobserved QTL. The null hypothesis is that the marker has no effect on the trait, while the alternative hypothesis is that the marker does affect the trait (because it is in LD with a QTL). The null hypothesis is rejected if $> F_{\alpha,\nu1,\nu2}$, where F is the F statistic calculated from the data and $F_{\alpha,\nu1,\nu2}$ is the value from an F distribution at α level of significance and ν1, ν2 degrees of freedom.

The F-value can be calculated as (Gondro et al., 2013):

$$F = \frac{(n-1)(\hat{g}X'y - \frac{1}{ny'y})}{y'y - \hat{g}X'y - \hat{u}1_n{}'y}$$

F-values can be transformed into p-values for comparison with significance thresholds. One common option is to adjust the significance level for the number of markers tested using a Bonferroni correction to obtain an experiment wise P-value of 0.05 (Gondro et al., 2013).

II - 1.6. Overcome the limitation of GWAS

Although nowadays GWAS is the principal tool used to reconnect the trait back to its underlying genetics, it presents some limitations connected with its nature: particularly multiple comparison problems and result portability. On one hand, in fact GWAS research testing for hundreds of thousands or even millions of SNPs simultaneously, have to pay a high statistical price. Typically, Bonferroni test is applied for GWAS multiple comparisons (Gondro et al., 2013) but it is often too conservative, so that many important loci may not pass the stringent criterion of the significance test, as well as its use in some cases is not considered completely appropriate (Chen and Liu, 2011). On the other hand the result portability across a population is limited by a variety of confounding factor, such as population structure, differential LD levels, breed specific selection targets and SNPs ascertainment bias (Clark et al., 2005). In this

context also the nature of SNP information itself could represent a limitation: e.g. one favourable allele may segregate in one breed and be fixated in another, the same allele segregates in both breeds but alleles may differ or the genetic background masks the effect of segregation (Capomaccio et al., 2015a). To improve results usability and low signal catching, various strategies can be applied. One could be using haplotypes instead of single markers, to better pinpoint the associated regions (Utsunomiya et al., 2017). Another option is the dissection of associated signals with post-GWAS analyses (Capomaccio et al., 2015b; Pegolo et al., 2018). With regards to this option, gene-based association strategies, that restrict GWA study only to genes and neighboring genomic regions (Liu et al., 2010; Capomaccio et al., 2015b; a), or pathway enrichment and network analysis for the prioritization of GWAS outcomes (Akula et al., 2011; Yoon et al., 2018) deserve particular mention. In this context, an interesting contribution may be represented by '*PIA: an R package for Pathways Interaction Analysis*' (see Chapter 4).

II - 1.7. Italian Simmental and Italian Holstein dairy breeds

The investigation of genetic differences in phenotypically diverse breeds is a recognized strategy to reveal genes and related pathways that underlie complex traits of interest (D'Andrea et al., 2011), including milk FA profiles (Buitenhuis et al., 2014). In this regard, Italian Simmental (IS) and Italian Holstein (IH) represent two breeds with different productive characteristics and divergent selective breeding history, as well as genetic backgrounds (Bomba et al., 2015; Marras et al., 2015). IS is a dual-purpose cattle type, well adaptable to extensive system such as in mountainous areas (www.anapri.it). IH is a dairy cow intensively selected for high yielding in intensive production systems (www.anafi.it). Little information is available in literature on the comparison between these two breeds in terms of milk FA composition, however, significant differences in milk FA content between the Polish Simmental and Polish Holstein Frisian were previously reported (Adamska et al., 2016). The identification of common genes involved in the control of the trait investigated between the two breeds is not an easy task due to breed-specific selection target, as mentioned before, that can lead to differences at the genomic level (Bomba et al., 2015; Marras et al., 2015). Moreover, alleles segregating in one breed may be fixed in the other or, even when the same alleles are segregating in both breeds, the genetic background may change their effects (Capomaccio et al., 2015a). At the same time,

it is known that multibreed GWAS helps QTL detection only if the same QTL are shared across breeds (van den Berg et al., 2016). For example, Raven and colleagues found decrease in power when Holstein and Jersey data were combined, compared with within-breed GWAS results (Raven et al., 2014).

II - 2. Aim of the study

Several GWAS on cattle production and morphological traits have been conducted and genomic regions associated to these traits have been found, but only a limited number found genes directly involved in milk FA biology and very few focused on Italian dairy breed populations (Capomaccio et al., 2015a; Macciotta et al., 2015). In this study a GWAS on bovine milk was performed with the aim to identify genomic regions or genes associated with FA profile and investigate genetic differences between IS and IH breeds. For this reason, milk from 416 IS and 436 IH cows were sampled and fat profile characterized. Subjects were genotyped with 150K SNP array and a single-marker regression model for GWAS was performed. In addition to the classical GWAS approach, with the aim to increase discovery power in both breeds, post-GWAS analyses were applied. First a gene-based approach, MUGBAS (MUlti species Gene-Based Association Suite) (Capomaccio et al., 2015b), and after the Pathways Interaction Analysis (PIA) (see chapter 4) were performed. The former used the single-SNP GWAS results to calculate a gene-wise p-value. Briefly, the gene-wise test statistic condenses p-values of a SNP subset (within gene boundaries) weighting local LD (Capomaccio et al., 2015b; a). As reported by Capomaccio and colleagues, the gene-centric approach improves the power of the analyses rescuing signal under the genome-wide threshold in single-SNP GWAS. The latter uses the list of significant genes obtained with MUGBAS and investigates their relations taking into account upstream and downstream pathways in interactions with those related to the trait of interest (i.e. ‘Lipid Metabolism pathways’). In this case, the gene is evaluated for its functionality in the pathways involved with the traits. This before mentioned pipeline was here applied on IS and IH lactating cows, to find new QTL and genes affecting breed-specific FA composition and to further elucidate the genetic differences or similarities between the two breeds.

II - 3. Materials and methods

II - 3.1. Experimental population

In this study 416 IS and 436 IH cows from 10 commercial farms were considered. Animals were selected from commercial dairy farms located in the North East part of the Po Valley (Italy) presenting homogeneous management and ration compositions. Farms were selected together with the local Farm and Breeder Association (Associazione Allevatori del Friuli Venezia Giulia, Codroipo, Italy; www.aafvg.it), which provided information of individual milk records, reproductive parameters and managerial aspects. The herd size ranged from 157 to 654 cows. The inclusion criteria considered for the cows was to be clinically healthy and preferentially in mid-lactation. The average days in milk (DIM) were 153 (± 70) and 167 (± 63) for IS and IH, respectively. All the lactating cows were housed in free stalls with cubicles and milking parlours and the management of the farms were similar. Cows had free access to water and an *ad libitum* total mixed ration (TMR), based on corn silage and formulated to cover nutrient requirements, was offered twice a day, after the morning and the afternoon milking.

The day of official milk recording of the Breeder Association, 100 ml of milk samples were collected in the parlour from each cow at the morning milking. An aliquot of 50 ml of milk was transferred into a tube containing preservative and was used for protein, fat, lactose analyses and for somatic cell count (SCC) determination. The other aliquot of milk was transferred to a tube without preservative, frozen within 2 hours and stored at -20°C for FA analyses. Peripheral blood samples were collected and stored at -20°C before DNA isolation. Animals were also classified for parity, DIM, milk yield, fat and protein percentage content; for both breeds data were provided by the National Breeder Associations (A.N.A.P.R.I. for IS and A.N.A.F.I. for IH).

II - 3.2. Fatty acid analysis

Milk fat was extracted according to Buccioni et al. (2010) and methyl esters of fatty acids (FAME) were prepared with a base-catalyzed transesterification according to Christie (1982). The FAME were separated and identified by gas-chromatography (Buccioni et al., 2015).

The desaturation index (DI) was calculated according to the following formulas:

$$DI = (cis\text{-}9\ 10{:}1) / (10{:}0 + cis\text{-}9\ 10{:}1)$$

$$DI = (cis\text{-}9\ 14{:}1) / (14{:}0 + cis\text{-}9\ 14{:}1)$$

$$DI = (cis\text{-}9\ 16{:}1) / (16{:}0 + cis\text{-}9\ 16{:}1)$$

$$DI = (cis\text{-}9\ 18{:}1) / (18{:}0 + cis\text{-}9\ 18{:}1)$$

$$DI = (cis\text{-}9,trans\text{-}11\ 18{:}2) / (trans\text{-}11\ 18{:}1 + cis\text{-}9,trans\text{-}11\ 18{:}2)$$

All results were expressed in grams per 100 grams of fatty acid (FA).

II - 3.3. Statistical analysis of phenotypes

FA traits were compared between the breeds with Welch Two Sample t-test R function to estimate significant differences (R Development Core Team, 2006). Variance components were calculated within breeds separately adopting a linear mixed model performed with MIXED procedure in SPSS software (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.) using the following model:

$$Y_{ijk} = \mu + yield_i + FKg_i + PKg_i + Log(SCC_i) + DIM_i + parity_j + farm_k + \varepsilon_{ijk}$$

where Y_{ijk} is the phenotype; μ is the overall mean; *yield* is the covariate effect of milk yield; *FKg* and *PKg* are the covariate effect of fat and protein content respectively; *SCC* is the covariate effect of somatic cell count; *DIM* is the covariate describing the effect of days in milk; *parity* is the fixed effect of calving, before statistical analysis animals were classified for parity with ordinal value of 1 for first calving, 2 for second calving and 3 for cows with more than 2 calving; *farm* is the random effect of farm distributed as $N(0, I\sigma^2_{farm})$, with identity matrix I and farm variance σ^2_{farm}; and ε is the random residual distributed as $N(0, I\sigma^2_{\varepsilon})$, with identity matrix I and farm variance σ^2_{ε}.

For each trait, the heritability was calculated by ASreml software (Gilmour, A.R. et al., 2009), using the same data and model described before. All phenotypic distributions were systematically diagnosed for normality using a Shapiro–Wilk test and non-normal phenotypes were adjusted by truncation of outliers (+/- 3 times SD and first five values +/- Q1Q3) or by log transformation.

II - 3.4. Genotyping and quality control

Genomic DNA was isolated from whole blood using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, California, USA). After extraction, quality and quantity of nucleic acid were assessed by electrophoresis and spectrophotometry. The DNA was stored at -20°C. Fifty ng of genomic DNA were sent to the genotyping facility (GeneSeek, Lincoln, NE) for marker analysis. A total of 152 IS subjects were genotyped with BovineHD Genotyping BeadChip (BovineHD, 777,000 SNP) (Illumina, San Diego, CA); the rest of IS subjects were genotyped with GeneSeek GGP Bovine 150K array from Neogene (Bovine150K, 150,000 SNP). All IH subjects were genotyped with GeneSeek GGP Bovine 150K array from Neogene (Bovine150K,). Before the imputation phase, genotypes were quality controlled excluding markers with more than 10% of missing data, minor allele frequency (MAF) less than 1% and with duplicate physical position. Subjects with more than 10% of missing data were also removed. Therefore the IS subjects genotyped with BovineHD chip were de-imputed to the lower density with BEAGLE software v4.0 (Browning and Browning, 2007, 2016) to uniform the SNPchip density among the breeds. Briefly, the common markers between Bovine150K chip and BovineHD chip were first extracted, and then the non-common ones were imputed. For IH only missing genotypes imputation and phasing were performed. Allelic r^2 estimated by BEAGLE was used to evaluate the imputation accuracy. Markers with r^2 lower than 0.75 were excluded. After imputation the genotypes were quality checked. SNP with MAF lower than 5% or Hardy Weinberg Equilibrium (HWE) P-value lower than 10^{-6} were excluded. Subjects with lower or higher observed heterozygosity (ObsHet) (average ± 4SD; 0.3352 ≥ ObsHet ≥ 0.4494) or not pure (i.e. more than 20% of other breed genetic component evaluated, using K2 from Admixture software v1.3 (Alexander et al., 2009) were excluded.

The final datasets consisted in 118,135 SNPs in 416 IS animals and 121,165 SNPs in 436 IH animals.

II - 3.5. Genome-wide association analysis

Genome-wide association analysis was carried out based on regression of phenotypes on the genotypes of animals for one SNP at a time. For this purpose a Genome-wide Rapid Analysis

using Mixed Model and Score test has been carried out (GRAMMAS) (Aulchenko et al., 2007a) in R package GenABEL (Aulchenko et al., 2007b) as described by the following general formula:

$$Y = \mu + Xb + Sa + Zu + \varepsilon$$

where *Y* is the vector of trait values (milk fatty acids – FA), μ is the overall mean, *b* is the vector of fixed effects (milk yield, fat and protein content, log(SCC), parity, DIM and farm effect), *a* is the fixed effect of the SNP genotype, *u* and ε are vectors of random additive polygenic effects and random residuals respectively, $u \sim N(0, A\sigma^2_a)$ and $\varepsilon \sim N(0, I\sigma^2_\varepsilon)$, where *A* is the additive genetic relationship matrix estimated from SNPs data using "ibs" function in GenABEL (Aulchenko et al., 2007b), while *I* is an identity matrix, σ^2_a and σ^2_ε are the additive genetic and residual error variance, respectively. *X*, *S* and *Z* are the related incidence matrices. The associations were deemed as being significant considering a Bonferroni-corrected genome-wide significant thresholds at 0.05 (equivalent to 4.23245×10^{-7} in IS population and 4.12660×10^{-7} in IH population). Furthermore, a GWAS analysis was performed for milk fat percentage content (FP) trait using the same model, but considering only parity, DIM and farm as fixed effects.

II - 3.6. Gene-based association analysis

As already discussed before, one of GWAS limitations is the stringent significance threshold often applied to correct for multiple testing. For this reason, a large proportion of genes with small effects are disregarded, with consequent overestimation of the effect of major genes (Capomaccio et al., 2015a). To overcome this limitation, the post-GWAS MUGBAS procedure was used to pinpoint candidate genes starting from single-SNP GWAS results (Capomaccio et al., 2015b). Briefly, MUGBAS takes into account the SNP significance results and a specific gene annotation information (*Bos taurus* reference genome assembly UMDv3.1), once defined gene boundaries MUGBAS condenses the p-values of a SNP subset weighting local LD and estimates a gene-based association p-value. In this analysis the gene boundaries were artificially increased in both sides (100 Kbp) in order to capture regulatory signals. For each gene, a 'gene-

wise' p-value and False Discovery Rate (FDR) q-value were calculated. The associations with the FDR q-values less than or equal to 0.05 were considered as being significant.

II - 3.7. Pathways interaction analysis

To infer the most probable candidate among the list of significant genes uncovered by MUGBAS, a Pathways Interaction Analysis (PIA) was performed (see chapter 4). Briefly, PIA is an in-house R-package based on gene relation investigations, taking into account upstream and downstream pathway interactions. For this purpose, the significant gene list is ordered into respective KEGG metabolic pathways (Kanehisa and Goto, 2000). Once pathways strongly related with the trait of interest are chosen (first degree interaction - FDI), an interaction network is automatically created selecting the relative up/downstream pathways (from 2 - second degree interaction, SDI - to *n* degree of interactions) based on information available on KEGG databases (Kanehisa and Goto, 2000). Ultimately, genes falling inside the pathways in interaction are considered good candidates for the trait of interest having both positional and functional evidences at their sight. In the present study, after KEGG 'Lipid Metabolism' pathways were selected as FDI, PIA analysis was built until the third degrees of interactions considering the complexity of gene networks driving bovine milk fat synthesis (Bionaz and Loor, 2008a).

II - 4. Results and discussion

A GWA approach was used to identify QTL affecting milk fatty acids composition in Italian IS and IH cows and to assess the genetic differences and similarities between those two breeds. In this regard, this is the first GWA study on FA composition in IS, since other similar studies were focused only on milk production and related quality traits (Capomaccio et al., 2015a) or lactation curve (Macciotta et al., 2015).

II - 4.1. Phenotype and genotype statistics

The two populations displayed normal, or near-normal, distribution for all fatty acids. A summary of FA profile measures and significant differences between the two populations is

reported in the Table II-1. The predominant FA were C16:0, C18:1c9, C14:0 and C18:0 in both breeds.

Although it is not possible to state which milk is healthier or better without knowing the final destination of consumption, the comparison between the two breeds supports the consideration that FA profile differs significantly. Interesting differences ($P < 0.05$) among breeds were observed in almost all FA considered (73 out of 83) with a particularly high statistical significance level ($P<0.0001$) for 65 traits (Table II-1). In this regard, no information is available in literature comparing milk fatty acids profiles between IS and IH breeds. In a more general perspective, milk fat comparison has a wide range of availability in literature for Holstein, Brown Swiss and Jersey cows, under different feeding regimen (DePeters et al., 1995; Kelsey et al., 2003; Moore et al., 2005; Carroll et al., 2006; Palladino et al., 2010; Nantapo et al., 2014). Whereas, to the best of our knowledge, less information is available about milk fat in Simmental cows. In this regard, Pilarczyk and colleagues (Pilarczyk et al., 2015), comparing the milk fat content of Simmental and Holstein-Friesian cows under the same feeding regimen (organic farming), reported that the concentration of PUFA n-6, was higher in Simmental cows, whereas there is no difference among the content of SFA and MUFA. A recent comparison of milk FA composition in Polish Simmental and Holstein is also available (Adamska et al., 2016). The authors reported higher contents of short chain SFA in Polish Simmental milk fat in comparison with Polish Holstein. In this study SFA milk concentration was higher (+2.5%) in the IS than IH, whereas MUFA concentration was lower (-6%). In particular, among SFA, short and medium chain FA were significantly higher in IS milk, whereas C18:0 content was higher in IH milk fat. Furthermore, the amounts of long chain FA were higher in IH. Considering the PUFA content, the n-6 FA were higher (+10%) in IS milk fat, whereas PUFA n-3 were higher (+20%) in IH.

Differences between the two breeds were also confirmed in terms of heritability (h^2) of the traits of interest, particularly regarding the estimated values for short-SFA (i.e. C6:0, C8:0, C10:0, C14:0). A moderate heritability was also found for C16:1 c9, C18:0, for DI 10-1/(10+10-1) and DI 14-1/(14+14-1), and for DI Rum/(vac+rum) only in IS. A summary of all heritability values is reported in Table II-2. Generally, estimated heritability results confirmed the good partitioning of observed variation into unobserved genetic and environmental factors. In particular, the estimated values indicated that genetic components contribute particularly to

SFA metabolic traits, as expected (Mele et al., 2007; Schennink et al., 2009b; Bilal et al., 2014). Considering single FA heritability values, our results were higher compared to those reported by other authors (Mele et al., 2009). In this regard, it must be taken into account that heritability estimation is always specific for the investigated population and, generally, is influenced by samples size and sampling purpose (in this case designed for GWAS analysis).

II - 4.2. Single-population GWAS results

After the single-marker genome-wide association study on 83 FA traits, 5 and 90 significant SNP were identified (i.e. over Bonferroni threshold) in IS and IH, respectively (Table II-3).

More specifically, few significant associations in IS were detected on BTA19 and BTA26 for C14:0; C14:1c9; DI 10-1/(10+10-1) and DI 14-1/(14+14-1) traits (Figures II-1, -2, -3 and -4), whereas, several signals were detected on IH BTA26 for C14:1c9, DI 10-1/(10+10-1) and DI 14-1/(14+14-1) traits (Figures II-5, -6 and -7). Overall, the 98% of significant SNPs were located in noncoding regions. In particular, except for intron variants (~58%), most of significant SNPs that predicted consequences were located at 5' of a gene (~29%). This result is not surprising because it is known by literature that most signals from GWAS map to the non-coding genome (Edwards et al., 2013; Zhu et al., 2017). Although functional interpretation of these associations remained challenging, it is possible to speculate as this indicates a complex regulatory mechanism for fatty acid metabolism.

The low number of SNP associated with IS FA profile trait is not surprising (Capomaccio et al., 2015a). Generally, complex traits such as fatty acids profile (Bionaz and Loor, 2008a; Buitenhuis et al., 2014) are affected by a few major genes with large effects and many others with moderate to low effects. The latter are not easily identified by genome-wide scans in modern cattle breeds due mainly to sample size limitation. In addition, the signals from the major genes are lost due to the fixation of favourable alleles (Capomaccio et al., 2015a). Differences between the two breeds, in terms of GWA results, are consistent with the consideration that IS has a different genetic background compared to IH (Bomba et al., 2015; Marras et al., 2015). The significant signals on BTA19 (~51.3 Mb) and BTA26 (~22.0 Mb) are likely related to the effect of *FASN* (chr 19, AC_000176.1, 51384892-51403614) and *SCD* (chr 26, AC_000183.1, 21137945-21148317) genes, involved in the fatty acid synthesis and unsaturated fatty acid biosynthetic process respectively. At the same time, it is interesting to

note that no signals were detected on BTA14, where *DGAT1* is located. The importance of the *DGAT1* gene in lactation has been widely described (Grisart et al., 2002b). This is consistent with the fact that considering the milk fat percentage content (FP) significant associations were found on IH BTA14 (~15.3 Mb, ~18.2 Mb, ~29.5 Mb, ~41.4 Mb) (Figure II-8, Table II-4). This association is clearly due to *DGAT1* effect (chr 14, AC_000171.1, 1795425-1804838).

In this regard, the absence of *DGAT1* signal in IS breed is in line with previous studies where DGAT1 p.232K allele was identified with a very low frequency in IS (Scotti et al., 2010), a condition that does not allow the association with milk fat percentage trait in IS (Capomaccio et al., 2015a).

II - 4.3. Gene-based association analysis results

As already explained before, MUGBAS gene-centric approach was performed (Capomaccio et al., 2015b) with the aim to overcome the stringent significance GWA threshold and amplify the single-marker association signals. A summary of all significant genes and the relative best SNP is reported in Supplementary Table II-S1. MUGBAS identified a list of 47 and 165 significant positional candidate genes, respectively associated to milk FA in IS and IH, showing a different pattern of genes in association with traits of interest (Supplemantary Table II-S2).

In this regard, *SCD* gene, responsible of all the conversions of saturated to mono-unsaturated fatty acids from C10:0 to C18:0, was found to be significantly associated with C14:1c9, DI 10-1/(10+10-1) and DI 14-1/(14+14-1) traits in both breeds, and with C10:1c9, C16:1c9, DI 16-1/(16+16-1) specifically in IS. These results are in line with other authors' outcomes (Mele et al., 2007; Moioli et al., 2007; Schennink et al., 2008; Conte et al., 2010; Bouwman et al., 2011; Buitenhuis et al., 2014). Whereas, *DGAT1* gene, pivotal in mammary gland triacylglycerol synthesis and known to underlie a large genetic variation in milk-fat production and composition of dairy cows (Grisart et al., 2002b), was associated only in IH with the C15:0, C16:0 and C16:1c9 traits, as expected (Scotti et al., 2010; Capomaccio et al., 2015a) and explained before. Our result in IH breed were consistent with outcomes obtained by Buitenhuis and colleagues (Buitenhuis et al., 2014) study on Danish Holstein and Danish Jersey breeds, reporting as *DGAT1* was associated to C16:0 and C16:1 but not with the desaturation indexes. Regarding this, it is woth noting that significant association between *DGAT1* and desaturation

indexes were reported in other studies on Italian Brown and Dutch Holstein breeds (Schennink et al., 2008; Conte et al., 2010). Moreover, it is worth nothing that the ARS-BFGL-NGS-4939, our best candidate SNP associated with the *DGAT1* effect in C15:0 (FA19) trait (Supplementary Table II-S2), was already shown to be in complete linkage disequilibrium (LD) with the DGAT1 p.232K polymorphism in German Holstein-Friesian (Wang et al., 2012b) and associated with milk fat percentage in IH (Capomaccio et al., 2015a). This SNP had a low frequency (0.063) in our IS population, confirming the previous findings (Scotti et al., 2010; Capomaccio et al., 2015a).

Another interesting and unexpected result was the association of *FASN*, encoding the enzyme responsible of *de novo* FA synthesis, only in IS with C10:0, C12:0 and C14:0 traits. *FASN* was reported in other studies as a candidate gene for milk fat percentage and fat composition (Schennink et al., 2009a; Bouwman et al., 2012). In particular, Schennink and colleagues (Schennink et al., 2009a) found a *FASN* association with C14:0 in Holstein Friesian. This may suggest that *FASN* variants might be of minor importance (i.e. small allelic effects) in the overall milk fat composition in IH, compared with other genes, or simply that the *FASN* effect might be masked by the major *SCD* and *DGAT1* mutation effects in our analysis. Regarding this, the BovineHD1900014372, our best SNP associated with the *FASN* effect in C10:0 and C12:0 traits (Supplementary Table II-S2), was already found to be significantly associated with FA in other breeds (Bouwman et al., 2012). This SNP had similar frequencies in both breeds (0.397 in IS and 0.375 in IH). Further analyses are required to investigate the *FASN* variants effects on bovine milk fat composition in IS and IH population but the fact remains that *FASN* is a candidate gene for milk production traits (Schennink et al., 2009a; Bouwman et al., 2012).

Other interesting candidate genes highlighted by MUGBAS but not by PIA were *coiled-coil domain containing 57* (*CCDC57*), *sorbin and SH3 domain containing 1* (*SORBS1*), and *conserved helixloop-helix ubiquitous kinase* (*CHUK*). The *CCDC57* on IS BTA19 was associated with the C12:0 and C14:0 traits. The *CCDC57* gene is expressed in cow mammary gland (Medrano et al., 2010), and the coiled-coil domains refer to protein structural motifs. This result is in agreement with the significant association outcomes on C14:0 obtained by (Bouwman et al., 2014). The *SORBS1* gene encodes an important protein in the insulin-signaling pathway in the adipose depots of humans and has a positive regulatory effect on lipid biosynthesis (Baumann et al., 2000; Yang et al., 2003; Li et al., 2014a). In Chinese Holstein

cows, Li and colleagues. (Li et al., 2014) found 2 SNPs associated with C14:1 cis-9 and DI 14–1/(14+14–1) located close to the *SORBS1* gene. In our study SORBS1 gene was associated with the C14:1 cis-9, DI 10–1/(10+10–1), and DI 14–1/(14+14–1) traits in IH on BTA 26. On BTA 26 the *CHUK* gene was associated with the DI 16–1/(16+16–1) trait in the IS population and with C14:1 cis-9 and DI 14–1/(14+14–1) in the IH population. Also, Li and colleagues (Li et al., 2014) found an association of this gene with the DI 14–1/(14+14–1) trait, but they considered their result to be an artifact due to the close proximity of *CHUK* to the *SCD* gene. The descriptions of the other significant genes discovered by MUGBAS but not revealed in PIA are reported in the 'Other significant genes' section.

II - 4.4. Pathway analysis results

To reduce false-positive signals and concentrate on finding meaningful results, a gene pathway analysis using an in-house method named Pathway Interaction Analysis (PIA) was performed (see chapter 4). This approach helped to confirm the functional role of significant genes obtained by MUGBAS. A summary of all genes obtained with PIA for the three degrees of interaction investigated in our analysis is reported in Table II-5. Among the MUGBAS significant genes and as a corollary of well-known principal lipogenic genes (discussed before), PIA highlighted another 8 and 10 possible candidate genes in IS and IH respectively, with functional roles in pathways related with 'Lipid Metabolism' in KEGG and/or in interaction with them (Figures II-9 and -10).

The originality of this approach lies in the possibility to investigate many levels of interactions among the pathways related to the trait of interest, revealing connections difficult to identify at a glance. Overall, PIA results confirmed the presence of differences between the two breeds in terms of genes associated with milk FA profile (Supplementary Table II-S2).

In this regard, at the first level of interaction, along with the expected effect of *SCD, FASN* and *DGAT1* genes, the significant signals of *ECI2*, *PCYT2*, *DCXR* genes in IS and of *CYP17A1* gene in IH were detected. In more detail, on IS BTA23 *ECI2* gene was associated with C7:0 trait. This gene (*enoyl-CoA delta isomerase 2*) encodes for an auxiliary mitochondrial enzyme involved in unsaturated fatty acid oxidation (van Weeghel et al., 2012), i.e. in the degradation of FA during fatty acid β-oxidation, as showed by PIA. In particular, ECI2 is an enzyme that

converts a cis-double bond to a trans-double bond so that β-oxidation may continue. It is interesting to note the significant association of this gene only in IS, where a higher milk n-6 FAs content was detected compared to IH. On IS BTA19, *DCXR* and *PCYT2* genes, involved in 'Arachidonic acid metabolism' and 'Glycerophospholipid metabolism' pathways, were associated with C14:0 trait. In particular, *DCXR* is a well-conserved gene among species and encodes for dicarbonyl L-xylulose reductase, a member of the short chain dehydrogenase/reductase superfamily, which reduces the various α-dicarbonyl compound (DCs) involved in the formation of advanced glycation end-products (AGEs) (Nakagawa et al., 2002). DCs are generated from sugars or lipid compounds in various biological systems by oxidative stress (Lee et al., 2013). Whereas, *PCYT2* encodes the ethanolamine-phosphate cytidylyltransferase 2, involved in the synthesis of phosphatidylethanolamine from diacylglycerol. In human adipose tissue, *PCYT2* expression is considered negatively correlated with fat mass percentage and body mass index (Sharma et al., 2013). In this regard, it is interesting to note that both genes are on BTA19, the same chromosome of *FASN* from which they are distant only ~63.1 Kb and ~122.8 Kb, respectively. This may support the hypothesis of possible false positive signals (Table 7). Nevertheless MUGBAS results showed that these two genes had two different best SNPs from *FASN* (i.e. ARS-BFGL-NGS-39983 and ARS-BFGL-NGS-90673). The LD among the 3 SNPs is very low (i.e. the maximum r^2 value is 0.09). This suggesting the possibility of a long haplotype associated with C14:0. On IH BTA26, *CYP17A1* gene was found to be associated with DI 14-1/(14+14-1) trait. This gene is involved in 'Steroid hormone biosynthesis' and, expressed in bovine ovary theca, encodes a steroid lyase that catalyzes the final step of androgen biosynthesis (Vanselow and Fürbass, 2011). This association is intriguing considering the known effect of hormones to increase milk yield in lactating animals (Mohammed and Johnson, 1985; Kalashnikova et al., 2009).

Focusing on the second degree of interaction (Table II-5), the significant and best candidate association of the *G6PC3* gene with C18:1 t6-8 trait was detected on IS BTA19 (Supplementary Table II-S2). PIA showed the *G6PC3* gene to be involved in 'Glycolysis/Gluconeogenesis pathway' and in 'Starch and sucrose metabolism' at the second and third level of interaction, respectively. This gene belongs to one of the gene families encoding the glucose-6-phosphatase enzyme, that catalyses the final step of gluconeogenesis with the hydrolysis of glucose-6-phosphate, ubiquitously expressed in humans (Banka and Newman, 2013). With regards to the IH breed, at the second level of interaction, the significant association of *ACO2* (*Aconitase 2*)

gene was detected on BTA5 with C15:0 and Odd-Chain Fatty Acids (OCFA) traits. We feel our results shows that *ACO2* should be considered as a best candidate gene which encodes a nuclear protein acting in the mitochondrion and catalysing the interconversion of citrate to isocitrate via cis-aconitate in the second step of the tricarboxylic acid cycle. On IH BTA26, a significant association was also found between *PI4K2A* (*Phosphatidylinositol 4-Kinase Type 2 Alpha*) gene with C14:1c9, ID 10-1/(10+10-1) and ID 14-1/(14+14-1) traits. *PI4K2A* was investigated in humans and an autosomal recessive mutation was found causing hereditary spastic paraplegia, the initial studies were completed in a knockout mouse model showing this phenotype (Cleeter et al., 2011). Its product is one of the enzymes involved in vesicle formation in the trans-Golgi network (TGN) and endosomes in mammalian cells (Albanesi et al., 2015). At the third level of interaction, PIA showed *PYCR1* and *ALG12* genes in IS, whereas *GOT1*, *GPT*, *NT5C2*, *PDE6G*, *POLR3H* and *COX15* genes in IH. *PYCR1* was found to be associated with the C14:0 trait on IS BTA19. PIA found it to be involved in the ‘Arginine and proline metabolism pathways’. The *PYCR1* (*pyrroline-5-carboxylate reductase 1*) gene synthetizes an enzyme involved in the proline metabolism and synthesis which are associated with the tricarboxylic acid cycle, urea cycle and the pentose phosphate pathway and, in several human studies, has been associated with tumor proliferation (Guernsey et al., 2009; Possemato et al., 2011; Cai et al., 2018). *ALG12* (a*sparagine-linked glycosylation 12 homolog*), was showed by PIA to be involved in the ‘N-Glycan biosynthesis pathway’ and considered by MUGBAS to be a best positional candidate gene associated with the C18:1 t6-8 trait on IS BTA5. It has been demonstrated that the *ALG12* promoter shows less transcriptional activity in response to endoplasmic reticulum stress, but its basic regulatory mechanism has not been characterized (Oh-Hashi et al., 2013). With regards to IH, PIA showed a cluster of genes involved in ‘Purine metabolism, ‘Pyrimidine metabolism’, ‘Alanine, aspartate and glutamate metabolism’: *NT5C2*, *GPT,* and *COX15.* In more detail, *NT5C2* was found to be associated with the DI 14-1/(14+14-1) trait and was considered by MUGBAS as a positional best candidate although on BTA26. *NT5C2* (*5'-Nucleotidase, Cytosolic II*) gene encodes protein involved in cellular purine metabolism. NT5C is in a family of enzymes that inhibit basal lipid oxidation and glucose transport in skeletal muscle. Reduction of *NT5C* expression or activity may promote metabolic flexibility in type 2 diabetes (Kulkarni et al., 2011). On IS BTA14 *GPT* (*glutamic-pyruvic transaminase*), was found to be associated with the C16:0, C16:1c9 and BCFA+OCFA (Branched Chain Fatty Acids + Odd-Chain Fatty Acids) traits. Serum GPT level is one of the

most clinically important indicators of liver function in many studies testing different diets or supplements (Liao et al., 2013). Whereas, *COX15* (*Cytochrome c oxidase*) was found to be associated with the C14:1c9, ID 10-1/(10+10-1) and ID 14-1/(14+14-1) traits on IS BTA26. *COX15* (*cytochrome c oxidase homologue*) is the terminal component of the mitochondrial respiratory chain, catalyzes the electron transfer from reduced cytochrome c to oxygen and functional and genetic studies demonstrated that its deficiency causes cardiomyopathy (Antonicka et al., 2003; Fedida et al., 2017). With regards to the IH breed, a significant association of *POLR3H* with C15:0 and OCFA were found on IH BTA5, although it close in proximity to *ACO2*. *POLR3H* (*Polymerase (RNA) III subunit H*) is a gene involved with the immune response and its expression level has been studied on pigs, where the authors supposed its function on immunity may contribute to feed efficiency (Gondret et al., 2017) and likely energetic metabolism. The description of the other significant genes discovered by MUGBAS but not revealed by PIA is reported in the 'Other significant genes' section.

II - 4.5. Other significant genes

Numerous other genes were highlighted by MUGBAS, some of them appeared in both breeds and many others exclusively associated with a single breed's fatty acid (FA) profile (Supplementary Table II-S2). Although these novel genes are not exclusively related to FA metabolism, they are still worthy of note. In fact, while often described in human diseases, these genes could potentially play a role in bovine fat physiology considering their possible pleiotropic effects or limited characterization. Clearly, further analyses are required to confirm the associations and roles of these genes in the bovine milk FA profile.

On chromosome BTA26, *GSTO* (*glutathione S-transferase omega*) was associated with the DI 10-1/(10+10-1) and DI 14-1/(14+14-1) traits in the IH breed. *GSTO* is a lipid metabolism-related gene whose expression in the liver was correlated with the FA content of the diet in pigs (Świątkiewicz et al., 2016). On IS BTA7, *SLC12A2* (*solute carrier family 12 member 2*) was associated with the C18:1c11 trait, which in chickens was found to be associated with abdominal fat (Zhang et al., 2012). On IS BTA14, *UQCRB* (*ubiquinol-cytochrome C reductase binding protein*) was associated with the C16:0 trait. *UQCRB* is a nuclear gene encoding the human ubiquinone-binding protein of Complex III (CIII), and its deficiency is associated with mitochondrial disease and one of the least common oxidative phosphorylation defects

(Fernández-Vizarra and Zeviani, 2015). *CCDC57* (*coiled-coil domain containing 57*) on IS BTA19 was associated with the C12:0 and C14:0 traits. The *CCDC57* gene is expressed in cow mammary gland (Medrano et al., 2010), and the coiled-coil domains refer to protein structural motifs. Our result is in agreement with the significant association outcomes on C14:0 obtained by Bouwman et al. (2014). The *PKD2L1* (*polycystin 2 like 1, transient receptor potential cation channel*) gene on BTA26, considered as the best candidate, was associated with C10:1c9, C14:1c9, DI 10-1/(10+10-1), C16:1 c9, DI 14-1/(14+14-1), and DI 16-1/(16+16-1) in IS cows. The *PKD2L1* gene was strongly associated with lysophosphatidylcholine (LPC) 16:1 phospholipids (Demirkan et al., 2012). Another interesting gene on IS BTA26, *CWF19L1* (*CWF19 like 1, cell cycle control*), was found to be associated with the C16:1 c9, DI 14-1/(14+14-1), and DI 16-1/(16+16-1) traits. In humans, this gene has been associated with nonalcoholic fatty liver disease (Kitamoto et al., 2014). The *CHUK* (*conserved helix-loop-helix ubiquitous kinase*) gene was associated with the DI 16-1/(16+16-1) trait in the IS population and with C14:1c9 and DI 14-1/(14+14-1) in the IH population. Also, Li *et al*. (2014) found an association of this gene with the DI 14-1/(14+14-1) trait, but they considered their result to be an artifact due to the close proximity of *CHUK* to the *SCD* gene. On BTA28, the *GHITM* (*growth hormone inducible transmembrane protein*) gene was associated with PUFA/SFA traits in IS. This gene encodes a polytopic membrane protein with 6 transmembrane domains; it is expressed in several cancer cell lines, but its physiological and probable pathological functions remain unknown (Reimers et al., 2007).

In the IH breed on BTA3, *ENSA* (*endosulfine alpha*) was associated with C14-iso. This gene was associated with a pericardial fat trait in humans (Chu et al., 2017). On BTA14, the *TONSL* (*tonsoku like DNA repair protein*) gene was associated with the C15:0, C16:0, C16:1 c9, and BCFA+OCFA traits in IH cows. *TONSL* was also considered significant by other authors, despite its proximity to *DAGT1* (Ibeagha-Awemu et al., 2016).

On BTA14, variants of *ARHGAP39* (*rho GTPase activating protein 39*), considered the best candidate, *MFSD3* (*major facilitator superfamily domain containing 3*), and *HSF1* (heat shock transcription factor 1), considered a possible false positive, were associated with the C16:0 and C16:1c9 traits, and *MFSD3* was further associated with BCFA+OCFA. *ARHGAP39* was found to be associated to C16:1 in Danish Holstein (Buitenhuis et al., 2014). *HSF1* is one four HSFs (HSF1 to HSF4) reported in vertebrates. HSF1, which participates in the heat shock response,

protects cells from various attacks, including exposure to elevated temperatures, heavy metals, proteasome inhibition, and oxidative stress (Budzyński et al., 2015).

On BTA26, several genes were found to be associated with the C14:1c9, DI 10-1/(10+10-1), and DI 14-1/(14+14-1) traits in IH (Table 1). For example, the *SORBS1* (*sorbin and SH3 domain containing 1*) gene encodes an important protein in the insulin-signaling pathway in the adipose depots in humans and has a positive regulatory effect on lipid biosynthesis (Baumann et al., 2000; Yang et al., 2003; Li et al., 2014a). In Chinese Holstein cows, Li *et al*. (2014) found two SNPs associated with C14:1c9 and DI 14-1/(14+14-1) located close to the *SORBS1* gene. The *CALHM1* and *CALHM2* genes in our study were associated with the DI 14-1/(14+14-1) trait, whereas *CALHM3* was associated with the C14:1c9 and DI 10-1/(10+10-1) traits in IH. The *CALHM* (*calcium homeostasis modulator 1, 2 and 3*) genes encode components of a brain calcium channel involved in cytosolic calcium homeostasis (Calero et al., 2012). The CALHM1 protein also plays a role in processing amyloid-beta precursor proteins (Dreses-Werringloer et al., 2008). Coding-region variants of all three human *CALMH* genes have been associated with Creutzfeldt-Jakob disease (Calero et al., 2012). On IH BTA26, the *ERLIN1*, *USMG5*, *BORCS7*, and *AS3MT* genes were associated with the DI 14-1/(14+14-1) trait. *AS3MT* (*arsenite methyltransferase*) encodes a protein that catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to trivalent arsenic (Lin et al., 2002). *USMG5* (*up-regulated during skeletal muscle growth 5 homolog*) encodes a protein that is a small subunit of the mitochondrial ATP synthase and the lysosomal V-ATPase (Kontro et al., 2012). Also, Duarte *et al*. (Duarte et al., 2016) found the *BORCS7*, *NT5C2* (already described above), and *AS3MT* genes to be associated with human schizophrenia. *ERLIN1* (*endoplasmic reticulum lipid raft associated 1*) was associated with FA metabolism in the IH breed but not in IS. In a meta-analysis conducted on nonalcoholic fatty liver disease, *ERLIN1* was shown to cluster with the *CHUK* and *CWF19L1* genes (mentioned above) (Feitosa et al., 2013) and was associated with early stages of fatty liver accumulation in hepatic inflammation.

II - 5. Conclusion

GWAS using a gene-centric approach and interaction pathways analysis allowed us to delineate a genomic region and novel genes associated with FA profile in IS and IH breeds. In particular, significant associations were detected on five chromosomes (BTA05, BTA14, BTA19, BTA23

and BTA26) for thirteen fatty acids (C7:0, C10:0, C10:1c9, C12:0, C14:0, C14:1c9, C15:0, C16:0, C16:1c9, C18:1 t6-8, C18:2 9-11 c/t, OCFA, BCFA + OCFA) and three desaturation indexes (ID 10-1/(10+10-1), ID 14-1/(14+14-1) and ID 16-1/(16+16-1)). According to previous results reported in literature, the effects of well-established genes associated with milk fat yield and content such as *SCD*, *DGAT1* and *FASN* were confirmed, with some differences among the breeds. Furthermore, this study revealed other possible candidate genes, several of them directly or indirectly involved in 'Lipid Metabolism' which include *ECI2*, *PCYT2*, *DCXR*, *G6PC3*, *PYCR1*, *ALG12, CYP17A1*, *ACO2*, *PI4K2A*, *GOT1*, *GPT*, *NT5C2*, *PDE6G*, *POLR3H* and *COX15*. In summary, the findings improve our understanding of genetic architecture in IS and IH cows and highlight breed-specific genomic features, in terms of milk fatty acids profile.

II - 6. Figures and tables

Figure II-1. Manhattan plots of GWAS results showing the significance of SNP associations for C14:0 fatty acid (FA) trait in the Italian Simmental (IS) breed. Negative $\log_{10}$ p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

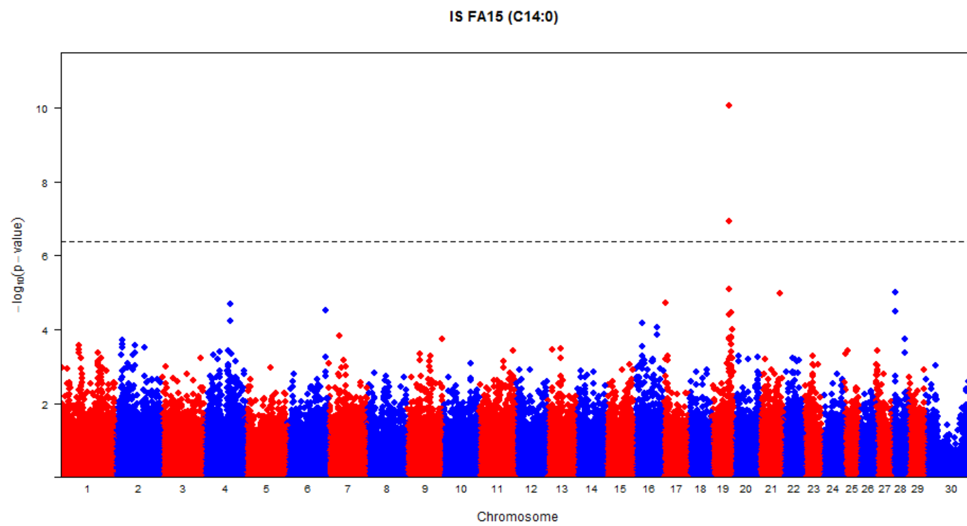


Figure II-2. Manhattan plots of GWAS results showing the significance of SNP associations for C14:1c9 fatty acid (FA) trait in the Italian Simmental (IS) breed. Negative $\log_{10}$ p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

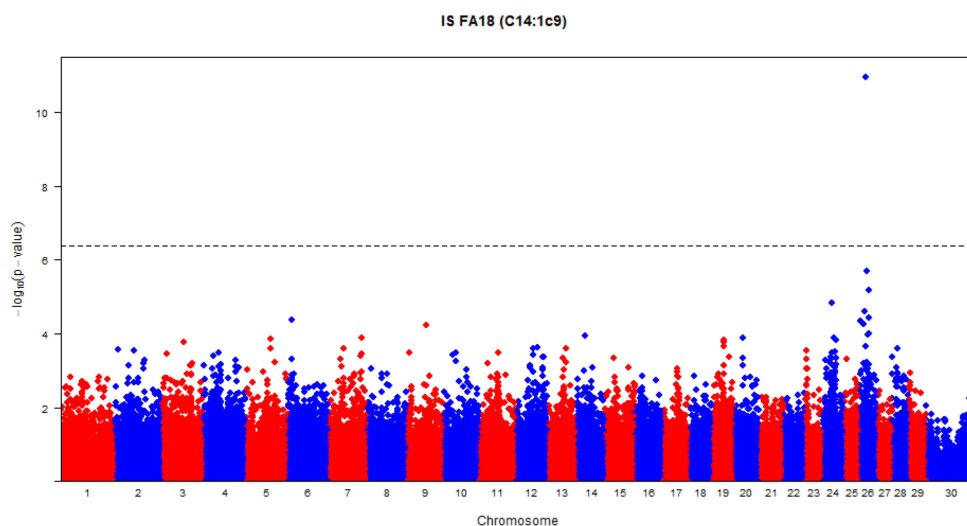


Figure II-3. Manhattan plots of GWAS results showing the significance of SNP associations for ID 10-1/(10+10-1) fatty acid (FA) trait in the Italian Simmental (IS) breed. Negative $\log_{10}$ p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

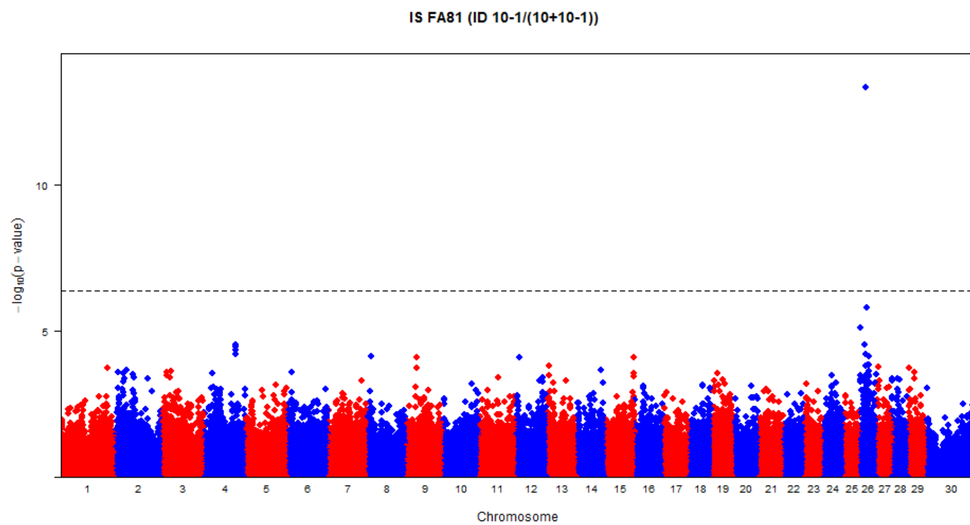


Figure II-4. Manhattan plots of GWAS results showing the significance of SNP associations for ID 14-1/(14+14-1) fatty acid (FA) trait in the Italian Simmental (IS) breed. Negative $\log_{10}$ p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

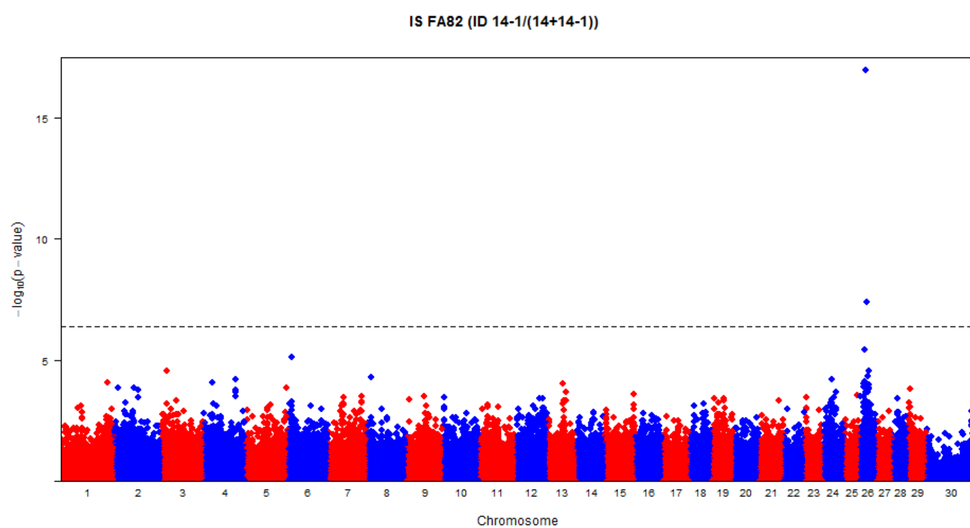


Figure II-5. Manhattan plots of GWAS results showing the significance of SNP associations for C14:1c9 fatty acid (FA) trait in the Italian Holstein (IH) breed. Negative log10 p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

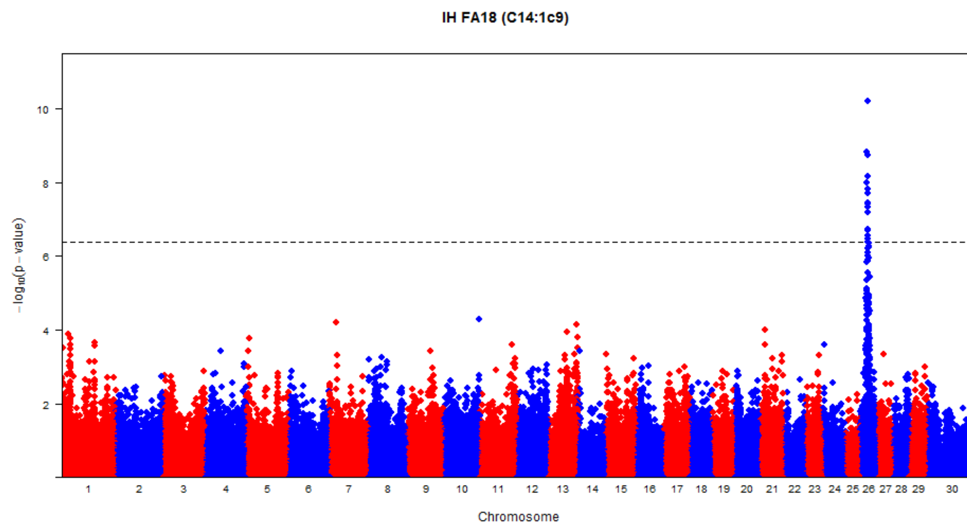


Figure II-6. Manhattan plots of GWAS results showing the significance of SNP associations for ID 10-1/(10+10-1) fatty acid (FA) trait in the Italian Holstein (IH) breed. Negative log10 p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

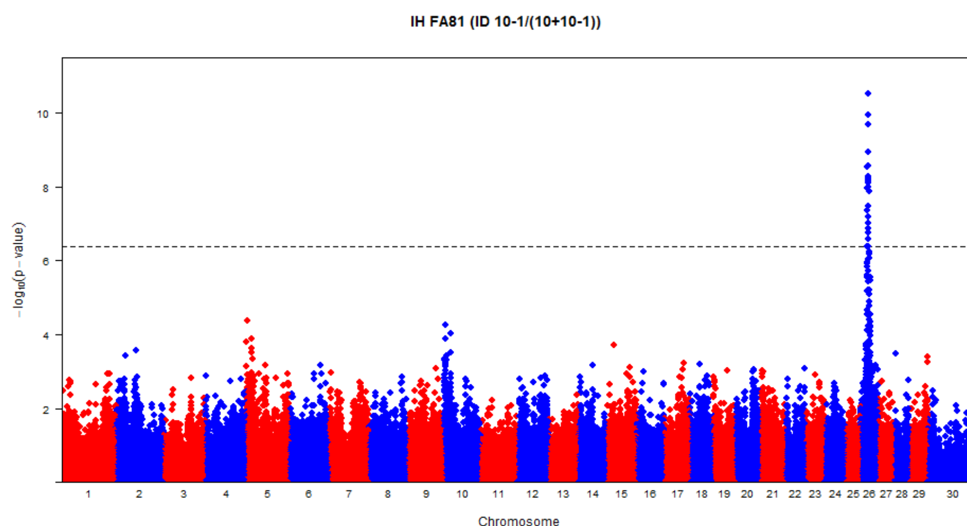


Figure II-7. Manhattan plots of GWAS results showing the significance of SNP associations for ID 14-1/(14+14-1) fatty acid (FA) trait in the Italian Holstein (IH) breed. Negative log10 p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

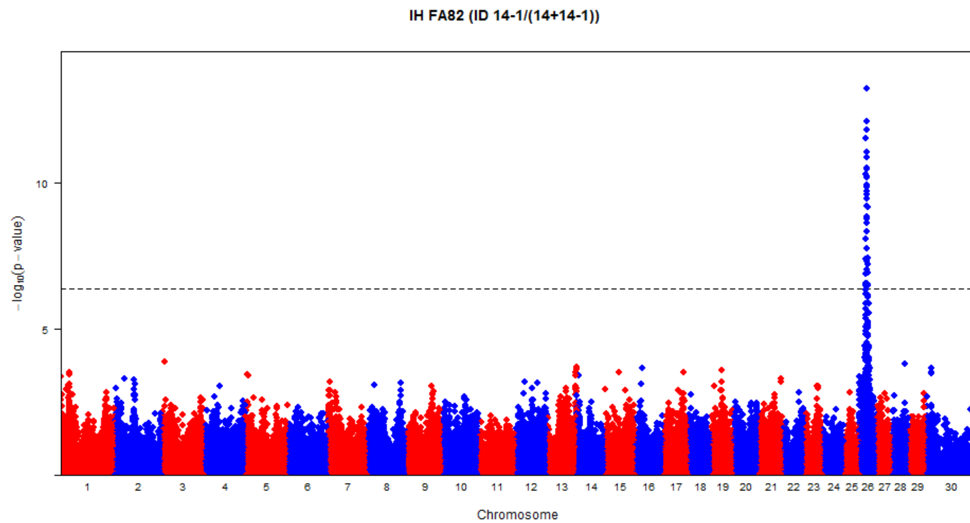


Figure II-8. Manhattan plots of GWAS results showing the significance of SNP associations for milk fat percentage content (FP) in Italian Holstein (IH) breed. Negative $\log_{10}$ p-values of all SNPs that passed quality control are plotted against their genomic positions. Different chromosomes are distinguished with blue and red colors. The dashed line indicates the Bonferroni-corrected genome-wide significance threshold at p-value 0.05

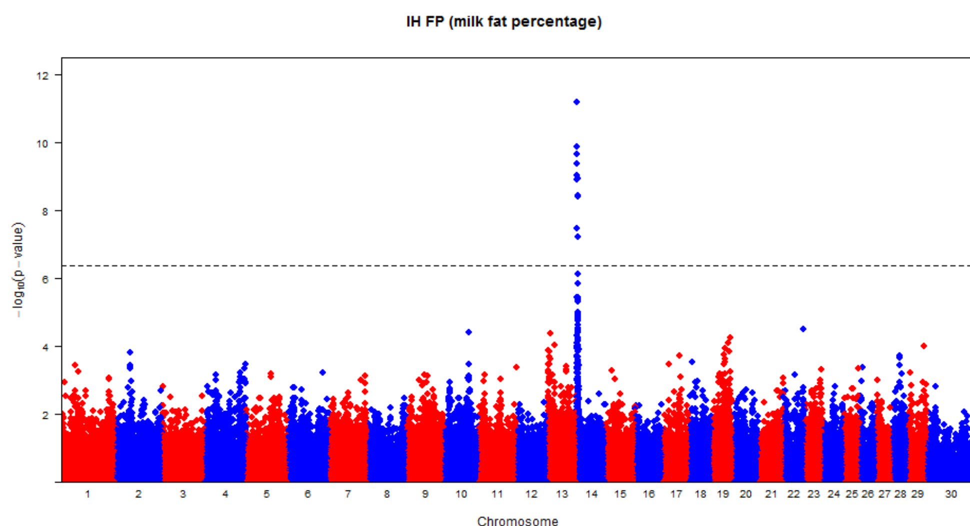


Figure II-9. Diagram showing the results obtained with Pathway Interaction Analysis (PIA) on MUGBAS significant genes for the Italian Simmental (IS) breed. The pink circles represent functional candidate genes falling inside the pathways associated with the trait of interest or interacting with the pathway. The red squares represent first-degree (FDI) interaction pathways, directly connected to the trait of interest (i.e. 'Lipid Metabolism' in KEGG). The green and blue diamond symbols represent second- and third-level pathways, respectively, interacting with FDI pathways, as highlighted by the PIA.

num	name
1	FASN
2	ECI2
3	PCYT2
4	ENSBTAG00000047043
5	LOC101906058
6	SCD1
7	G6PC3
8	DCXR
9	PYCR1
10	ALG12
11	Fatty acid biosynthesis
12	Fatty acid degradation
13	Glycerophospholipid metabolism
14	Glycolysis / Gluconeogenesis
15	Arachidonic acid metabolism
16	Biosynthesis of unsaturated fatty acids
17	Pentose and glucuronate interconversions
18	Arginine and proline metabolism
19	Starch and sucrose metabolism
20	N-Glycan biosynthesis

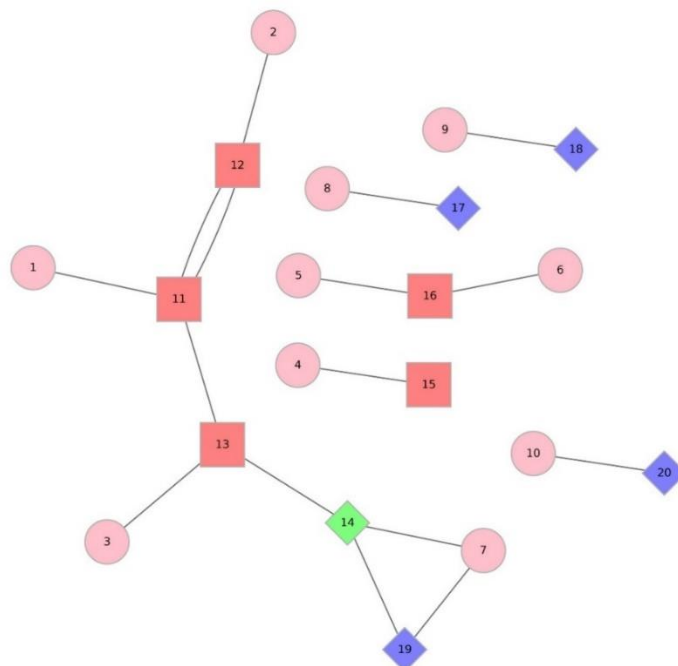
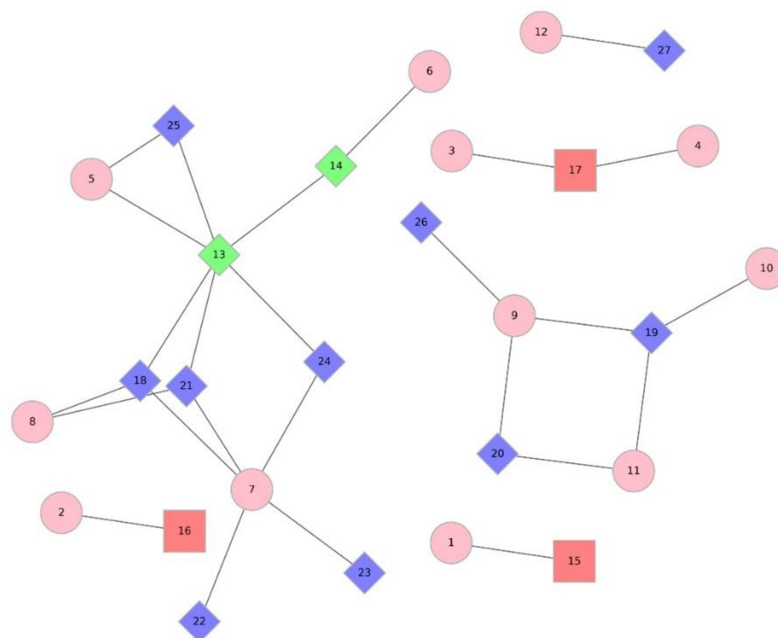


Figure II-10. Diagram showing the results obtained with Pathway Interaction Analysis (PIA) on MUGBAS significant genes for the Italian Holstein (IH) breed. The pink circles represent functional candidate genes falling inside the pathways associated with the trait of interest or interacting with the pathway. The red squares represent the first-degree (FDI) interaction pathway, directly connected to the trait of interest (i.e. 'Lipid Metabolism' in KEGG). The green and blue diamond symbols represent the second- and third-level pathways, respectively, interacting with FDI pathways, as highlighted by the PIA.



num	name
1	CYP17A1
2	DGAT1
3	SCD
4	SCD1
5	ACO2
6	ENSBTAG00000002010
7	GOT1
8	GPT
9	NT5C2
10	PDE6G
11	POLR3H
12	COX15
13	Citrate cycle (TCA cycle)
14	Inositol phosphate metabolism
15	Steroid hormone biosynthesis
16	Glycerolipid metabolism
17	Biosynthesis of unsaturated fatty acids
18	Arginine biosynthesis
19	Purine metabolism
20	Pyrimidine metabolism
21	Alanine, aspartate and glutamate metabolism
22	Cysteine and methionine metabolism
23	Arginine and proline metabolism
24	Tyrosine metabolism
25	Glyoxylate and dicarboxylate metabolism
26	Nicotinate and nicotinamide metabolism
27	Porphyrin and chlorophyll metabolism

Table II-1. Mean and standard deviation (SD) of individual fatty acids (FA), grouped fatty acids and desaturation index (DI) (g /100 g of FA) in milk of Italian Simmental (IS) and Italian Holstein (IH) breeds determined by gas-chromatography.

		IS		IH		Signif.[1]
FA	FA code	Mean	SD	Mean	SD	p-value
C4:0	FA01	2.804	0.351	2.812	0.523	ns
C5:0	FA02	0.037	0.014	0.038	0.017	ns
C6:0	FA03	1.867	0.205	1.760	0.254	***
C7:0	FA04	0.029	0.014	0.029	0.015	ns
C8:0	FA05	1.116	0.138	0.985	0.143	***
C10:0	FA06	2.586	0.397	2.122	0.370	***
C10:1c9	FA07	0.206	0.043	0.187	0.043	***
C11:0	FA08	0.063	0.032	0.056	0.030	***
C12:0	FA09	2.944	0.494	2.377	0.430	***
C13-iso	FA10	0.022	0.007	0.019	0.010	***
C13-anteiso	FA11	0.056	0.016	0.048	0.016	***
C12:1c11	FA12	0.067	0.020	0.055	0.019	***
C13:0	FA13	0.115	0.040	0.104	0.038	***
C14-iso	FA14	0.100	0.034	0.084	0.035	***
C14:0	FA15	12.132	1.323	10.722	1.060	***
C15-iso	FA16	0.224	0.044	0.190	0.029	***
C15-anteiso	FA17	0.458	0.074	0.386	0.059	***
C14:1c9	FA18	0.797	0.224	0.818	0.234	ns
C15:0	FA19	1.197	0.278	1.108	0.260	***
C16-iso	FA20	0.254	0.074	0.210	0.071	***
C16:0	FA21	32.320	3.868	31.436	3.046	***
C16-1t6-7	FA22	0.030	0.006	0.032	0.008	***
C16:1t9	FA23	0.043	0.013	0.044	0.012	ns
C17-iso	FA24	0.482	0.065	0.441	0.064	***
C16:1 c9	FA25	1.433	0.376	1.320	0.357	***
C17-anteiso	FA26	0.434	0.075	0.373	0.056	***
C17:0	FA27	0.562	0.088	0.518	0.072	***
C17:1 c9	FA28	0.230	0.070	0.201	0.068	***
C18:0	FA29	8.678	2.003	10.876	1.906	***
C18:1 t4	FA30	0.017	0.005	0.025	0.009	***
C18:1 t5	FA31	0.017	0.026	0.023	0.031	**

C18:1 t6-8	FA32	0.313	0.063	0.378	0.093	***
C18:1 t9	FA33	0.251	0.073	0.291	0.067	***
C18:1 t10	FA34	0.475	0.168	0.535	0.284	***
C18:1 t11	FA35	0.858	0.239	0.955	0.269	***
C18:1 t12	FA36	0.601	0.246	0.739	0.310	***
C18:1c9	FA37	20.145	3.000	21.469	2.597	***
C18:1 c11	FA38	0.821	0.179	0.985	0.180	***
C18:1 c12	FA39	0.465	0.177	0.555	0.183	***
C18:1t16	FA40	0.335	0.091	0.429	0.090	***
C18:2 t9.t12	FA41	0.250	0.064	0.240	0.048	**
C18:2 t11.c15	FA42	0.076	0.031	0.119	0.067	***
C18:2 c9.c12	FA43	2.531	0.555	2.278	0.458	***
C20:0	FA44	0.146	0.036	0.150	0.037	ns
C18:3 c9.c12.c15	FA45	0.435	0.167	0.553	0.168	***
C18:2 9.11 c/t	FA46	0.457	0.119	0.432	0.126	**
C18:4 c6.c9.c12.c15	FA47	0.019	0.007	0.020	0.010	*
C20:2 c11.c14	FA48	0.020	0.008	0.018	0.008	**
C18-3 c9.t11.c15	FA49	0.016	0.009	0.019	0.011	***
C20:3 c8.c11.c14	FA50	0.128	0.039	0.122	0.029	*
C20:4 c5.c8.c11.c14	FA51	0.177	0.032	0.152	0.036	***
C20:5 c5. c8. c11. c14. c17	FA52	0.038	0.010	0.045	0.015	***
C22:4 c7.c10.c13.c16	FA53	0.032	0.011	0.027	0.011	***
C22:5 c7.c10.c13.C16.c19	FA54	0.078	0.022	0.080	0.024	ns
C22:6 c4.c7.c10.c13.16.c19	FA55	0.014	0.010	0.010	0.009	***
SFA	FA56	68.510	3.809	66.740	3.507	***
UFA	FA57	31.490	3.809	33.260	3.507	***
PUFA	FA58	4.270	0.788	4.116	0.651	**
MUFA	FA59	27.220	3.395	29.144	3.144	***
PUFA n6	FA60	2.887	0.589	2.596	0.482	***
PUFA n3	FA61	0.584	0.178	0.709	0.189	***
SCFA (C<10)	FA62	8.645	0.903	7.933	1.083	***
MCFA (10<C<17)	FA63	53.963	5.107	50.542	3.980	***
LCFA (C>17)	FA64	37.393	5.440	41.526	4.302	***
BCFA	FA65	1.546	0.243	1.311	0.199	***
BCFAiso	FA66	1.081	0.174	0.944	0.153	***
BCFAanteiso	FA67	0.465	0.126	0.367	0.086	***
OCFA	FA68	3.842	0.558	3.445	0.439	***
trans 18-1 tot	FA69	2.868	0.673	3.374	0.806	***

trans totali	FA70	3.191	0.691	3.690	0.837	***
UFA/SFA	FA71	0.464	0.084	0.503	0.081	***
PUFA/SFA	FA72	0.063	0.014	0.062	0.012	ns
PUFA/(SFA-C18:0)	FA73	0.073	0.018	0.074	0.015	ns
n6/n3	FA74	5.126	0.931	3.898	1.089	***
DHA/EPA	FA75	0.376	0.305	0.231	0.173	***
AA/DHA	FA76	18.552	14.790	20.954	11.579	**
AA/(DHA+EPA+DPA)	FA77	1.409	0.322	1.183	0.347	***
BCFA + OCFA	FA78	5.388	0.719	4.756	0.529	***
BCFA/OCFA	FA79	0.405	0.052	0.384	0.059	***
BCFAiso/BCFAanteiso	FA80	2.518	1.217	2.643	1.119	ns
DI 10-1/(10+10-1)	FA81	0.074	0.014	0.082	0.016	***
DI 14-1/(14+14-1)	FA82	0.061	0.015	0.071	0.018	***
DI 16-1/(16+16-1)	FA83	0.042	0.009	0.040	0.009	***
DI 18-1/(18+18-1)	FA84	0.701	0.041	0.665	0.043	***
DI Rum/(vac+rum)	FA85	0.350	0.049	0.313	0.048	***

[1] FA traits were compared between the breeds by using the Welch Two-Sample t-test.

ns = non-significant at $P > 0.05$; Significant at * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. SD = standard deviation.

Table II-2. Heritability (h^2) and standard errors (e) of individual fatty acids (FA), grouped fatty acids, and desaturation index (DI) in the Italian Simmental (IS) and Italian Holstein (IH) breeds.

		IS		IH	
FA	FA code	h^2	e	h^2	e
C4:0	FA01	0.4034	0.1384	0.3711	0.2166
C5:0	FA02	0.4311	0.1409	0.0935	0.1851
C6:0	FA03	0.2299	0.1255	0.4891	0.2085
C7:0	FA04	0.3158	0.135	0.1903	0.1993
C8:0	FA05	0.2099	0.122	0.4371	0.2007
C10:0	FA06	0.2201	0.128	0.416	0.2045
C10:1c9	FA07	0.1973	0.1165	0.2174	0.2135
C11:0	FA08	0.3361	0.1363	0.1304	0.1811
C12:0	FA09	0.2432	0.133	0.2493	0.1939
C13-iso	FA10	0.2252	0.1405	0.0482	0.1594
C13-anteiso	FA11	0.1897	0.1203	0.3253	0.2041
C12:1c11	FA12	0.2273	0.1312	0.158	0.1973

C13:0	FA13	0.3566	0.1409	0.1849	0.191
C14-iso	FA14	0.2506	0.1259	0	0
C14:0	FA15	0.2397	0.1342	0.3998	0.2102
C15-iso	FA16	0.2725	0.1398	0.3533	0.2064
C15-anteiso	FA17	0.0944	0.1234	0.4067	0.242
C14:1c9	FA18	0.3525	0.1222	0.7266	0.2334
C15:0	FA19	0.3335	0.1388	0.3588	0.1966
C16-iso	FA20	0.2156	0.1263	0.1116	0.175
C16:0	FA21	0.1314	0.1262	0.3716	0.2058
C16-1t6-7	FA22	0.4001	0.1356	0.0119	0.1715
C16:1t9	FA23	0.2665	0.1318	0.2515	0.1773
C17-iso	FA24	0.0108	0.1195	0.183	0.1919
C16:1 c9	FA25	0.4186	0.1416	0.6043	0.2147
C17-anteiso	FA26	0.1514	0.1249	0	0
C17:0	FA27	0.3995	0.1326	0.1622	0.198
C17:1 c9	FA28	0.1322	0.1204	0.0996	0.203
C18:0	FA29	0.4385	0.1399	0.3636	0.2088
C18:1 t4	FA30	0.1567	0.1231	0.0581	0.1579
C18:1 t5	FA31	0	0	0	0
C18:1 t6-8	FA32	0	0	0	0
C18:1 t9	FA33	0	0	0.132	0.1477
C18:1 t10	FA34	0.0509	0.1094	0.0823	0.1735
C18:1 t11	FA35	0.2686	0.1289	0.0254	0.1676
C18:1 t12	FA36	0.1974	0.1261	0.1818	0.1758
C18:1c9	FA37	0.0331	0.1134	0.3749	0.2146
C18:1 c11	FA38	0.1172	0.1221	0.166	0.1739
C18:1 c12	FA39	0.1648	0.1253	0.2058	0.184
C18:1t16	FA40	0.1238	0.1297	0.1846	0.1902
C18:2 t9.t12	FA41	0.1473	0.1418	0	0
C18:2 t11.c15	FA42	0	0	0.2775	0.2064
C18:2 c9.c12	FA43	0.3116	0.134	0.0505	0.2212
C20:0	FA44	0.4328	0.1371	0	0
C18:3 c9.c12.c15	FA45	0.0264	0.1035	0.3209	0.1998
C18:2 9.11 c/t	FA46	0.3495	0.1297	0.1701	0.2137
C18:4 c6.c9.c12.c15	FA47	0	0	0.3856	0.2136
C20:2 c11.c14	FA48	0	0	0.2278	0.1848
C18-3 c9.t11.c15	FA49	0.1348	0.1252	0.5146	0.2012
C20:3 c8.c11.c14	FA50	0.3723	0.1493	0.1612	0.178

C20:4 c5.c8.c11.c14	FA51	0.1042	0.1151	0.3588	0.2033
C20:5 c5. c8. c11. c14. c17	FA52	0	0	0.2779	0.2036
C22:4 c7.c10.c13.c16	FA53	0.0779	0.1192	0.0511	0.1513
C22:5 c7.c10.c13.C16.c19	FA54	0.0404	0.1016	0.1977	0.2144
C22:6 c4.c7.c10.c13.16.c19	FA55	0.0864	0.1051	0.395	0.2052
SFA	FA56	0	0	0.395	0.2052
UFA	FA57	0	0	0.35	0.2041
PUFA	FA58	0.1912	0.1247	0.3726	0.1938
MUFA	FA59	0	0	0.2647	0.2004
PUFA n6	FA60	0.3052	0.1343	0.043	0.1514
PUFA n3	FA61	0	0	0.4912	0.2081
SCFA (C<10)	FA62	0.1748	0.1229	0.2152	0.1884
MCFA (10<C<17)	FA63	0.1301	0.127	0.2674	0.1936
LCFA (C>17)	FA64	0.1164	0.1308	0.1971	0.1955
BCFA	FA65	0.1812	0.1352	0.2875	0.207
BCFAiso	FA66	0.2405	0.1321	0.3232	0.2226
BCFAanteiso	FA67	0.0819	0.1308	0.2825	0.1918
OCFA	FA68	0.3328	0.1332	0.2323	0.1647
trans 18-1 tot	FA69	0.1678	0.1229	0.2347	0.1658
trans totali	FA70	0.1717	0.123	0.2742	0.192
UFA/SFA	FA71	0	0	0.3573	0.2028
PUFA/SFA	FA72	0.1046	0.1133	0.3754	0.2068
PUFA/(SFA-C18:0)	FA73	0.058	0.1122	0	0
n6/n3	FA74	0	0	0	0
DHA/EPA	FA75	0.047	0.0992	0	0
AA/DHA	FA76	0	0	0.14	0.1815
AA/(DHA+EPA+DPA)	FA77	0	0	0.2836	0.1892
BCFA + OCFA	FA78	0.3392	0.1392	0.0424	0.1598
BCFA/OCFA	FA79	0.3822	0.1344	0.5815	0.2265
BCFAiso/BCFAanteiso	FA80	0.0294	0.1184	0.4728	0.2218
DI 10-1/(10+10-1)	FA81	0.4163	0.1339	0.848	0.2175
DI 14-1/(14+14-1)	FA82	0.4259	0.1283	0.7328	0.2143
DI 16-1/(16+16-1)	FA83	0.4138	0.1448	0.5039	0.2159
DI 18-1/(18+18-1)	FA84	0.3317	0.137	0.3837	0.2084
DI Rum/(vac+rum)	FA85	0.3602	0.1339	0.0612	0.2058

Table II-3. Most significant SNPs associated with milk fatty acid traits in Italian Simmental (IS) and Italian Holstein (IH) breeds. SNP name (SNP), chromosome (Chr), genome position (Position) and GWAS significance p-value (pvalue) are reported.

Breed	Trait	SNP	Chr	Position	pvalue*
IS	C14:0	BovineHD1900014364	19	51349695	9.20E-08
IS	C14:1c9	BovineHD1900014364	19	51349695	9.20E-08
IS	ID 10-1/(10+10-1)	BovineHD2600005467	26	21149234	4.64E-14
IS	ID 14-1/(14+14-1)	BovineHD2600005467	26	21149234	1.15E-17
IS	ID 14-1/(14+14-1)	BTB-00931586	26	21409429	3.99E-08
IH	C14:1c9	BovineHD2600005302	26	20463679	1.04E-08
IH	C14:1c9	BovineHD2600005467	26	21149234	1.54E-09
IH	C14:1c9	BTB-00931481	26	21226405	6.85E-09
IH	C14:1c9	BovineHD2600005491	26	21278993	1.98E-08
IH	C14:1c9	ARS-BFGL-NGS-110077	26	21322557	3.99E-08
IH	C14:1c9	BTB-00931586	26	21409429	1.86E-07
IH	C14:1c9	BovineHD2600005557	26	21479224	6.44E-11
IH	C14:1c9	BovineHD2600005579	26	21555707	1.90E-09
IH	C14:1c9	BovineHD2600005581	26	21564772	1.52E-08
IH	C14:1c9	BovineHD2600005591	26	21598269	2.73E-07
IH	C14:1c9	BovineHD2600005595	26	21629048	3.34E-07
IH	C14:1c9	BovineHD2600005633	26	21878305	6.50E-08
IH	C14:1c9	BovineHD2600005648	26	21926490	3.60E-08
IH	C14:1c9	BTB-00932332	26	22118554	4.71E-08
IH	C14:1c9	ARS-BFGL-NGS-107403	26	22889812	1.99E-07
IH	ID 10-1/(10+10-1)	BovineHD2600005302	26	20463679	1.09E-08
IH	ID 10-1/(10+10-1)	BovineHD2600005467	26	21149234	2.89E-09
IH	ID 10-1/(10+10-1)	Hapmap33073-BTA-162864	26	21180893	4.52E-08
IH	ID 10-1/(10+10-1)	BTB-00931481	26	21226405	2.15E-10
IH	ID 10-1/(10+10-1)	BovineHD2600005491	26	21278993	7.70E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005497	26	21298468	4.09E-07
IH	ID 10-1/(10+10-1)	ARS-BFGL-NGS-110077	26	21322557	2.78E-09
IH	ID 10-1/(10+10-1)	ARS-BFGL-NGS-108305	26	21363670	2.63E-07
IH	ID 10-1/(10+10-1)	BovineHD2600005535	26	21385652	9.60E-08
IH	ID 10-1/(10+10-1)	BTB-00931586	26	21409429	7.61E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005557	26	21479224	3.19E-11
IH	ID 10-1/(10+10-1)	BovineHD2600005567	26	21508165	5.24E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005579	26	21555707	1.18E-10

IH	ID 10-1/(10+10-1)	BovineHD2600005581	26	21564772	1.20E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005591	26	21598269	6.64E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005595	26	21629048	6.43E-08
IH	ID 10-1/(10+10-1)	BovineHD2600005633	26	21878305	1.02E-08
IH	ID 10-1/(10+10-1)	BovineHD2600005648	26	21926490	6.03E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005654	26	21954328	1.32E-07
IH	ID 10-1/(10+10-1)	ARS-BFGL-NGS-116481	26	21977581	1.32E-07
IH	ID 10-1/(10+10-1)	Hapmap24832-BTA-138805	26	22016380	1.32E-07
IH	ID 10-1/(10+10-1)	BovineHD2600005678	26	22018949	1.32E-07
IH	ID 10-1/(10+10-1)	BovineHD2600005686	26	22037112	1.32E-07
IH	ID 10-1/(10+10-1)	ARS-BFGL-NGS-6259	26	22059103	1.32E-07
IH	ID 10-1/(10+10-1)	BovineHD4100017766	26	22094866	1.74E-07
IH	ID 10-1/(10+10-1)	BTB-00932332	26	22118554	7.12E-09
IH	ID 10-1/(10+10-1)	BovineHD2600005698	26	22122641	1.73E-07
IH	ID 10-1/(10+10-1)	ARS-BFGL-NGS-107403	26	22889812	3.41E-08
IH	ID 10-1/(10+10-1)	BovineHD2600006436	26	24918578	1.38E-08
IH	ID 14-1/(14+14-1)	BovineHD2600004833	26	18761989	3.95E-08
IH	ID 14-1/(14+14-1)	BovineHD2600004938	26	19015156	2.69E-07
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-23064	26	20365711	3.13E-07
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-77668	26	20393457	3.13E-07
IH	ID 14-1/(14+14-1)	BovineHD2600005288	26	20427852	1.28E-07
IH	ID 14-1/(14+14-1)	BovineHD2600005302	26	20463679	5.26E-11
IH	ID 14-1/(14+14-1)	BovineHD2600005467	26	21149234	3.01E-12
IH	ID 14-1/(14+14-1)	Hapmap33073-BTA-162864	26	21180893	8.06E-09
IH	ID 14-1/(14+14-1)	BTB-00931481	26	21226405	7.91E-13
IH	ID 14-1/(14+14-1)	BovineHD2600005491	26	21278993	6.53E-11
IH	ID 14-1/(14+14-1)	BovineHD2600005497	26	21298468	4.42E-09
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-110077	26	21322557	3.49E-11
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-108305	26	21363670	4.87E-08
IH	ID 14-1/(14+14-1)	BovineHD2600005535	26	21385652	1.82E-08
IH	ID 14-1/(14+14-1)	BTB-00931586	26	21409429	1.18E-10
IH	ID 14-1/(14+14-1)	BovineHD2600005557	26	21479224	6.13E-14
IH	ID 14-1/(14+14-1)	BovineHD2600005567	26	21508165	2.34E-09
IH	ID 14-1/(14+14-1)	BovineHD2600005579	26	21555707	1.51E-12
IH	ID 14-1/(14+14-1)	BovineHD2600005581	26	21564772	3.18E-11
IH	ID 14-1/(14+14-1)	BovineHD2600005591	26	21598269	1.43E-10
IH	ID 14-1/(14+14-1)	BovineHD2600005595	26	21629048	2.43E-10
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-114149	26	21702714	9.10E-08

IH	ID 14-1/(14+14-1)	BovineHD2600005633	26	21878305	1.40E-11
IH	ID 14-1/(14+14-1)	BovineHD2600005648	26	21926490	8.73E-12
IH	ID 14-1/(14+14-1)	BovineHD2600005654	26	21954328	3.49E-10
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-116481	26	21977581	3.49E-10
IH	ID 14-1/(14+14-1)	Hapmap24832-BTA-138805	26	22016380	3.49E-10
IH	ID 14-1/(14+14-1)	BovineHD2600005678	26	22018949	3.49E-10
IH	ID 14-1/(14+14-1)	BovineHD2600005686	26	22037112	3.49E-10
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-6259	26	22059103	3.49E-10
IH	ID 14-1/(14+14-1)	BovineHD4100017766	26	22094866	1.31E-10
IH	ID 14-1/(14+14-1)	BTB-00932332	26	22118554	5.85E-11
IH	ID 14-1/(14+14-1)	BovineHD2600005698	26	22122641	5.97E-10
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-107403	26	22889812	1.97E-10
IH	ID 14-1/(14+14-1)	BovineHD2600006067	26	23497760	1.46E-09
IH	ID 14-1/(14+14-1)	BovineHD2600006134	26	23847594	1.75E-09
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-111090	26	23920913	2.79E-07
IH	ID 14-1/(14+14-1)	BovineHD2600006239	26	24238250	2.95E-07
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-1092	26	24531763	3.79E-08
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-18194	26	24575207	3.16E-07
IH	ID 14-1/(14+14-1)	BovineHD2600006436	26	24918578	6.66E-10
IH	ID 14-1/(14+14-1)	BovineHD2600006913	26	26152575	3.04E-07
IH	ID 14-1/(14+14-1)	BovineHD2600006920	26	26182662	1.23E-07
IH	ID 14-1/(14+14-1)	BovineHD2600006943	26	26242200	2.98E-07
IH	ID 14-1/(14+14-1)	BTB-00935537	26	26585557	5.90E-08
IH	ID 14-1/(14+14-1)	ARS-BFGL-NGS-71848	26	27213271	1.19E-07

Table II-4. Most significant SNPs (Bonferroni cut-off p-value at 0.05) associated with milk fat percentage (FP) in Italian Holstein (IH) cows. SNP name (SNP), chromosome (Chr), genome position (Position) and GWAS significance p-value (pvalue) are reported.

SNP	Chr	Position	pvalue*
Hapmap30383-BTC-005848	14	1489496	9.29E-10
BovineHD1400000188	14	1588879	1.26E-09
ARS-BFGL-NGS-57820	14	1651311	1.26E-09
BovineHD1400000206	14	1679844	4.24E-10
UFL-rs134432442	14	1736599	1.38E-10

ARS-BFGL-NGS-4939	14	1801116	2.29E-10
BovineHD1400000262	14	1967325	3.36E-08
BovineHD1400000286	14	2069181	6.45E-12
BovineHD1400000467	14	2898515	4.10E-09
ARS-BFGL-NGS-18858	14	2909929	3.93E-09
BovineHD1400000480	14	2936478	3.76E-09
BovineHD1400024350	14	3048650	1.21E-09
BovineHD1400000870	14	4136087	6.10E-08

* Bonferroni cut-off p-value at 0.05

Table II-5. Genes obtained with PIA analysis considering three degrees of interaction with KEGG Lipid Metabolism pathways.

breed	***degree of interaction***	***ensembl gene ID***	***gene symbol***	***pathway***
IS	1DI	ENSBTAG00000015980	*FASN*	Fatty acid biosynthesis
		ENSBTAG00000015178	*ECI2*	Fatty acid degradation
		ENSBTAG00000001868	*PCYT2*	Glycerophospholipid metabolism
		ENSBTAG00000047043	*DCXR*	Arachidonic acid metabolism
		ENSBTAG00000047957	*SCD*	Biosynthesis of unsaturated fatty acids
		ENSBTAG00000045728	*SCD1*	Biosynthesis of unsaturated fatty acids
IS	2DI	ENSBTAG00000016253	*G6PC3*	Glycolysis / Gluconeogenesis
IS	3DI	ENSBTAG00000008747	*DCXR*	Pentose and glucuronate interconversions
		ENSBTAG00000000042	*PYCR1*	Arginine and proline metabolism
		ENSBTAG00000016253	*G6PC3*	Starch and sucrose metabolism
		ENSBTAG00000046173	*ALG12*	N-Glycan biosynthesis
IH	1DI	ENSBTAG00000014335	*CYP17A1*	Steroid hormone biosynthesis
		ENSBTAG00000026356	*DGAT1*	Glycerolipid metabolism
		ENSBTAG00000047957	*SCD*	Biosynthesis of unsaturated fatty acids
		ENSBTAG00000045728	*SCD1*	Biosynthesis of unsaturated fatty acids

IH	2DI	ENSBTAG00000006429	*ACO2*	Citrate cycle (TCA cycle)
		ENSBTAG00000002010	*PI4K2A*	Inositol phosphate metabolism
IH	3DI	ENSBTAG00000011960	*GOT1*	Arginine biosynthesis
		ENSBTAG00000007835	*GPT*	Arginine biosynthesis
		ENSBTAG00000012858	*NT5C2*	Purine metabolism
		ENSBTAG00000000354	*PDE6G*	Purine metabolism
		ENSBTAG00000005311	*POLR3H*	Purine metabolism
		ENSBTAG00000012858	*NT5C2*	Pyrimidine metabolism
		ENSBTAG00000005311	*POLR3H*	Pyrimidine metabolism
		ENSBTAG00000011960	*GOT1*	Alanine, aspartate and glutamate metabolism
		ENSBTAG00000007835	*GPT*	Alanine, aspartate and glutamate metabolism
		ENSBTAG00000011960	*GOT1*	Cysteine and methionine metabolism
		ENSBTAG00000011960	*GOT1*	Arginine and proline metabolism
		ENSBTAG00000011960	*GOT1*	Tyrosine metabolism
		ENSBTAG00000006429	*ACO2*	Glyoxylate and dicarboxylate metabolism
		ENSBTAG00000012858	*NT5C2*	Nicotinate and nicotinamide metabolism
		ENSBTAG00000045703	*COX15*	Porphyrin and chlorophyll metabolism

II - 7. References

Adamska, A., J. Rutkowska, and W. Przybylski. 2016. Comparison of Fatty Acid Composition of Milk from Simmental and Polish Holstein-Friesian Cows in Different Production Seasons. Ann. Anim. Sci. 16:1211–1225. doi:10.1515/aoas-2016-0040.

Akula, N., A. Baranova, D. Seto, J. Solka, M.A. Nalls, A. Singleton, L. Ferrucci, T. Tanaka, S. Bandinelli, Y.S. Cho, Y.J. Kim, J.-Y. Lee, B.-G. Han, B.D.G.S. (BiGS) Consortium, T.W.T.C.-C. Consortium, and F.J. McMahon. 2011. A network-based approach to prioritize results from genome-wide association studies. PLOS ONE 6:e24220. doi:10.1371/journal.pone.0024220.

Albanesi, J., H. Wang, H.-Q. Sun, B. Levine, and H. Yin. 2015. GABARAP-mediated targeting of PI4K2A/PI4KIIα to autophagosomes regulates PtdIns4P-dependent autophagosome-lysosome fusion. Autophagy 11:2127–2129. doi:10.1080/15548627.2015.1093718.

Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. doi:10.1101/gr.094052.109.

Antonicka, H., A. Mattman, C.G. Carlson, D.M. Glerum, K.C. Hoffbuhr, S.C. Leary, N.G. Kennaway, and E.A. Shoubridge. 2003. Mutations in COX15 produce a defect in the mitochondrial heme biosynthetic pathway, causing early-onset fatal hypertrophic cardiomyopathy. Am. J. Hum. Genet. 72:101–114. doi:10.1086/345489.

Arnould, V.M.-R., and H. Soyeurt. 2009. Genetic variability of milk fatty acids. J. Appl. Genet. 50:29–39. doi:10.1007/BF03195649.

Aulchenko, Y.S., D.-J. de Koning, and C. Haley. 2007a. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177:577–585. doi:10.1534/genetics.107.075614.

Aulchenko, Y.S., S. Ripke, A. Isaacs, and C.M. van Duijn. 2007b. GenABEL: an R library for genome-wide association analysis. Bioinforma. Oxf. Engl. 23:1294–1296. doi:10.1093/bioinformatics/btm108.

Banka, S., and W.G. Newman. 2013. A clinical and molecular review of ubiquitous glucose-6-phosphatase deficiency caused by G6PC3 mutations. Orphanet J. Rare Dis. 8:84. doi:10.1186/1750-1172-8-84.

Baumann, C.A., V. Ribon, M. Kanzaki, D.C. Thurmond, S. Mora, S. Shigematsu, P.E. Bickel, J.E. Pessin, and A.R. Saltiel. 2000. CAP defines a second signalling pathway required for insulin-stimulated glucose transport. Nature 407:202–207. doi:10.1038/35025089.

van den Berg, I., D. Boichard, and M.S. Lund. 2016. Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. J. Dairy Sci. 99:8932–8945. doi:10.3168/jds.2016-11073.

Bilal, G., R.I. Cue, A.F. Mustafa, and J.F. Hayes. 2014. Short communication: Genetic parameters of individual fatty acids in milk of Canadian Holsteins. J. Dairy Sci. 97:1150–1156. doi:10.3168/jds.2012-6508.

Bionaz, M., and J.J. Loor. 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366. doi:10.1186/1471-2164-9-366.

Bomba, L., E.L. Nicolazzi, M. Milanesi, R. Negrini, G. Mancini, F. Biscarini, A. Stella, A. Valentini, and P. Ajmone-Marsan. 2015. Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. Genet. Sel. Evol. GSE 47:25. doi:10.1186/s12711-015-0113-9.

Bouwman, A.C., H. Bovenhuis, M.H.P.W. Visker, and J.A.M. van Arendonk. 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet. 12:43. doi:10.1186/1471-2156-12-43.

Bouwman, A.C., M.H. Visker, J.A. van Arendonk, and H. Bovenhuis. 2012. Genomic regions associated with bovine milk fatty acids in both summer and winter milk samples. BMC Genet. 13:93. doi:10.1186/1471-2156-13-93.

Bouwman, A.C., M.H.P.W. Visker, J.M. van Arendonk, and H. Bovenhuis. 2014. Fine mapping of a quantitative trait locus for bovine milk fat composition on Bos taurus autosome 19. J. Dairy Sci. 97:1139–1149. doi:10.3168/jds.2013-7197.

Browning, B.L., and S.R. Browning. 2016. Genotype Imputation with Millions of Reference Samples. Am. J. Hum. Genet. 98:116–126. doi:10.1016/j.ajhg.2015.11.020.

Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987.

Buccioni, A., M. Pauselli, C. Viti, S. Minieri, G. Pallara, V. Roscini, S. Rapaccini, M.T. Marinucci, P. Lupi, G. Conte, and M. Mele. 2015. Milk fatty acid composition, rumen microbial population, and animal performances in response to diets rich in linoleic acid supplemented with chestnut or quebracho tannins in dairy ewes. J. Dairy Sci. 98:1145–1156. doi:10.3168/jds.2014-8651.

Budzyński, M.A., M.C. Puustinen, J. Joutsen, and L. Sistonen. 2015. Uncoupling Stress-Inducible Phosphorylation of Heat Shock Factor 1 from Its Activation. Mol. Cell. Biol. 35:2530–2540. doi:10.1128/MCB.00816-14.

Buitenhuis, B., L.L. Janss, N.A. Poulsen, L.B. Larsen, M.K. Larsen, and P. Sørensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15. doi:10.1186/1471-2164-15-1112.

Cai, F., Y. Miao, C. Liu, T. Wu, S. Shen, X. Su, and Y. Shi. 2018. Pyrroline-5-carboxylate reductase 1 promotes proliferation and inhibits apoptosis in non-small cell lung cancer. Oncol. Lett. 15:731–740. doi:10.3892/ol.2017.7400.

Calero, O., M.J. Bullido, J. Clarimón, R. Hortigüela, A. Frank-García, P. Martínez-Martín, A. Lleó, M.J. Rey, I. Sastre, A. Rábano, J. de Pedro-Cuesta, I. Ferrer, and M. Calero. 2012. Genetic variability of the gene cluster CALHM 1-3 in sporadic Creutzfeldt-Jakob disease. Prion 6:407–412. doi:10.4161/pri.20785.

Capomaccio, S., M. Milanesi, L. Bomba, K. Cappelli, E.L. Nicolazzi, J.L. Williams, P. Ajmone-Marsan, and B. Stefanon. 2015a. Searching new signals for production traits through gene-based association analysis in three Italian cattle breeds. Anim. Genet. 46:361–370. doi:10.1111/age.12303.

Capomaccio, S., M. Milanesi, L. Bomba, E. Vajana, and P. Ajmone-Marsan. 2015b. MUGBAS: a species free gene-based programme suite for post-GWAS analysis. Bioinforma. Oxf. Engl. 31:2380–2381. doi:10.1093/bioinformatics/btv144.

Carroll, S.M., E.J. DePeters, S.J. Taylor, M. Rosenberg, H. Perez-Monti, and V.A. Capps. 2006. Milk composition of Holstein, Jersey, and Brown Swiss cows in response to increasing levels of dietary fat. Anim. Feed Sci. Technol. 131:451–473. doi:10.1016/j.anifeedsci.2006.06.019.

Chen, Z., and Q. Liu. 2011. A new approach to account for the correlations among single nucleotide polymorphisms in genome-wide association studies. Hum. Hered. 72:1–9. doi:10.1159/000330135.

Chu, A.Y., X. Deng, V.A. Fisher, A. Drong, Y. Zhang, M.F. Feitosa, C.-T. Liu, O. Weeks, A.C. Choh, Q. Duan, T.D. Dyer, J.D. Eicher, X. Guo, N.L. Heard-Costa, T. Kacprowski, J.W. Kent, L.A. Lange, X. Liu, K. Lohman, L. Lu, A. Mahajan, J.R. O'Connell, A. Parihar, J.M. Peralta, A.V. Smith, Y. Zhang, G. Homuth, A.H. Kissebah, J. Kullberg, R. Laqua, L.J. Launer, M. Nauck, M. Olivier, P.A. Peyser, J.G. Terry, M.K. Wojczynski, J. Yao, L.F. Bielak, J. Blangero, I.B. Borecki, D.W. Bowden, J.J. Carr, S.A. Czerwinski, J. Ding, N. Friedrich, V. Gudnason, T.B. Harris, E. Ingelsson, A.D. Johnson, S.L.R. Kardia, C.D. Langefeld, L. Lind, Y. Liu, B.D. Mitchell, A.P. Morris, T.H. Mosley, J.I. Rotter, A.R. Shuldiner, B. Towne, H. Völzke, H. Wallaschofski, J.G. Wilson, M. Allison, C.M. Lindgren, W. Goessling, L.A. Cupples, M.L. Steinhauser, and C.S. Fox. 2017. Multiethnic genome-wide meta-analysis of ectopic fat depots

identifies loci associated with adipocyte development and differentiation. Nat. Genet. 49:125–130. doi:10.1038/ng.3738.

Clark, A.G., M.J. Hubisz, C.D. Bustamante, S.H. Williamson, and R. Nielsen. 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15:1496–1502. doi:10.1101/gr.4107905.

Cleeter, M., H. Houlden, P. Simons, R. Al-Shawi, G. Stevanin, A. Durr, J. Hsuan, and T.T. Warner. 2011. Screening for mutations in the phosphatidylinositol 4-kinase 2-alpha gene in autosomal recessive hereditary spastic paraplegia. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 12:148–149. doi:10.3109/17482968.2010.543689.

Conte, G., M. Mele, S. Chessa, B. Castiglioni, A. Serra, G. Pagnacco, and P. Secchiari. 2010. Diacylglycerol acyltransferase 1, stearoyl-CoA desaturase 1, and sterol regulatory element binding protein 1 gene polymorphisms and milk fatty acid composition in Italian Brown cattle. J. Dairy Sci. 93:753–763. doi:10.3168/jds.2009-2581.

D'Andrea, M., S. Dal Monego, A. Pallavicini, M. Modonut, R. Dreos, B. Stefanon, and F. Pilla. 2011. Muscle transcriptome profiling in divergent phenotype swine breeds during growth using microarray and RT-PCR tools. Anim. Genet. 42:501–509. doi:10.1111/j.1365-2052.2010.02164.x.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82 E-Suppl:E313-328. doi:10.2527/2004.8213_supplE313x.

Demirkan, A., C.M. van Duijn, P. Ugocsai, A. Isaacs, P.P. Pramstaller, G. Liebisch, J.F. Wilson, Å. Johansson, I. Rudan, Y.S. Aulchenko, A.V. Kirichenko, A.C.J.W. Janssens, R.C. Jansen, C. Gnewuch, F.S. Domingues, C. Pattaro, S.H. Wild, I. Jonasson, O. Polasek, I.V. Zorkoltseva, A. Hofman, L.C. Karssen, M. Struchalin, J. Floyd, W. Igl, Z. Biloglav, L. Broer, A. Pfeufer, I. Pichler, S. Campbell, G. Zaboli, I. Kolcic, F. Rivadeneira, J. Huffman, N.D. Hastie, A. Uitterlinden, L. Franke, C.S. Franklin, V. Vitart, DIAGRAM Consortium, C.P. Nelson, M.

Preuss, CARDIoGRAM Consortium, J.C. Bis, C.J. O'Donnell, N. Franceschini, CHARGE Consortium, J.C.M. Witteman, T. Axenovich, B.A. Oostra, T. Meitinger, A.A. Hicks, C. Hayward, A.F. Wright, U. Gyllensten, H. Campbell, G. Schmitz, and EUROSPAN consortium. 2012. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. PLoS Genet. 8:e1002490. doi:10.1371/journal.pgen.1002490.

DePeters, E.J., J.F. Medrano, and B.A. Reed. 1995. Fatty acid composition of milk fat from three breeds of dairy cattle. Can. J. Anim. Sci. 75:267–269. doi:10.4141/cjas95-040.

Dreses-Werringloer, U., J.-C. Lambert, V. Vingtdeux, H. Zhao, H. Vais, A. Siebert, A. Jain, J. Koppel, A. Rovelet-Lecrux, D. Hannequin, F. Pasquier, D. Galimberti, E. Scarpini, D. Mann, C. Lendon, D. Campion, P. Amouyel, P. Davies, J.K. Foskett, F. Campagne, and P. Marambaud. 2008. A polymorphism in CALHM1 influences Ca2+ homeostasis, Abeta levels, and Alzheimer's disease risk. Cell 133:1149–1161. doi:10.1016/j.cell.2008.05.048.

Duarte, R.R.R., C. Troakes, M. Nolan, D.P. Srivastava, R.M. Murray, and N.J. Bray. 2016. Genome-wide significant schizophrenia risk variation on chromosome 10q24 is associated with altered cis-regulation of BORCS7, AS3MT, and NT5C2 in the human brain. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. 171:806–814. doi:10.1002/ajmg.b.32445.

Edwards, S.L., J. Beesley, J.D. French, and A.M. Dunning. 2013. Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. 93:779–797. doi:10.1016/j.ajhg.2013.10.012.

Esposito, G., F. Masucci, F. Napolitano, A. Braghieri, R. Romano, N. Manzo, and A. Di Francia. 2014. Fatty acid and sensory profiles of Caciocavallo cheese as affected by management system. J. Dairy Sci. 97:1918–1928. doi:10.3168/jds.2013-7292.

FAO. 2010. Fats and fatty acids in human nutrition. Report of an expert consultation. FAO Food Nutr. Pap. 91:1–166.

Fedida, J., V. Fressart, P. Charron, E. Surget, T. Hery, P. Richard, E. Donal, B. Keren, G. Duthoit, F. Hidden-Lucet, E. Villard, and E. Gandjbakhch. 2017. Contribution of exome sequencing for genetic diagnostic in arrhythmogenic right ventricular cardiomyopathy/dysplasia. PloS One 12:e0181840. doi:10.1371/journal.pone.0181840.

Feitosa, M.F., M.K. Wojczynski, K.E. North, Q. Zhang, M.A. Province, J.J. Carr, and I.B. Borecki. 2013. The ERLIN1-CHUK-CWF19L1 gene cluster influences liver fat deposition and hepatic inflammation in the NHLBI Family Heart Study. Atherosclerosis 228:175–180. doi:10.1016/j.atherosclerosis.2013.01.038.

Fenelon, M.A., and T.P. Guinee. 1999. The Effect of Milk Fat on Cheddar Cheese Yield and Its Prediction, Using Modifications of the Van Slyke Cheese Yield Formula. J. Dairy Sci. 82:2287–2299. doi:10.3168/jds.S0022-0302(99)75477-9.

Fernández-Vizarra, E., and M. Zeviani. 2015. Nuclear gene mutations as the cause of mitochondrial complex III deficiency. Front. Genet. 6:134. doi:10.3389/fgene.2015.00134.

Fisher, R.A. 1919. The Correlation between relatives on the supposition of mendelian inheritance. Earth Environ. Sci. Trans. R. Soc. Edinb. 52:399–433. doi:10.1017/S0080456800012163.

Fox, P.F. 2003. Milk proteins: general and historical aspects. P.F. Fox and P.L.H. McSweeney, ed. Springer US, Boston, MA.

German, J.B., and C.J. Dillard. 2006. Composition, structure and absorption of milk lipids: a source of energy, fat-soluble nutrients and bioactive molecules. Crit. Rev. Food Sci. Nutr. 46:57–92. doi:10.1080/10408690590957098.

Gilmour, A.R., Gogel, B. J., Cullis, B. R., and Thompson, R. 2009. ASReml User Guide Release 3.0. VSN Int. Ltd Hemel Hempstead HP1 1ES.

Gondret, F., A. Vincent, M. Houée-Bigot, A. Siegel, S. Lagarrigue, D. Causeur, H. Gilbert, and I. Louveau. 2017. A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. BMC Genomics 18:244. doi:10.1186/s12864-017-3639-0.

Gondro, C., J. van der Werf, and B. Hayes eds. 2013. Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology. Humana Press.

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

Guernsey, D.L., H. Jiang, S.C. Evans, M. Ferguson, M. Matsuoka, M. Nightingale, A.L. Rideout, S. Provost, K. Bedard, A. Orr, M.-P. Dubé, M. Ludman, and M.E. Samuels. 2009. Mutation in pyrroline-5-carboxylate reductase 1 gene in families with cutis laxa type 2. Am. J. Hum. Genet. 85:120–129. doi:10.1016/j.ajhg.2009.06.008.

Haug, A., A.T. Høstmark, and O.M. Harstad. 2007. Bovine milk in human nutrition – a review. Lipids Health Dis. 6:25. doi:10.1186/1476-511X-6-25.

Hayes, B., and M.E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. GSE 33:209–229. doi:10.1051/gse:2001117.

Hill, W.G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. TAG Theor. Appl. Genet. Theor. Angew. Genet. 38:226–231. doi:10.1007/BF01245622.

Hill, W.G. 1981. Estimation of effective population size from data on linkage disequilibrium1. Genet. Res. 38:209–216. doi:10.1017/S0016672300020553.

Hu, F.B., and W.C. Willett. 2002. Optimal diets for prevention of coronary heart disease. JAMA 288:2569–2578.

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

Jensen, R.G. 2002. The composition of bovine milk lipids: January 1995 to December 2000. J. Dairy Sci. 85:295–350. doi:10.3168/jds.S0022-0302(02)74079-4.

Kalashnikova, L., Y. A. Khabibrakhmanova, and A. Sh. Tinaev. 2009. Effect of polymorphism of milk protein and hormone genes on milk productivity of black pied cows. Russ. Agric. Sci. 35:192–195. doi:10.3103/S1068367409030185.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28:27–30.

Kelsey, J.A., B.A. Corl, R.J. Collier, and D.E. Bauman. 2003. The effect of breed, parity, and stage of lactation on conjugated linoleic acid (CLA) in milk fat from dairy cows. J. Dairy Sci. 86:2588–2597. doi:10.3168/jds.S0022-0302(03)73854-5.

Kitamoto, A., T. Kitamoto, T. Nakamura, Y. Ogawa, M. Yoneda, H. Hyogo, H. Ochi, S. Mizusawa, T. Ueno, K. Nakao, A. Sekine, K. Chayama, A. Nakajima, and K. Hotta. 2014. Association of polymorphisms in GCKR and TRIB1 with nonalcoholic fatty liver disease and metabolic syndrome traits. Endocr. J. 61:683–689.

Knutsen, T.M., H.G. Olsen, V. Tafintseva, M. Svendsen, A. Kohler, M.P. Kent, and S. Lien. 2018. Unravelling genetic variation underlying de novo-synthesis of bovine milk fatty acids. Sci. Rep. 8. doi:10.1038/s41598-018-20476-0.

Kontro, H., J.J. Hulmi, P. Rahkila, and H. Kainulainen. 2012. Cellular and tissue expression of DAPIT, a phylogenetically conserved peptide. Eur. J. Histochem. EJH 56:e18. doi:10.4081/ejh.2012.18.

Kulkarni, S.S., H.K.R. Karlsson, F. Szekeres, A.V. Chibalin, A. Krook, and J.R. Zierath. 2011. Suppression of 5'-nucleotidase enzymes promotes AMP-activated protein kinase (AMPK) phosphorylation and metabolism in human and mouse skeletal muscle. J. Biol. Chem. 286:34567–34574. doi:10.1074/jbc.M111.268292.

Lander, E.S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Lee, S.-K., L.T. Son, H.-J. Choi, and J. Ahnn. 2013. Dicarbonyl/l-xylulose reductase (DCXR): The multifunctional pentosuria enzyme. Int. J. Biochem. Cell Biol. 45:2563–2567. doi:10.1016/j.biocel.2013.08.010.

Lewontin, R.C. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics 49:49–67.

Li, C., D. Sun, S. Zhang, S. Wang, X. Wu, Q. Zhang, L. Liu, Y. Li, and L. Qiao. 2014. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. PloS One 9:e96186. doi:10.1371/journal.pone.0096186.

Li, C., D. Sun, S. Zhang, S. Yang, M.A. Alim, Q. Zhang, Y. Li, and L. Liu. 2016. Genetic effects of FASN, PPARGC1A, ABCG2 and IGF1 revealing the association with milk fatty acids in a Chinese Holstein cattle population based on a post genome-wide association study. BMC Genet. 17. doi:10.1186/s12863-016-0418-x.

Liao, C.-C., T.-T. Ou, C.-H. Wu, and C.-J. Wang. 2013. Prevention of diet-induced hyperlipidemia and obesity by caffeic acid in C57BL/6 mice through regulation of hepatic lipogenesis gene expression. J. Agric. Food Chem. 61:11082–11088. doi:10.1021/jf4026647.

Lin, S., Q. Shi, F.B. Nix, M. Styblo, M.A. Beck, K.M. Herbin-Davis, L.L. Hall, J.B. Simeonsson, and D.J. Thomas. 2002. A novel S-adenosyl-L-methionine:arsenic(III) methyltransferase from rat liver cytosol. J. Biol. Chem. 277:10795–10803. doi:10.1074/jbc.M110246200.

Liu, J.Z., A.F. McRae, D.R. Nyholt, S.E. Medland, N.R. Wray, K.M. Brown, AMFS Investigators, N.K. Hayward, G.W. Montgomery, P.M. Visscher, N.G. Martin, and S. Macgregor. 2010. A versatile gene-based test for genome-wide association studies. Am. J. Hum. Genet. 87:139–145. doi:10.1016/j.ajhg.2010.06.009.

Macciotta, N.P.P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175–8185. doi:10.3168/jds.2015-9500.

Marras, G., G. Gaspa, S. Sorbolini, C. Dimauro, P. Ajmone-Marsan, A. Valentini, J.L. Williams, and N.P.P. Macciotta. 2015. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46:110–121. doi:10.1111/age.12259.

Martini, M., F. Salari, and I. Altomonte. 2016. The Macrostructure of Milk Lipids: The Fat Globules. Crit. Rev. Food Sci. Nutr. 56:1209–1221. doi:10.1080/10408398.2012.758626.

McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P.A. Ioannidis, and J.N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9:356–369. doi:10.1038/nrg2344.

Medrano, J., G. Rincon, and A. Islas-Trejo. 2010. Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. 9th World Congr. Genet. Appl. Livest. Prod. Leipz. Ger. 852.

Mele, M., G. Conte, B. Castiglioni, S. Chessa, N.P.P. Macciotta, A. Serra, A. Buccioni, G. Pagnacco, and P. Secchiari. 2007. Stearoyl-coenzyme A desaturase gene polymorphism and milk fatty acid composition in Italian Holsteins. J. Dairy Sci. 90:4458–4465. doi:10.3168/jds.2006-617.

Mele, M., R. Dal Zotto, M. Cassandro, G. Conte, A. Serra, A. Buccioni, G. Bittante, and P. Secchiari. 2009. Genetic parameters for conjugated linoleic acid, selected milk fatty acids, and milk fatty acid unsaturation of Italian Holstein-Friesian cows. J. Dairy Sci. 92:392–400. doi:10.3168/jds.2008-1445.

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Meuwissen, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie 124:321–322. doi:10.1111/j.1439-0388.2007.00708.x.

Mohammed, M.E., and H.D. Johnson. 1985. Effect of growth hormone on milk yields and related physiological functions of Holstein cows exposed to heat stress. J. Dairy Sci. 68:1123–1133. doi:10.3168/jds.S0022-0302(85)80938-3.

Moioli, B., G. Contarini, A. Avalli, G. Catillo, L. Orrù, G. De Matteis, G. Masoero, and F. Napolitano. 2007. Short communication: Effect of stearoyl-coenzyme A desaturase polymorphism on fatty acid composition of milk. J. Dairy Sci. 90:3553–3558. doi:10.3168/jds.2006-855.

Moore, C.E., J.K. Kay, R.J. Collier, M.J. Vanbaale, and L.H. Baumgard. 2005. Effect of supplemental conjugated linoleic acids on heat-stressed brown swiss and holstein cows. J. Dairy Sci. 88:1732–1740. doi:10.3168/jds.S0022-0302(05)72846-0.

Nakagawa, J., S. Ishikura, J. Asami, T. Isaji, N. Usami, A. Hara, T. Sakurai, K. Tsuritani, K. Oda, M. Takahashi, M. Yoshimoto, N. Otsuka, and K. Kitamura. 2002. Molecular characterization of mammalian dicarbonyl/L-xylulose reductase and its localization in kidney. J. Biol. Chem. 277:17883–17891. doi:10.1074/jbc.M110703200.

Nantapo, C.T.W., V. Muchenje, and A. Hugo. 2014. Atherogenicity index and health-related fatty acids in different stages of lactation from Friesian, Jersey and Friesian×Jersey cross cow

milk under a pasture-based dairy system. Food Chem. 146:127–133. doi:10.1016/j.foodchem.2013.09.009.

Oh-Hashi, K., T. Tejima, Y. Hirata, and K. Kiuchi. 2013. Characterization of the 5'-flanking region of the mouse asparagine-linked glycosylation 12 homolog gene. Cell. Mol. Biol. Lett. 18:315–327. doi:10.2478/s11658-013-0091-2.

Palladino, R.A., F. Buckley, R. Prendiville, J.J. Murphy, J. Callan, and D.A. Kenny. 2010. A comparison between Holstein-Friesian and Jersey dairy cows and their F(1) hybrid on milk fatty acid composition under grazing conditions. J. Dairy Sci. 93:2176–2184. doi:10.3168/jds.2009-2453.

Palmquist, D.L., A. Denise Beaulieu, and D.M. Barbano. 1993. Feed and animal factors influencing milk fat composition. J. Dairy Sci. 76:1753–1771. doi:10.3168/jds.S0022-0302(93)77508-6.

Palmquist, D.L. 2006. Milk fat: origin of fatty acids and influence of nutritional factors thereon. Adv. Dairy Chem. Vol. 2 Lipids 43–92. doi:10.1007/0-387-28813-9_2.

Parodi, P.W. 1999. Conjugated linoleic acid and other anticarcinogenic agents of bovine milk fat. J. Dairy Sci. 82:1339–1349. doi:10.3168/jds.S0022-0302(99)75358-0.

Pegolo, S., A. Cecchinato, M. Mele, G. Conte, S. Schiavon, and G. Bittante. 2016. Effects of candidate gene polymorphisms on the detailed fatty acids profile determined by gas chromatography in bovine milk. J. Dairy Sci. 99:4558–4573. doi:10.3168/jds.2015-10420.

Pegolo, S., N. Mach, Y. Ramayo-Caldas, S. Schiavon, G. Bittante, and A. Cecchinato. 2018. Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. Sci. Rep. 8:566. doi:10.1038/s41598-017-18916-4.

Petrini, J., L.H.S. Iung, M. a. P. Rodriguez, M. Salvian, F. Pértille, G.A. Rovadoscki, L.D. Cassoli, L.L. Coutinho, P.F. Machado, G.R. Wiggans, and G.B. Mourão. 2016. Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie 133:384–395. doi:10.1111/jbg.12205.

Pilarczyk, R., J. Wójcik, P. Sablik, and P. Czerniak. 2015. Fatty acid profile and health lipid indices in the raw milk of Simmental and Holstein-Friesian cows from an organic farm. South Afr. J. Anim. Sci. 45:30–38. doi:10.4314/sajas.v45i1.4.

Possemato, R., K.M. Marks, Y.D. Shaul, M.E. Pacold, D. Kim, K. Birsoy, S. Sethumadhavan, H.-K. Woo, H.G. Jang, A.K. Jha, W.W. Chen, F.G. Barrett, N. Stransky, Z.-Y. Tsun, G.S. Cowley, J. Barretina, N.Y. Kalaany, P.P. Hsu, K. Ottina, A.M. Chan, B. Yuan, L.A. Garraway, D.E. Root, M. Mino-Kenudson, E.F. Brachtel, E.M. Driggers, and D.M. Sabatini. 2011. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476:346–350. doi:10.1038/nature10350.

Pulina, G., A.H.D. Francesconi, B. Stefanon, A. Sevi, L. Calamari, N. Lacetera, V. Dell'Orto, F. Pilla, P.A. Marsan, M. Mele, F. Rossi, G. Bertoni, G.M. Crovetto, and B. Ronchi. 2017. Sustainable ruminant production to help feed the planet. Ital. J. Anim. Sci. 16:140–171. doi:10.1080/1828051X.2016.1260500.

R Development Core Team, R. 2006. A language and environment for statistical computing. Computing 1. doi:10.1890/0012-9658(2002)083[3097:CFHIWS]2.0.CO;2.

Raven, L.-A., B.G. Cocks, and B.J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi:10.1186/1471-2164-15-62.

Reimers, K., C.Y. Choi, V. Bucan, and P.M. Vogt. 2007. The growth-hormone inducible transmembrane protein (Ghitm) belongs to the Bax inhibitory protein-like family. Int. J. Biol. Sci. 3:471–476.

Schennink, A., W.M. Stoop, M.H.P.W. Visker, J.M.L. Heck, H. Bovenhuis, J.J. van der Poel, H.J.F. van Valenberg, and J. a. M. van Arendonk. 2007. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Anim. Genet. 38:467–473. doi:10.1111/j.1365-2052.2007.01635.x.

Schennink, A., J.M.L. Heck, H. Bovenhuis, M.H.P.W. Visker, H.J.F. van Valenberg, and J. a. M. van Arendonk. 2008. Milk fatty acid unsaturation: genetic parameters and effects of stearoyl-CoA desaturase (SCD1) and acyl CoA: diacylglycerol acyltransferase 1 (DGAT1). J. Dairy Sci. 91:2135–2143. doi:10.3168/jds.2007-0825.

Schennink, A., H. Bovenhuis, K.M. Léon-Kloosterziel, J. a. M. van Arendonk, and M.H.P.W. Visker. 2009a. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. Anim. Genet. 40:909–916. doi:10.1111/j.1365-2052.2009.01940.x.

Schennink, A., W.M. Stoop, M.H.P.W. Visker, J.J. van der Poel, H. Bovenhuis, and J. a. M. van Arendonk. 2009b. Short communication: Genome-wide scan for bovine milk-fat composition. II. Quantitative trait loci for long-chain fatty acids. J. Dairy Sci. 92:4676–4682. doi:10.3168/jds.2008-1965.

Schwingshackl, L., and G. Hoffmann. 2012. Monounsaturated Fatty Acids and Risk of Cardiovascular Disease: Synopsis of the Evidence Available from Systematic Reviews and Meta-Analyses. Nutrients 4:1989–2007. doi:10.3390/nu4121989.

Scotti, E., L. Fontanesi, F. Schiavini, V.L. Mattina, A. Bagnato, and V. Russo. 2010. DGAT1 p.K232A polymorphism in dairy and dual purpose Italian cattle breeds. Ital. J. Anim. Sci. 9:e16. doi:10.4081/ijas.2010.e16.

Sharma, N.K., K.A. Langberg, A.K. Mondal, and S.K. Das. 2013. Phospholipid Biosynthesis Genes and Susceptibility to Obesity: Analysis of Expression and Polymorphisms. PLoS ONE 8. doi:10.1371/journal.pone.0065303.

Shingfield, K.J., M. Bonnet, and N.D. Scollan. 2013. Recent developments in altering the fatty acid composition of ruminant-derived foods. Anim. Int. J. Anim. Biosci. 7 Suppl 1:132–162. doi:10.1017/S1751731112001681.

Soyeurt, H., P. Dardenne, A. Gillon, C. Croquet, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler. 2006. Variation in fatty acid contents of milk and milk fat within and across breeds. J. Dairy Sci. 89:4858–4865. doi:10.3168/jds.S0022-0302(06)72534-6.

Soyeurt, H., A. Gillon, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler. 2007. Estimation of heritability and genetic correlations for the major fatty acids in bovine milk. J. Dairy Sci. 90:4435–4442. doi:10.3168/jds.2007-0054.

Stoop, W.M., J. a. M. van Arendonk, J.M.L. Heck, H.J.F. van Valenberg, and H. Bovenhuis. 2008. Genetic parameters for major milk fatty acids and milk production traits of Dutch Holstein-Friesians. J. Dairy Sci. 91:385–394. doi:10.3168/jds.2007-0181.

Stoop, W.M., H. Bovenhuis, J.M.L. Heck, and J. a. M. van Arendonk. 2009a. Effect of lactation stage and energy status on milk fat composition of Holstein-Friesian cows. J. Dairy Sci. 92:1469–1478. doi:10.3168/jds.2008-1468.

Stoop, W.M., A. Schennink, M.H.P.W. Visker, E. Mullaart, J. a. M. van Arendonk, and H. Bovenhuis. 2009b. Genome-wide scan for bovine milk-fat composition. I. Quantitative trait loci for short- and medium-chain fatty acids. J. Dairy Sci. 92:4664–4675. doi:10.3168/jds.2008-1966.

Stull, J.W., and W.H. Brown. 1964. Fatty acid composition of milk. 2. Some differences in common dairy breeds.. J. Dairy Sci. 47.

Świątkiewicz, M., M. Oczkowicz, K. Ropka-Molik, and E. Hanczakowska. 2016. The effect of dietary fatty acid composition on adipose tissue quality and expression of genes related to lipid metabolism in porcine livers. Anim. Feed Sci. Technol. 216:204–215. doi:10.1016/j.anifeedsci.2016.03.020.

Syrstad, O., N. Standal, and Ø. Karijord. 1982. Sources of variation in composition of milk fat. Ann. Génétique Sélection Anim. 14:112. doi:10.1186/1297-9686-14-1-112.

Tullo, E., E. Frigo, A. Rossoni, R. Finocchiaro, M. Serra, N. Rizzi, A.B. Samorè, F. Canavesi, M.G. Strillacci, R.T.M.M. Prinsen, and A. Bagnato. 2014. Genetic Parameters of Fatty Acids in Italian Brown Swiss and Holstein Cows. Ital. J. Anim. Sci. 13:3208. doi:10.4081/ijas.2014.3208.

Utsunomiya, Y.T., M. Milanesi, A.T.H. Utsunomiya, R.B.P. Torrecilha, E.-S. Kim, M.S. Costa, T.S. Aguiar, S. Schroeder, A.S. do Carmo, R. Carvalheiro, H.H.R. Neves, R.C.M. Padula, T.S. Sussai, L.B. Zavarez, R.S. Cipriano, M.M.T. Caminhas, G. Hambrecht, L. Colli, E. Eufemi, P. Ajmone-Marsan, D. Cesana, M. Sannazaro, M. Buora, M. Morgante, G. Liu, D. Bickhart, C.P. Van Tassell, J. Sölkner, T.S. Sonstegard, and J.F. Garcia. 2017. A PLAG1 mutation contributed to stature recovery in modern cattle. Sci. Rep. 7:17140. doi:10.1038/s41598-017-17127-1.

Vanselow, J., and R. Fürbass. 2011. The bovine genome contains three differentially methylated paralogous copies of the P450c17 encoding gene (CYP17A1). Gen. Comp. Endocrinol. 170:475–479. doi:10.1016/j.ygcen.2010.11.003.

Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reinhardt, J. Tetens, G. Thaller, and R. Fries. 2012. Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. PloS One 7:e40711. doi:10.1371/journal.pone.0040711.

van Weeghel, M., H. te Brinke, H. van Lenthe, W. Kulik, P.E. Minkler, M.S.K. Stoll, J.O. Sass, U. Janssen, W. Stoffel, K.O. Schwab, R.J.A. Wanders, C.L. Hoppel, and S.M. Houten. 2012. Functional redundancy of mitochondrial enoyl-CoA isomerases in the oxidation of unsaturated fatty acids. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 26:4316–4326. doi:10.1096/fj.12-206326.

Yang, W.-S., W.-J. Lee, K.-C. Huang, K.-C. Lee, C.-L. Chao, C.-L. Chen, T.-Y. Tai, and L.-M. Chuang. 2003. mRNA levels of the insulin-signaling molecule SORBS1 in the adipose depots of nondiabetic women. Obes. Res. 11:586–590. doi:10.1038/oby.2003.82.

Yoon, S., H.C.T. Nguyen, Y.J. Yoo, J. Kim, B. Baik, S. Kim, J. Kim, S. Kim, and D. Nam. 2018. Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. Nucleic Acids Res. 46:e60–e60. doi:10.1093/nar/gky175.

Zhang, H., X. Hu, Z. Wang, Y. Zhang, S. Wang, N. Wang, L. Ma, L. Leng, S. Wang, Q. Wang, Y. Wang, Z. Tang, N. Li, Y. Da, and H. Li. 2012. Selection signature analysis implicates the PC1/PCSK1 region for chicken abdominal fat content. PloS One 7:e40736. doi:10.1371/journal.pone.0040736.

Zhang, W., J. Zhang, L. Cui, J. Ma, C. Chen, H. Ai, X. Xie, L. Li, S. Xiao, L. Huang, J. Ren, and B. Yang. 2016. Genetic architecture of fatty acid composition in the longissimus dorsi muscle revealed by genome-wide association studies on diverse pig populations. Genet. Sel. Evol. GSE 48:5. doi:10.1186/s12711-016-0184-2.

Zhao, H., D. Nettleton, M. Soller, and J.C.M. Dekkers. 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genet. Res. 86:77–87. doi:10.1017/S001667230500769X.

Zhu, Y., C. Tazearslan, and Y. Suh. 2017. Challenges and progress in interpretation of non-coding genetic variants associated with human disease. Exp. Biol. Med. 242:1325–1334. doi:10.1177/1535370217713750.

Chapter III - TRANSCRIPTIONAL PROFILING OF SWINE MAMMARY GLAND DURING THE TRANSITION FROM COLOSTROGENESIS TO LACTOGENESIS USING RNA SEQUENCING

Abstract

Colostrum and milk are essential sources of antibodies and nutrients for the neonate, playing a key role in their survival and growth. Slight abnormalities in the timing of colostrogenesis/lactogenesis potentially threatens piglet survival. To further delineate the genes and transcription regulators implicated in the control of the transition from colostrogenesis to lactogenesis, RNASeq analysis of swine mammary gland tissue from late-gestation to farrowing was performed.

Three 2nd parity sows were used for mammary tissue biopsies on days 14, 10, 6 and 2 before (-) parturition and on day 1 after (+) parturition. A total of 15 mRNA libraries were sequenced on a HiSeq2500 (Illumina Inc.). The Dynamic Impact Approach and the Ingenuity Pathway Analysis were used for pathway analysis and gene network analysis, respectively.

A large number of differentially expressed genes were detected very close to parturition (-2d) and at farrowing (+1d). The results reflect the extraordinary metabolic changes in the swine mammary gland once it enters into the crucial phases of lactogenesis and underscores a strong transcriptional component in the control of colostrogenesis. There was marked upregulation of genes involved in synthesis of colostrum and main milk components (i.e. proteins, fat, lactose and antimicrobial factors) with a pivotal role of *CSN1S2*, *LALBA*, *WAP*, *SAA2*, and *BTN1A1*. The sustained activation of transcription regulators such as SREBP1 and XBP1 suggest they help coordinate these adaptations.

Overall, the precise timing for the transition from colostrogenesis to lactogenesis in swine mammary gland remains uncharacterized. However, the transcriptomic data results support the hypothesis that the transition occurs before parturition. This is likely attributable to upregulation of a wide array of genes including those involved in 'Protein and Carbohydrate Metabolism', 'Immune System', 'Lipid Metabolism', 'PPAR signaling pathway' and 'Prolactin signaling pathway' along with the activation of transcription regulators controlling lipid synthesis and endoplasmic reticulum biogenesis and stress response.

Index of chapter

III - 1. Introduction....90
III - 1.1. Piglet survival....90
III - 1.2. Swine colostrum....90
III - 1.3. Formation of colostrum....91
III - 1.4. Longitudinal transcriptomic study....92
III - 1.5. RNA sequencing....92
III - 1.6. Differentially expressed genes (DEG) analysis....93
III - 2. Aim of the study....94
III - 3. Materials and methods....94
III - 3.1. Animal sampling and RNA extraction....94
III - 3.2. RNA sequencing....94
III - 3.3. Bioinformatics analysis....95
III - 3.3.1. Identification of differentially expressed genes....95
III - 3.3.2. Dynamic impact approach (DIA)....95
III - 3.3.3. Gene network analysis....96
III - 3.3.4. Verification by real-time PCR....96
III - 4. Results....96
III - 4.1. RNASeq analysis and DEG....96
III - 4.2. DIA results....97
III - 4.2.1. Overall summary of KEGG categories....97
III - 4.2.2. Most impacted pathways....98
III - 4.2.3. Enrichment analysis of genes in most recurrent pathway categories....99
III - 4.3. Gene network analysis results....99
III - 5. Discussion....99
III - 5.1. Protein and Carbohydrate Metabolism....101
III - 5.2. Immune System....104
III - 5.3. Lipid Metabolism....109
III - 5.4. Transcription factors....115
III - 6. Conclusion....119
III - 7. Figures and tables....120
III - 8. References....129

List of chapter figures and tables

Figure III-1. Total number of DEGs due to time resulting from DE analysis of RNASeq.

Figure III-2. KEGG main categories resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data.

Figure III-3. KEGG 'Lipid Metabolism' pathways resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data.

Figure III-4. KEGG ’Endocrine system’ pathways resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data.

Figure III-5. Top 10 upregulated KEGG pathways in -2d vs -14 d comparison resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data.

Figure III-6. Top 10 upregulated KEGG pathways in +1d vs -14 d comparison resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data.

Table III-1. Summary of RNA extraction and quality check for all the samples.

Table III-2. Quantitative real time PCR (qPCR) validation of sequencing (Seq) results.

Table III-3. RNA sequencing and alignment for all the samples.

Table III-4. Sequencing read alignments to the reference genome.

Table III-5. Top ten upregulated genes in -2vs-14 comparison.

Table III-6. Summary of upregulated genes of most recurrent subcategory in both and specific comparisons.

Table III-7. Summary Transcription Regulators (TRs) in both and specific comparisons.

In supplementary material section

Supplementary Table III-S1. Summary of upstream transcription regulators (TR) in -2vs-14 and +1vs-14 time comparisons obtained by IPA.

III - 1. Introduction

III - 1.1. Piglet survival

Piglet survival is a major problem especially in modern pig production (Theil et al., 2014). When a piglet is being born, its risk of dying is greater than at any other stage of life. In fact, piglets are born deficient of energy, but at the same time, they have a very high energy requirements. To be able to survive, newborn piglets rely upon three different sources of energy; glycogen, colostrum and transient milk, which help to cover their energy requirements. Energy originating from oxidized glycogen, colostrum and transient milk all contribute to maintain a constant body temperature and to keep the piglets alive. If one of these energy sources fails to supply adequate amounts of energy, the piglet ends up dying either because of hunger or because it eventually is crushed by the sow because the piglet is too weak for adequate movements.

Any change in nutritional or management strategies for late gestating sows that favors transfer of energy from the sow to the offspring is particularly important. This is so crucial in pigs, more than in other species, because piglets do not have fat depots or brown adipose tissue that is present in calves, lambs and rodents (Pastorelli et al., 2009). Nevertheless, although several studies have been done with the goal of increasing glycogen depots with changes to sow nutrition over the years (Seerley et al., 1974; Newcomb et al., 1991; Jean and Chiang, 1999; Pastorelli et al., 2009; Theil et al., 2011), it is still questionable whether glycogen deposition in fetuses can be stimulated by sow nutrition in late gestation (Theil et al., 2014). At the same time, currently, not much is known on how sow nutrition affects colostrum yield (Theil et al., 2014), although attempts to alter macrochemical composition of colostrum by sow nutrition have been made (Nissen et al., 1994; Dividich et al., 2005; Loisel et al., 2013). In view of this, an interesting contribution could be offered by breeding selection programs for udder quality traits and in particular for colostrum quality and quantity (Balzani et al., 2016).

III - 1.2. Swine colostrum

Colostrum is an essential source of antibodies and nutrients for the neonate, playing a key role in their survival and growth (Salmon, 2000). Sow colostrum can be defined as the mammary secreta ingested by neonatal piglets until 24 h after birth of the first piglet (Devillers et al., 2004b). Colostrum can be considered as the 'elixir of life' because of the high abundance of

many different components (macronutrients and micronutrients and bioactive molecules such as immunoglobulins, growth factors and enzymes), and many of these components are important for survival of the newborns and proper development of the gastrointestinal tract (Mei et al., 2006; Bjornvad et al., 2007). Lactose and fat in colostrum serve a main purpose of supplying energy, whereas the protein fraction promotes the transfer of immunity (immunoglobulins), stimulation of growth (growth factors) and facilitation of fat digestion (enzymes) (Theil et al., 2014). Approximately, one-third of sows produce less colostrum than the recommended level of 250 g colostrum/piglet which would be adequate for survival and proper growth (Quesnel et al., 2012).

At present, it is not known exactly when or at which rate colostrum is being synthesized in the mammary gland, or when colostrum synthesis starts and ceases. Most of the colostrum is produced before the first piglet is born and, consequently, colostrum yield cannot be dependent only on piglets suckling (Theil et al., 2014).

III - 1.3. Formation of colostrum

Mammogenesis occurs during prepuberty, puberty and gestation and continues during lactation as long as the teats are suckled (Farmer et al., 2004). If the teats are not suckled, involution will occur. This involution is especially rapid if the teats are not suckled during the first seven to ten days of lactation (Kim et al., 2001). Mammogenesis is slow during the first two thirds of gestation and more rapid during the last third (Ji et al., 2006). In particular, the development of mammary gland is crucial during the final stages of gestation when alveoli begin to distend (Ji et al., 2006) and there is an abrupt increase in the concentration of colostrum and milk constituents just prior to parturition (Kensinger et al., 1982). These stages coincides with the early 'peripatum' period, which goes from roughly one week pre-farrowing to several days post-farrowing. One of the dynamic shifts in mammary tissue function occurs in the peripartum period, with the exponential growth during late gestation culminating in the process of colostrogenesis/lactogenesis, and then shifting to initiation of lactation (galactopoiesis) (Farmer, 2006). Lactogenesis is generally subdivided in two stages: lactogenesis I, which in swine is initiated in late-pregnancy (around day 105) and is linked to the initiation of synthesis of milk-specific components and to structural and metabolic differentiation of the mammary glands, and lactogenesis II, which is characterized by the onset of copious milk secretion (Hartmann et al., 1997). There is some discrepancy in the literature about the description of the

switching from lactogenesis I to lactogenesis II in sow (Farmer, 2006). Nevertheless, it is generally accepted that colostrum production takes place during lactogenesis I (Farmer, 2006). The presence of this incomplete timing information is justifiable since colostrogenesis and lactogenesis are complex biological processes. It is clear that in this context, characterizing the transcriptome profile and the metabolic and signaling pathways during this essential period of reproduction, when any abnormalities in the timing of colostrogenesis/lactogenesis might threaten piglet survival (Kensinger et al., 1982), could be of great interest.

III - 1.4. Longitudinal transcriptomic study

Longitudinal (or time-course) transcriptomic studies are ideally-suited for unravelling complex biological behavior at a genome-wide level and provide a more detailed view of the underlying physiological adaptations over time (Zhao et al., 2013). Longitudinal designs have two principal motivations. (1) Increase the precision of a treatment by eliminating individual variation. (2) Examine the individual's changing response over time (Cook and Ware, 1983). In this regard, the development of high-throughput technologies has revolutionized time-course study (Zhang and Davis, 2014), as well as transcriptome analysis in general (Fontanesi et al., 2011). Particularly, RNASeq technology enables the generation of more extensive transcriptome information providing an advantage over microarray analyses, due to its capability to quantify all transcripts (Kukurba and Montgomery, 2015) and not only those present on the arrays.

III - 1.5. RNA sequencing

RNA molecules are essential components of all living cells. Understanding the identity and abundance of each RNA molecule in a given cell under a specific condition is the ultimate goal of gene expression analysis. The first decade of this millennium witnessed the advent of Next Generation Sequencing (NGS) technologies: a revolution in biology researcher for its ability to acquire an unprecedented amount of data in a short time. Nowadays, RNASeq is the method of choice to study gene expression profile and identify novel RNA species (Hrdlickova et al., 2017). Compared to DNA microarray-based methods, RNASeq offers less background noise and a greater dynamic range for detection. Most importantly, RNASeq directly reveals sequence identity, crucial for analysis of unknown genes and novel transcript isoforms. Although, several different technologies have been developed for RNASeq (Nookaew et al., 2012; Adiconis et al., 2013; Li et al., 2014b; Han et al., 2015), generically a typical RNASeq experiment consists

of isolating RNA, converting it to complementary DNA (cDNA), preparing the sequencing library, and sequencing it on an NGS platform. Depending on the experimental goals, there are several possible choices particularly referred to library construction. (Kukurba and Montgomery, 2015). Currently, several NGS platforms are commercially available (Metzker, 2010). The majority of high-throughput sequencing platforms use a sequencing-by-synthesis method to sequence tens of millions of sequence clusters in parallel. In recent years, the sequencing industry has been dominated by Illumina, which applies an ensemble-based sequencing-by-synthesis approach (Bentley et al., 2008).

III - 1.6. Differentially expressed genes (DEG) analysis

The correct identification of differentially expressed genes (DEGs) between specific conditions is a key in the understanding phenotypic variation. As already explained, RNASeq has become the main option for these studies. The number of methods and software for differential expression analysis data also increased rapidly during the last decade (Costa-Silva et al., 2017). Overall, these methods can be grouped into two main subsets: parametric and non-parametric. Parametric methods capture all information about the data within the parameters. In these cases, it is possible to predict the value of unknown data from observing the adopted model and its parameters. In other words it is assumed that each expression value for a given gene is mapped into a particular distribution, such as Poisson (Marioni et al., 2008; Bullard et al., 2010; Hardcastle and Kelly, 2010) or negative binomial (Robinson and Smyth, 2007; Anders and Huber, 2010; Robinson et al., 2010). Regarding the RNASeq differential expression analysis, some tools such as edgeR (Robinson et al., 2010) and baySeq (Hardcastle and Kelly, 2010), adopt the negative binomial model as the main approach. Other software tools, such as NOIseq (Tarazona et al., 2015) and SAMseq (Li and Tibshirani, 2013), adopt non-parametric methods, i.e. not imposing a rigid model to be fitted. To be thorough it is worth noting other methods, based on transcript detection, specifically developed for the identification of unknown transcripts or isoforms, can also be applied to DEG analysis, such as EBSeq (Leng et al., 2013) and Cuffdiff2 (Trapnell et al., 2013). Nowadays, there is not a consensus about which methodology is most appropriate or which approach ensures the validity of the results in terms of robustness, accuracy and reproducibility (Costa-Silva et al., 2017). This topic in Bioinformatics research is still developing (Zhang et al., 2014).

III - 2. Aim of the study

In recent years, RNASeq technology has been applied to the study of lactating mammary glands in several species (Suárez-Vega et al., 2016). Although previous studies using microarrays have provided some preliminary insights into the differential expression of genes (DEG) in sow mammary glands during the peripartum period (Zhao et al., 2013), our understanding of metabolic or signaling pathways in this species is still limited. The aim of this study was to provide a comprehensive transcriptome profiling of the sow mammary gland from 14 days prior to parturition to day 1 in lactation using RNASeq analysis and functional bioinformatics tools such as the Dynamic Impact Approach (DIA) (Bionaz et al., 2012b) and Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA).

III - 3. Materials and methods

III - 3.1. Animal sampling and RNA extraction

Details of the experiment design are reported elsewhere (Krogh et al., 2017a). Briefly, all procedures involving animals were in compliance with Danish laws and regulations for the humane care and use of animals in research (The Danish Ministry of Justice. 1995. Animal testing act, consolidation act no. 726 of September 9, 1993 (as amended by act no. 1081 of December 20, 1995). The Danish Ministry of Justice, Copenhagen, Denmark, 1995). Mammary tissue collected on days 14, 10, 6 and 2 before (-) parturition and on day 1 after (+) parturition from three 2nd parity crossbred sows (Danish Lande race × Yorkshire) with the highest colostrum yield (among 9 sows) were used for the present analysis. RNA isolation and quality evaluation was performed following the protocols previously described (Tramontana et al., 2008). The average yield of total RNA (from 20.3 ± 6.9 mg tissue) was 44 ± 19 µg, and the average RNA integrity number (Agilent Bioanalyzer) was 8.2 ± 0.8. An aggregate summary of RNA extraction and quality check for all the samples is reported in Table III-1.

III - 3.2. RNA sequencing

Sequencing was performed using "High-Throughput Sequencing and Genotyping Unit" of the W. M. Keck Biotechnology Center at the University of Illinois at Urbana Champaign (Urbana, IL, USA). A total of 15 mRNA libraries were quantified by qPCR and sequenced on two lanes for 101 cycles from one end of the fragments on a HiSeq2500 (Illumina Inc.), using v4 HiSeq

SBS reagents. In total approximately 403 million single-read sequences of 100 nt in length were collected. Quality control metrics were performed on raw sequencing reads using the FASTQC v0.11.15 application. Using STAR (v2.5.1b), an index of the reference genome was built and single-end clean reads for each individual were aligned to the reference genome. Reads were mapped and annotated to the *Sus scrofa* genome (v10.2.86), downloaded from the EnsemblGenome website (Nov. 2016). Aligned reads were quantified with the Subread package (v1.5.0) based on the Refseq gene annotation.

III - 3.3. Bioinformatics analysis

III - 3.3.1. Identification of differentially expressed genes

Non-expressed and weakly expressed genes, defined as having <1 read per million in *n* of the samples, where *n* is the size of time group replicates, were removed prior to differential expression (DE) analysis (Anders et al., 2013) A TMM (trimmed mean of M-values) normalization was applied to all samples using edgeR (Robinson and Oshlack, 2010). Following log transformation of the data, limma-voom method (Bioconductor packages) was used to conduct DE analyses. The limma module utilizes a standard variance moderated across all genes using a Bayesian model and produces p-values with greater degrees of freedom (Ritchie et al., 2015). The voom module was used to transform the data based on observational level weights derived from the mean-variance relationship prior to statistical modeling, where time was considered as fixed effect and animal as random effect (Law et al., 2014). Differentially expressed genes across different time points were defined as genes with a Benjamini–Hochberg multiple-testing adjusted p-value of ≤ 0.05. To identify the longitudinal transcriptional gene response close to parturition, the time point -14 day was used as baseline for each time comparison. In order to highlight the metabolic processes underlying mammary changes associated with the colostrogenesis and the onset of lactogenesis in the last stages of gestation leading up to parturition, we relied on DEGs between -10vs-14, -6vs-14, -2vs-14 and +1vs-14 time comparisons.

III - 3.3.2. Dynamic impact approach (DIA)

The DIA software (Bionaz et al., 2012b) was used for functional analyses. Briefly, DIA uses the systems information from the KEGG database and ranks pathways calculating the overall impact (e.g., biological importance of a given pathway as a function of the change in expression

of genes composing the pathway) and flux (direction of impact; e.g, average change in expression as up-regulation/activation, down-regulation/inhibition, or no change) of biological pathways. For this purpose, the whole dataset with Entrez gene IDs, FDR, FC, and p-values of each time comparison were uploaded in DIA and an overall cut-off (FDR and p-value ≤ 0.05) was applied as the threshold.

III - 3.3.3. Gene network analysis

Ingenuity Pathway Analysis (IPA) was performed to identify transcription regulators and their networks with other genes, within the list of significant DEG (similar cut-off as DIA analysis; FDR and p-value ≤ 0.05) at each time comparison. Software features and the IPA knowledge base were used for the analysis (https://www.qiagenbioinformatics.com).

III - 3.3.4. Verification by real-time PCR

The expression of *LALBA*, *CSN2*, *PAEP*, and *LTF* was analyzed to verify the physiologic response of the mammary gland as farrowing approached. These genes are well-established markers of mammary-specific genes. Complete information about cDNA synthesis and qPCR performance are reported elsewhere (Vailati-Riboni et al., 2016). After normalization with the geometric mean of three internal control genes (*API5*, *VABP*, and *MRPL39*), qPCR data were log2 transformed prior to statistical analysis to obtain a normal distribution. Statistical analysis was performed with SAS (v 9.4). Normalized, log2 transformed data were subjected to ANOVA with PROC MIXED. The statistical model included time (-14, -10, -6, -2, and +1 day from farrowing) as the fixed effect, and sow as the random effect. The Kenward-Roger statement was used for computing the denominator degrees of freedom. Fold change for the time comparisons -10vs-14, -6vs-14, -2vs-14, and 1vs-14 were then calculated from the estimates of the model. For each of the four genes and comparisons FDR, fold change, and p values are reported in Table III-2, together with the respective results from the sequencing analysis.

III - 4. Results

III - 4.1. RNASeq analysis and DEG

An aggregate summary of RNA sequencing and alignment for all the samples is reported in Table III-3. Illumina sequencing was effective at producing large numbers of high-quality reads

from all samples. On average, 92% of the total reads were mapped successfully. Among the aligned reads, 91.8% were mapped to unique genomic regions. Results for total number of DEG due to time are presented in figure 1. Considering an FDR and p-value ≤ 0.05 among the 9393 genes (after annotation with the entrez genes ID) a total of 0, 17 (15 upregulated and 2 downregulated), 788 (451 upregulated and 337 downregulated), and 2884 (1508 upregulated and 1376 downregulated) were differentially expressed for -10vs-14, -6vs-14, -2vs-14, +1vs-14 time comparisons, respectively.

For further analysis with DIA (Bionaz et al., 2012b) and IPA we focused on DEG between -2vs-14 and +1vs-14 time comparisons, where the largest numbers of activated and inhibited genes were detected. The -2vs-14 comparison represents the difference in gene expression patterns between a gland with limited growth and a gland that is near full-term, i.e. genes that encompass the last stages of functional differentiation. In contrast, the +1vs-14 comparison represents the difference in mammary tissue between a stage with limited mammary growth and a functional mammary gland which had entered into the lactogenesis stage (Zhao et al., 2013). To highlight the overall weight of genes in each comparison, the top ten upregulated genes were underscored (Tables III-4 and -5). *CSN1S2* and *LALBA* were the most-expressed genes at 2d prepartum, whereas *WAP*, *CSN1S2*, *SAA2* and *LALBA* had a marked upregulation at 1d postpartum. The overlap and specific upregulated genes between the last two time comparisons are reported in Table III-4.

III - 4.2. DIA results

III - 4.2.1. Overall summary of KEGG categories

The DIA results are summarized in Figure III-2. They provide an overview of impact and flux for each KEGG category calculated following DIA procedures (Bionaz et al., 2012b). We clearly observed no significant changes in -10vs-14 and -6vs-14 time comparisons because of the lack of DEG associated with these comparisons (data not shown). However, closer to parturition (-2vs-14 comparison), we detected an evident activation of all main categories and in particular 'Metabolism category' and 'Organismal Systems' pathways, which became stronger considering the postpartum stage (+1vs-14). Focusing only on these main categories and considering the related subcategories with flux value at least 50% of impact value, clearly within 'Metabolism' the subcategory 'Lipid Metabolism' was the most-impacted and recurrent in both comparisons, followed by 'Metabolism of Other Amino Acids'. It was not possible to

highlight a recurrent subcategory within 'Organismal Systems' with flux value at least 50% of impact value in the last 2 comparisons, thus, we chose the recurrent subcategory with the highest impact and upregulated flux: 'Endocrine system'. Regarding the other KEGG pathway categories (i.e. 'Genetic Information Processing', 'Environmental Information Processing' and 'Cellular Processes'), we observed a marked downregulation of all 'Genetic Information Processing' subcategories and a marked upregulation of 'Environmental Information Processing' at 1 day postpartum. The above general results only provide information about the overall impact and the general direction of the impact (flux) of each significant category and subcategory in the dataset. With the aim of reaching a better understanding of the biological relevance of each significantly impacted category/subcategories, we focused on the single metabolic pathways falling within the main subcategories of interest particularly 'Lipid Metabolism' and 'Endocrine system' (Figures III-3 and -4).

Except 'Primary bile acid biosynthesis', for the 'Lipid Metabolism' subcategory we uncovered a clear upregulation of all pathways. In particular, 'Fatty acid biosynthesis' was the most impacted and upregulated pathway in both comparisons followed by 'Arachidonic acid metabolism' and 'Steroid hormone biosynthesis' at 2d prepartum, and 'Steroid biosynthesis' and 'Synthesis and degradation of ketone bodies' at 1d postpartum. Within 'Endocrine system', we detected a marked upregulation of 'Prolactin signaling pathway' recurrent in the last 2 comparisons, followed by 'Ovarian steroidogenesis' and 'PPAR signaling pathway' at 2d prepartum and 'PPAR signaling pathway' and 'Thyroid hormone synthesis' at 1d postpartum. There was a marked downregulation of 'Adipocytokine signaling pathway' at 1d postpartum and 'Insulin signaling pathway' in the last 2 comparisons.

III - 4.2.2. Most impacted pathways

To highlight the overall most impacted and upregulated pathways, we considered all pathways without any category classification and with flux value at least 50% of impact value (Figures III-5 and -6). When the last stage (-2d vs -14d) of gestation was considered, 'Fatty acid biosynthesis', 'Retinol metabolism', 'Drug metabolism – other enzymes', 'PPAR signaling pathway', 'Galactose metabolism', 'Steroid hormone biosynthesis', 'Metabolism of xenobiotics by cytochrome P450', 'Chemical carcinogenesis', 'Fatty acid degradation', 'Arachidonic acid metabolism' were the most impacted and upregulated pathways. When the parturition stage (+1d vs -14d) was considered, 'PPAR signaling pathway', 'Steroid

biosynthesis', 'Fatty acid biosynthesis', 'Synthesis and degradation of ketone bodies', 'Mineral absorption', 'beta-Alanine metabolism', 'Galactose metabolism', 'Fatty acid degradation', 'Drug metabolism - other enzyme' and 'Histidine metabolism' were the most impacted and upregulated pathways.

III - 4.2.3. Enrichment analysis of genes in most recurrent pathway categories

As shown above, 'Lipid metabolism' and 'Endocrine system' were the most recurrent pathway subcategories. A KEGG enrichment analysis was performed to identify the most upregulated genes among these subcategories (FDR and p-value ≤ 0.05). To underscore the weight of a specific gene within the most recurrent KEGG pathway categories, the genes used by DIA for impact and flux calculations were extrapolated and divided into three groups: upregulated genes in '+1vs-14', '-2vs-14' and 'both' time comparisons. The summary of upregulated genes is reported in Table III-6.

III - 4.3. Gene network analysis results

IPA allowed the uncovering of relationships between transcription factors and DEG. Considering a ± 2 'Activation z-score' value and p-value cut-off of 0.01, we identified 6 and 55 upstream transcription regulators (TR) in -2vs-14 and +1vs-14 time comparisons, respectively (Supplementary Table III-S1). To highlight and summarize similarities and differences in the activation of TR, the overlap between the two comparisons was performed and the activated TR were extracted. The results are reported in Table III-7.

III - 5. Discussion

Although there is some discrepancy in the literature as to the specific timing of colostrogenesis and lactogenesis, the consensus is that swine lactogenesis is activated in late-pregnancy between day 100 and 110 of gestation (Farmer et al., 2006)(Zhao et al., 2013). Lactogenesis is further subdivided in two stages: lactogenesis I, which is initiated in late-pregnancy and is linked to the initiation of synthesis of colostrum and milk specific components and to structural and metabolic differentiation of the mammary gland (MG); and lactogenesis II, which is characterized by the onset of copious milk secretion (Farmer et al., 2006). It is generally accepted that colostrum production takes place during lactogenesis I and that the transition from

colostrum to mature milk occurs within 1 to 2 days postpartum, when removal of colostrum from the mammary glands enhances the rate of fat secretion and accelerates the increase in lactose concentrations towards that of mature milk (Farmer et al., 2006). Our results partly confirmed these timings. Our results indirectly suggest that at 14 days prior to parturition (day 100 of gestation) the MG had already entered lactogenesis stage I, hence, the extremely low number of DEG when comparing both -10d, and -6d to -14d stage. The large numbers (788) of DEG uncovered in the -2dvs-14d time comparison reflects a strong activation of many metabolic processes compatible with the shifting from stage I to stage II of lactogenesis occurring before parturition in this species (Bussmann et al., 1996). This is consistent with the consideration that MG reached the greatest degree of structural development at that time, and the preparation for copious milk synthesis and secretion had begun (Farmer, 2012).
The transition period from a non-lactating to lactating state requires important metabolic changes to enable the shift of nutrient prioritization from body reserves towards the mammary gland for milk production. It is clear that the marked number of DEG (2884) detected at day 1 postpartum vs 14 days prepartum reflects the extraordinary metabolic changes in the swine mammary gland once it fully entered into crucial phases of lactogenesis. A general overview of DIA results confirm this conclusion. In fact, the overall activation of all 'Metabolic pathways' and in particular of 'Lipid Metabolism', 'Metabolism of Other Amino Acids', 'Carbohydrate Metabolism' and 'PPAR signaling pathways' is compatible with the transition from a non-lactating to lactating state. This initial stage involves cellular development of the milk synthesis apparatus, and expression of genes associated with synthesis of milk components (i.e. milk proteins, fat and lactose) (Kensinger et al., 1982). At the same time, the general inhibition of 'Genetic Information Processing', the specific inhibition of 'Cell Growth and Death' in 'Cellular Process' and of 'Development' in 'Organismal System' is consistent with the fact that the mammary gland has already significantly grown in mass (Ji et al., 2006b).
At farrowing, the MG is fully involved in the accumulation and secretion of colostrum and milk, with their nutritional and immunological proprieties confirmed by the activation of 'Immune System', 'Endocrine System' and 'Excretory System'. The role of prolactin secretion, which peaks around farrowing (Devillers et al., 2004a) stimulates the gland to switch from formation and accumulation of colostrum to synthesis and secretion of milk components, is supported by the upregulation of 'Prolactin signaling pathway'. In summary, our results indicate that the transition from colostrogenesis to lactogenesis occurs between 6 and 2 days

before expected parturition. This is likely attributable to upregulation of a wide array of genes including those involved in 'Protein and Carbohydrate Metabolism', 'Immune System', 'Lipid Metabolism', 'PPAR signaling pathways' and 'Prolactin signaling pathway'.

III - 5.1. Protein and Carbohydrate Metabolism

It is known that concentrations of total protein in sow mammary secretions are highest at parturition (Hurley, 2014). The evident upregulation in the last comparison of 'Amino Acid Metabolism' and 'Metabolism of Other Amino Acids' pathways is consistent with the strong activation of synthesis of the major milk protein during the onset of lactation. These changes in protein concentrations mirror the changes in immunoglobulin content, highly abundant in colostrum with a gradual decrease in milk, and concurrently a lower casein content in colostrum followed by a high increase during the postpartum period (Hurley, 2014). While immunoglobulin concentrations decline significantly and β-lactoglobulin levels are relatively constant from the colostrum period through lactation, the proportion of casein and α-lactalbumin increases considerably in the postpartum period (Hurley, 2014). These previously reported responses are confirmed in our results by the strong upregulation of *CSN2*, *CSN1S2*, *CSN1S1* and *LALBA*, which are among the overall top upregulated genes in the last 2 comparisons along with the marked upregulation of *WAP* at 1 day postpartum. The upregulation of *LALBA* deserves particular consideration because of its involvement in lactose biosynthesis (Ramakrishnan et al., 2001). Lactose is the major carbohydrate in sow milk and the major osmole in milk, responsible for drawing water into the secretory vesicles (Hurley, 2014). Lactose concentrations are low in colostrum then increase gradually over the first 2 to 3 days of lactation (Hurley, 2014). Our transcriptomic results confirmed this evidence, showing a marked upregulation of 'Carbohydrate Metabolism' during the transition from late pregnancy to parturition, driven by upregulation of many genes and particularly by *LALBA*, *B4GALT1* and *HK2* activity all involved in 'Galactose Metabolism' pathway.

Milk Caseins. There was a marked upregulation of CSN1S2 (alpha-S2-casein) at 2d prepartum and 1d postpartum [FC = 629.57 and 3858.51]. Several studies reported CSN1S2 as one of the most up-regulated caseins increasing in expression during lactation in bovine (Gao et al., 2013), pig and mouse (Bionaz et al., 2012a). It was also found to be expressed in colostrum and mid lactation milk in goats (Crisà et al., 2016). The temporal expression pattern of α-casein genes is similar in many species with CSN1S2 as the most upregulated followed by CSN1S1. Our

results also confirmed this pattern in pig with a lower expression of CSN1S1 [FC = 4.00 and 5.25] compared with CSN1S2.

An expected result was the marked upregulation of CSN2 [FC = 31.01 and 33.52], a member of the β-casein family. Beta casein is the principal protein in human milk and the primary source of essential amino acids for a suckling infant. The increases in expression from 2d prepartum to 1d postpartum of *CSN2* was not surprising because it is known that the proportion casein of total milk protein sharply increases by 24 h postpartum (Hurley, 2014) and because caseins, together with the whey proteins, represent the 90% of milk protein fraction (Farrell et al., 2004). In this regard, it is also known that the mRNA abundance of WAP in monogastrics appears to be as high or higher than caseins (Bionaz et al., 2012a). This is in agreement with our results at 1d postpartum where we found the abrupt upregulation of *WAP* (whey acidic protein) [FC = 4205.80]. Whey acidic protein is the major whey protein in the milk of many species, including the pig where it is secreted at a consistent level throughout lactation (Simpson et al., 1998). The increase in expression of *WAP* in monogastrics was proportional to *LALBA* (Bionaz et al., 2012a). Our result confirmed this relationship [*LALBA*, FC = 128.51 and 1275.53].

It is well established that milk protein is affected by energy content of the diet and at the same time by the availability of amino acids (AA) (Bionaz et al., 2012a). In this sense, the transport of AA is one of the major limitations for milk protein synthesis. A comprehensive review of AA transporters in the mammary gland and their functional and molecular regulation was recently conducted by Shennan and Boyd (Shennan and Boyd, 2014). Furthermore, the marked upregulation of *SLC7A4* in the last comparison [FC = 32.27] is noteworthy. *SLC7A4* codes for the CAT-4 protein, which is related to other members of the SLC7 family of cationic amino acid transporters found highly expressed in swine placental tissue (Vallet et al., 2014). Considering that *SLC7A4* is known as an important paralog of *SLC7A1*, coding for the CAT-1 protein, which was identified in porcine MG where its abundance increases at early lactation compared with prepartum and it is positively correlated to β-CN and α-LA (Manjarin et al., 2011), we speculate that *SLC7A4* in mammary epithelial cells (MEC) could enhance the mammary uptake of leucine (Leu), hence, stimulating protein synthesis through activation of the mTOR cell signaling pathway (Rezaei et al., 2016; Gao et al., 2015). Krogh et al. (Krogh et al., 2017b) showed recently that Leu is the most extracted AA by the sow mammary gland in early lactation (d+3). In this study, we detected an upregulation [FC = 2.04] at 1d postpartum of another solute carrier family: *SLC7A8*, coding for LAT2, that, together with LAT1, have

been proposed to be involved in Leu uptake in the mammary gland (Shennan et al., 2002). In the same way, insulin signaling plays an important role in the control of milk protein synthesis by inducing translation via activation of the mTOR pathway (Bionaz et al., 2012a). The insulin effect prevents mTOR inhibition by blocking (via phosphorylation) the tuberous sclerosis proteins (i.e., TSC1 and TSC2), which are the main inhibitors of mTOR. In this regard, it was interesting to note the downregulation of *TSC1* at the postpartum stage [FC = -1.24]. The marked upregulation of *CTGF* (connective tissue growth factor) at 1d postpartum [FC = 13.16], which contributes to and is required for lactogenic differentiation in mouse mammary gland (Morrison et al., 2010), also was noteworthy.

Lactose synthesis. Expression of *LALBA*, encoding α-lactalbumin, was strongly upregulated at 2d prepartum and 1d postpartum [FC = 128.51 and 1275.53]. This is one of the main milk proteins involved in 'Carbohydrate Metabolism' via activation of 'Galactose metabolism'. In fact, LALBA is a component of the lactose synthetase complex that uses glucose and UDP-galactose as substrates for the synthesis of lactose in the Golgi complex (Messer and Elliott, 1987). Our result is in agreement with findings of other studies, where *LALBA* upregulation was detected towards the end of gestation, just before parturition (Robinson et al., 1995; Theil et al., 2005). This result is also consistent with the low α-lactalbumin concentrations in swine colostrum, and the gradual increase along with lactose through the first days of lactation (Hurley, 2014). However, if glands are not suckled by 12 th after parturition (i.e. during the colostrum period), expression of LALBA is decreased 24 h after parturition in response to lack of colostrum removal (Theil et al., 2006).

Regarding the lactose synthase enzyme complex, it was noteworthy that in the last 2 comparisons *B4GALT1* was upregulated [FC = 2.40 and 3.27]. The *B4GALT1* gene encodes one of seven beta-1,4-galactosyltransferase (beta4GalT) proteins of the complex and is unique because it participates both in glycoconjugate and lactose biosynthesis. In fact, the first enzyme in the pathway adds galactose to N-acetylglucosamine residues that are either monosaccharides or the nonreducing ends of glycoprotein carbohydrate chains. The second enzyme is restricted to lactating mammary tissue where it forms a heterodimer with α-lactalbumin to catalyze UDP-galactose + D-glucose <=> UDP + lactose (Ramakrishnan et al., 2001).

The transport of UDP-galactose into the Golgi is regulated by *SLC35A2*, which is considered a rate-limiting process in lactose synthesis (Mohammad et al., 2012). The expression of this gene was upregulated in the 2 last comparisons [FC = 2.05 and 2.01]. The marked upregulation of

HK2 [FC = 5.02] at 1d postpartum is important in the context of lactose biosynthesis. In fact, hexokinase (HK) is considered to have a potential controlling step for glucose availability for lactose synthesis (Mohammad et al., 2012). In rodents, *HK2* is detected only after parturition and it was speculated that the presence of *HK2* during lactation may lead to both an increase in free glucose for lactose synthesis and increased activity of the pentose phosphate shunt to generate reducing equivalents for lipogenesis (Kaselonis et al., 1999).

III - 5.2. Immune System

The concept that milk, mammary secretions, and the mammary gland have major roles in immune defense has long been proposed (Wheeler et al., 2007). It is well-established that both colostrum and milk proteins have nutritive and immunological functions for the newborn (Sanchez et al., 1992). This is crucial for pigs that have an epitheliochorial placenta impermeable to immunoglobulins (Ig) (Salmon, 2000), thus, neonate survival depends upon the passive acquisition of maternal immunity. (Kruse, 1983). Immunoglobulins are the primary protein components of colostrum with an immunological function (Hurley and Theil, 2013). Immunoglobulin G, in particular, is the major immunoglobulin in sow colostrum and its concentration remains elevated for the initial hours postpartum and then starts to decline consistently (Hurley, 2014). In bovine it is known that a large amount of IgG immunoglobulins are transferred from the blood stream across the mammary barrier into colostrum and milk by a specific transport mechanism (Larson et al., 1980). In pigs it would also appear that colostrum is not a true mammary secretion since 90% of its immunoglobulin content is of serum origin (Bourne and Curtis, 1973).

The transport of immunoglobulins from the maternal plasma across the mammary barrier into the colostrum is highly-selective (Mayer et al., 2005) and it is known that FcRn plays an important role in the IgG transport during colostrum formation in several species (Mayer et al., 2002; Lu et al., 2007). In this regard, our results showed no-differential expression of *FCGRT* (Fc fragment of IgG receptor and transporter) among time comparisons. Considering the time-window of our experiment, we speculate that this result is consistent with the need for sustained expression of *FCGRT* as a way to support colostrum synthesis. In fact, it is known that FcRn expression coincides with Stage 1 lactogenesis (the onset of colostrogenesis) (Barrington et al., 1999). Although our data did not reveal transcriptional activation of immunoglobulin synthesis, there was an evident overall increase of MG innate immune response and production of

antimicrobial factors during lactation. In fact, it should be noted that colostrum and milk not only contain immunoglobulins, but also contain a range of antimicrobial factors and factors that may impact the immune system (Hurley and Theil, 2011). In this sense, our results highlighted the upregulation from late pregnancy to parturition of several genes that are involved in innate immune response in swine MG. Some have direct or indirect antimicrobial, chemoattractant and pathogen recognition activity including lactoferrin, haptoglobin, serum amyloid A-2 protein, chemokine, osteopontin, toll like receptor, ceruloplasmin.

***Antimicrobial components and chemoattractant activity*.** Antimicrobial proteins naturally present in colostrum and milk have the ability to kill and inhibit a broad spectrum of bacteria. In this regard, the marked upregulation of *HP* (haptoglobin) in the last 2 comparisons [FC = 29.27 and 204.86] was noteworthy. Haptoglobin is an acute-phase protein responsive to inflammation and infection (Lai et al., 2009) that has already been shown to exert immune modulating functions on the innate and adaptive immune system of the pig (Hiss-Pesch et al., 2011). At 1d postpartum there was also a significant upregulation of *LTF* (lactotransferrin) [FC = 3.92]. This gene is a member of the transferrin gene family and is a major iron-binding protein in milk and body secretions with an antimicrobial activity, making it an important component of the non-specific immune system (Wheeler et al., 2007). Our result is consistent with the fact that lactotransferrin concentrations in swine colostrum at parturition are high and remain elevated through day 3 of lactation, and then decline by day 7 (Hurley, 2014).

Milk is also known to exert a potent chemotactic activity on neutrophils (Rainard et al., 2008). The prompt recruitment of neutrophils is crucial for the containment of a number of pathogens at sites of infection, and is considered an important arm of innate host defenses against pathogenic microorganisms (Nathan, 2006). In this sense, the role of the chemokine superfamily that encode secreted proteins involved in immunoregulatory and inflammatory processes must be underscored. Both *CXCL2* and *CXCL10* encode chemokine antimicrobial proteins with a marked upregulation at 1d postpartum [FC = 17.87 and 5.34, respectively]. Bovine colostrum contains all main chemokines (CXCL1, CXCL2 and CXCL3), but concentrations of CXCL2 are generally the lowest and decrease sharply such that it is undetectable in milk after few days of lactation (Rainard et al., 2008). From that standpoint, the strong upregulation of *growth-regulated protein homolog gamma* (also known as *CXCL3*) at 1d postpartum [FC = 36.64] is noteworthy. CXCL3/GRO-gamma is involved in the chemokine signaling pathway and (in the absence of inflammation) is considered the major chemotactic

factor for neutrophils secreted constitutively into milk (Rainard et al., 2008). Our results emphasized a major role of *CXCL2* and *CXCL3* in the transition from colostrum to mature milk in swine, probably to help in the prompt recruitment of neutrophils.

We also detected a marked upregulation of *C7* [FC = 7.43] at 1d postpartum. This gene encodes a serum glycoprotein that forms a membrane attack complex together with complement components C5b, C6, C8, and C9 as part of the terminal complement pathway of the innate immune system. In bovine milk, these complement components are found in high concentrations in the first 2 days after parturition and then decrease during the following days (Zhang et al., 2015). In the last 2 time comparisons we also detected the upregulation of *C4A* (Complement C4A) [FC = 2.05 and 5.73], which acts in concert with other complement components to hasten the destruction of pathogens by phagocytes (Janeway CA Jr et al., 2001). It is known that milk and colostrum are rich in host-resistance factors, among the others C4 and C3 proactivators (Goldman, 1977). Even in the absence of cognate interactions, the complement system participates in innate immunity providing efficient and rapid protection (Trégoat et al., 1999). The levels of complement fractions C3 and C4 have been studied in the human transition from colostrum to mature milk, where C3 and C4 decrease over lactation with a highest concentration of C3 in colostrum and a highest concentration of C4 in mature milk (Trégoat et al., 1999). Thus, our results confirm a similar trend in the pig.

The marked upregulation of ceruloplasmin (*CP*) [FC = 258.47] was consistent with previous studies in pigs, where expression of *CP* increases in late pregnancy and especially upon lactation, with a correlation between the degree of mammary mRNA expression and the content of milk ceruloplasmin (Cerveza et al., 2000). Although a specific function for CP in the mammary gland is unknown, it may participate in the metabolism of copper (Cerveza et al., 2000).

The antimicrobial protein encoded by *LYZ* (lysozyme) was downregulated in the last 2 comparisons [FC = -3.22 and -2.66]. Lysozyme has nonspecific antimicrobial activity that is present in many secretions, tissues, and phagocytic cells of mammals but the role in swine mammary secretions is not clearly understood, even if it is thought to contribute to overall antibacterial activity (Wagstrom et al., 2000). Krakowski et al. (Krakowski et al., 2002) reported lysozyme activity in sow colostrum immediately after parturition, however Chandan et al. (Chandan et al., 1968) did not find lysozyme activity in sow milk. Proinflammatory cytokines mediate the early local and systemic responses to microbial challenges and may play

a key role in development of the neonatal immune system (Nguyen et al., 2007). In this regard, it was also interesting that IPA results showed a pattern of cytokines predicted to be activated: TNF, IFNG, OSM, IL6, IL1B, TNFSF11, IL5, IFNL1, CSF3, TNFSF13B, IL13, IFNB1, IL1A, IFNA2, IFNA1/IFNA13, IL15, IFNL4, THPO, and IFNK. We also detected the upregulation at 1d postpartum of *IL13RA1* (Interleukin 13 Receptor Subunit Alpha 1) [FC = 1.95], *IL15* (Interleukin 15) [FC = 1.57], *IL17RB* (Interleukin 17 Receptor B) [FC = 2.91], *TNFSF13* (Tumor Necrosis Factor Superfamily Member 13) [FC = 2.31], *TNFRSF12A* (TNF Receptor Superfamily Member 12A) [FC = 1.76], and *TNFRSF1A* (TNF Receptor Superfamily Member 1A) [FC = 1.66]. The presence or transfer of these cytokines has not been studied in porcine colostrum/milk and there is also a lack of information for humans and other species about the persistence or function of these maternal cytokines in neonates after transfer via suckling (Nguyen et al., 2007). The upregulation of *F7* and *F10* (Coagulation Factor VII and X) at 1d postpartum [FC = 6.44 and 5.01] also is noteworthy because it is known that the coagulation system is part of the innate immune system and its local activation has been found to play an important role in the early host response to infection (van der Poll and Herwald, 2014).

There was an abrupt upregulation of *SAA2* (serum amyloid A-2 protein) [FC = 3039.23] at 1d postpartum. This isoform is considered the most predominant member of the *SAA* family expressed in the swine mammary gland (Rodriguez et al., 2009). Porcine SAA mRNA production increased during lactation and stimulates the neonatal immune response by enhancing the recruitment of mucosal gut B lymphoblasts (potentially influencing Ig concentrations) conferring active and passive protection on neonates and providing local protection for the mammary gland (Rodriguez et al., 2009). The upregulation of *CD14* [FC = 24.65] at 1d postpartum was noteworthy. The protein encoded by this gene is a surface antigen that cooperates with other proteins to mediate the innate immune response to bacterial lipopolysaccharide. Colostrum has high concentrations of soluble CD14 that decrease over time, with the highest concentration detected in “transitional” milk (0 to 4 d postpartum) (Lee et al., 2003). Considering the enrichment of sCD14 in colostrum and milk, Filipp et al. (Filipp et al., 2001) speculated it plays a role in actively stimulating the immune system and homeostasis of IgM of the suckling neonate.

The marked upregulation of *SPP1* (also known as *OST*) at 1d postpartum [FC = 14.61] is in agreement with data from RNA isolated from colostrum and mid lactation milk from goats in which it was the most upregulated gene (Crisà et al., 2016). *SPP1*, encoding the osteopontin

protein, is considered essential for mammary gland development, milk production, local mammary gland immunity and seems to have a significant role in the modulation of milk protein gene expression (Dudemaine et al., 2014; Sheehy et al., 2009). Despite these biologic associations, the precise role of osteopontin in the mammary gland is unclear. Several studies have shown an association between the expression of the *SPP1* and milk yield by enhancing the expression of CSN2 (Sheehy et al., 2009), but it has also been related to mammary gland morphogenesis (Nemir et al., 2000) and newborn immunity (Alain et al., 2009). Our result seems to suggest a biologic role of this gene during swine lactation but further analyses are required.

Pathogen recognition. We detected the upregulation of *TLR2* [FC = 4.45 and 7.32] at 2d prepartum and 1d postpartum. This gene encodes a protein member of the Toll-like receptor (TLR) family and plays a fundamental role in pathogen recognition and activation of innate immunity (He et al., 2016). The main bacterial ligands for TLR2 are peptidoglycan and lipoteichoic acid (LTA) of Gram-positive bacteria (Rainard and Riollet, 2006). *TLR2* is known to be expressed in mammary epithelial cells in bovine, where, the recognition of specific molecular motifs (i.e. PAMP), determines a rapid and complex innate cascade (Stelwagen et al., 2009). We also detected the upregulation at 1d postpartum of *TLR4* [FC = 2.04], the main signaling receptor for most bacterial LPS, and the major component of the outer membrane of Gram-negative bacteria. TLR4 also acts as the signal-transducing receptor for whole Gram-negative bacteria and for the fusion protein from respiratory syncytial virus (LeBouder et al., 2003).

The marked upregulation of *LBP* in the last 2 comparisons was surprising [FC = 372.18 and 20.89]. The lipopolysaccharide-binding protein (LBP) is one of the most-abundant proteins during infections with Gram-negative bacteria, and is involved in the acute-phase immunologic response. The main function of this protein is to bind bacterial lipopolysaccharides (LPS) expressed on the outer cell wall of bacteria, acting as a carrier for LPS and to help control LPS-dependent monocyte responses (Stelwagen et al., 2009). The expression of LBP was also demonstrated in mouse mammary gland early during involution, accompanied by a strong increase in the expression of CD14 protein (Stein et al., 2004). Cow colostrum also contains LBP (Nissen et al., 2012). Our result seems to confirm an important role of LBP in swine mammary gland.

The upregulation of *LY96* (Lymphocyte Antigen 96) [FC = 5.02] also appears biologically-relevant in the context of pathogen recognition. This gene encodes a protein associated with TLR4 on the cell surface and confers responsiveness to LPS, thus, providing a link between the receptor and LPS signaling. It is known that TLR4 cooperates with LY96 and CD14, both of which were upregulated at 1d postpartum and could indicate a response to mediate the innate immune response to bacterial LPS (Poltorak et al., 1998; Tsukamoto et al., 2010). The upregulation of *ICAM1* at 1d postpartum was significant [FC = 4.42]. This gene encodes a cell surface glycoprotein, which is typically expressed on endothelial cells and cells of the immune system. The fact that human milk contains substantial amounts of slCAM-1 indicates that it could affect the immune system of the neonate (Xyni et al., 2000). This gene could also have a similar role in swine colostrum and milk.

III - 5.3. Lipid Metabolism

There was an evident activation of all lipid-related pathways very close to the parturition (2d prepartum). This is consistent with the consideration that mammary tissue is preparing to begin copious milk synthesis and secretion. The further upregulation of Lipid Metabolism pathways at 1d postpartum confirmed this. This is consistent with the fact that the mammary gland retains fat in late gestation and synthesizes great amounts of *de novo* fat in early lactation (d+3) (Krogh et al., 2017b; Hurley, 2014). Our results underscored that this transition is likely attributable to upregulation of many genes, including those involved in *de novo* fatty acid (FA) synthesis, FA activation and desaturation, cholesterol synthesis and ketone body utilization.

FA de novo synthesis. We observed at 1d postpartum the upregulation of *ACACB* (acetyl-CoA carboxylase-β) [FC = 3.65]. ACC is a complex multifunctional enzyme system, catalyzing the carboxylation of acetyl-CoA to malonyl-CoA, the rate-limiting step in fatty acid synthesis. This result is consistent with expression profiles of genes involved in *de novo* FA synthesis of human mammary gland during secretory activation (Mohammad and Haymond, 2013), where a progressive increase of ACACB activity by day 4 postpartum was shown. In contrast, in mouse mammary gland *ACACA* was the only isoform with significant upregulation of expression during lactation, while *ACACB* expression did not differ between pregnancy and lactation (Han et al., 2010). *ACACA* and *ACACB* are distinct genes that respectively encode the isoenzymic ACC proteins ACCα and ACCβ. The *ACACA* gene is expressed at its highest levels in the lipogenic tissues and provides cytoplasmic malonyl-CoA for FA synthesis. The *ACACB* gene

is implicated in the regulation of β-oxidation of FA in the mitochondria (Abu-Elheiga et al., 2000). *FASN* encodes another rate-controlling enzyme in lipogenesis that works in concert with ACACA activity. Both genes play a key role in regulating *de novo* FA synthesis in bovine mammary gland (Bionaz and Loor, 2008a). However, we did not detect differential expression in *FASN* or *ACACA* between late-gestation and early lactation (d+1). Whether this represents a unique feature of the swine mammary gland will have to be established in future experiments.

FA desaturation genes. The primary enzyme involved in monounsaturated FA synthesis is stearoyl-CoA desaturase (SCD), an important enzyme in the mammary gland, which introduces a double bond in the Δ-9 position of myristoyl-, palmitoyl-, and stearoyl-CoA, primarily (Bionaz and Loor, 2008a). The expression of *SCD* was upregulated at 1d postpartum [FC = 5.99], and appears to be central during milk fat synthesis at the onset of lactation in swine mammary gland. This result is consistent with the expression of desaturases in bovine during lactation (Kinsella, 1972; Bionaz and Loor, 2008a) but is contrary to data from human mammary epithelial cells where its expression decreased over the first 72 h and then gradually increased by day 21 of lactation (Mohammad and Haymond, 2013).

The synthesis of very-long-chain FA is carried out by fatty acid desaturase 1 (*FADS1*) and 2 (*FADS2*), which adds double bonds at the Δ-5 and Δ-6 position of PUFA (Xie and Innis, 2008). FADS1 is involved in the synthesis of the long-chain PUFA arachidonic acid, eicosapentaenoic acid and docosahexaenoic acid. The stage of lactation alters mammary *FADS1* and *FADS2* expression in bovine (Bionaz and Loor, 2008a), rat (Rodriguez-Cruz et al., 2006; Rodriguez-Cruz et al., 2011), and mouse, with a marked upregulation after parturition. In the latter, FADS1 compared with FADS2 mRNA had a more pronounced and significant upregulation after parturition (Han et al., 2010). Yantao Lv et al. (Lv et al., 2015) suggested that from late-pregnancy and throughout lactation the swine mammary gland participates in LC-PUFA synthesis by altering the expression of FADS1 and FADS2. The authors speculated that *FADS1* instead of FADS2-3 might play a major role in the biosynthesis of LC-PUFA in the lactating porcine mammary gland. Our results are in agreement with this consideration, since we found *FADS1* gradually increased in the last 2 comparisons (2d prepartum and 1d postpartum) [FC = 2.59 and 4.35], while *FADS2* was significantly activated only in the postpartum period [FC = 2.32].

Glycerol backbone activation. To synthesize triacylglycerides (TAG), both fatty acyl-CoAs and glycerol 3-phosphate must be readily available (Gonzalez-Baró et al., 2007). The major

steps in the pathway of TAG synthesis in mammary gland have been elucidated (West et al., 1972; Lin et al., 1976; Harvatine et al., 2009). The activation of the glycerol carbon backbone, which is needed for further acylation, is the first and crucial step for further TAG assembly, and enzymes encoded by glycerol kinase (*GK)* and diacylglycerol kinase alpha (*DGKA*) play an important role. In particular, glycerol can enter the mammary epithelial cells from the plasma to be phosphorylated by GK (Mohammad and Haymond, 2013), and in the last comparison we detected the upregulation of *GK* [FC = 2.41]. We also detected a moderate upregulation in expression of *DGKA* in the last 2 time comparisons [FC = 1.42 and 1.75]. DGKA plays an important role in the resynthesis of phosphatidylinositol and phosphorylation of diacylglycerol to phosphatidic acid. This indicates that the activation of the glycerol carbon backbone, which is needed for further acylation, is crucial during the onset of lactation in swine as in humans (Mohammad and Haymond, 2013).

FA internalization and activation. Passive diffusion of FA across membranes plays a minor role compared with protein-mediated FA uptake and the flip-flop mechanism (Bionaz and Loor, 2008a). The main proteins involved in FA uptake in non-ruminant cells include fatty acid translocase FAT/CD36 (CD36) and fatty acid transport proteins (FATP or SLC27A) (Doege and Stahl, 2006). In bovine, *CD36* was associated with mammary fatty acid uptake from the blood after parturition (Bionaz and Loor, 2008a). Our results underscored a strong upregulation of CD36 [FC = 7.47], hence, confirming its pivotal role in swine mammary gland. The strong upregulation of *ACSL6* (acyl-CoA synthetase long-chain family member 6) in the last 2 comparisons [FC = 26.04 and 37.58] confirmed the importance of ACSL family member isoforms for FA activation during the onset of lactation in swine as in humans (Mohammad and Haymond, 2013) and bovine (Bionaz and Loor, 2008b). In fact internalized FAs must be esterified with CoA in the inner face of the plasma membrane via acyl-CoA (ASC) prior to participating in further metabolism (Bionaz and Loor, 2008a). In this regard, the upregulation of *ACSL3* at 1d postpartum [FC = 5.62] is noteworthy given results reported by Yantao Lv et al. (Lv et al., 2015). They reported that ACSL3 is the most abundant isoform in the porcine mammary gland, in contrast to ACSL1 which is the main isoform in lactating bovine (Bionaz and Loor, 2008b) and human (Mohammad and Haymond, 2013) mammary cells. The authors speculated that in swine mammary gland ACSL3 channels LCFA mainly towards TAG synthesis during lactation. This consideration was based on the fact that in the rats ACSL3 prefers C16-C20 unsaturated FA (Fujino et al., 1996), which are major constituents of FA in

sow milk (Lv et al., 2015). The downregulation of *ACSL1* and *ACSL5* at 2 days prepartum [FC = -1.44 and -1.84] seems to support the idea of *ACSL3* being more important during lactation.

Acyltransferases and TG assembly. From late pregnancy to onset of lactation, we detected the upregulation in the last 2 comparisons of a cluster of genes involved in the first and rate-limiting step in the TAG biosynthesis pathway, i.e. *GPAT3*, *GPAT4* and *AGPAT1* [FC = 2.04 , 2.09 and 1.44; 3.13, 4.46 and 1,47 respectively]. The first committed step in TAG synthesis via the glycerol phosphate pathway is catalyzed by GPAT (glycerol-3-phosphate acyltransferase) enzymes, which reside in the endoplasmic reticulum (ER) and mitochondria (Takeuchi and Reue, 2009). These enzymes add fatty-acyl groups to the sn-1 position of glycerol-3-phosphate, leading to the production of monoacylglycerols (MAG) (Gonzalez-Baró et al., 2007). The second acylation step in the glycerol phosphate pathway is the conversion of lysophosphatidate to phosphatidate via AGPAT, which adds an acyl group to the sn-2 position of the glycerol backbone. (Takeuchi and Reue, 2009).

Regarding specific isoforms uncovered in our results, *GPAT3* is a gene with a controversial identity and current evidence suggests that it has both GPAT and AGPAT activities (Takeuchi and Reue, 2009). Similar to *GPAT3*, *GPAT4* was also initially classified as an *AGPAT6* based on high amino acid similarity to *AGPAT1* and *AGPAT2*. After careful examination of enzyme activity, it was found instead to be a second ER-localized GPAT and renamed as GPAT4 (Takeuchi and Reue, 2009). Nagle et al. (Nagle et al., 2008) revealed that, when expressed in cultured cells, GPAT4 can utilize a variety of substrates, including C12:0-, C16:0-, C18:0-, C18:1-, C18:2-, and C20:4-CoA substrates (Nagle et al., 2008). *AGPAT1*, which was upregulated in the last 2 time comparisons is a well-established AGPAT isoform, for which the enzyme activity has been validated (Leung, 2001) and with a preference for C12–16:0, C16:1, C18:2, and C18:3, followed by C18:0, C18:1, and C20:4, but with a poor activity for C20:0 and C24:0 (Takeuchi and Reue, 2009). Interestingly, *AGPAT1* can also catalyze ATP-independent acyl-CoA and LPA (lysophosphatidic acid) synthesis from PA (phosphatidic acid), the reverse of the normal AGPAT reaction (Yamashita et al., 2001). This reversible activity suggested that it may be involved in the regulation of the levels of LPA and PA available to act as signaling molecules (Takeuchi and Reue, 2009). *AGPAT1* (1-Acylglycerol-3-Phosphate O-Acyltransferase 1) was also discovered to have a crucial role also during *de novo* synthesis of triacylglycerol in bovine mammary gland during lactation (Bionaz and Loor, 2008b).

Once synthesized and activated, FAs are esterified to glycerol-3-phosphate to produce TAG (Lv et al., 2015). Both *GPAM* and *DGAT1* are responsible for the first and last step of esterification leading to TAG synthesis (Bionaz and Loor, 2008a). *GPAM* (glycerol-3-phosphate acyltransferase, mitochondrial) is a well-known gene, mostly expressed in tissues with high lipogenic activity and plays a key role in phospholipid and TAG biosynthesis (Tomàs et al., 2003). *DGAT1* (diacylglycerol acyltransferase 1) is well-characterized gene and catalyzes the esterification of the last FA to diacylglycerol leading to TAG synthesis. In the present study both *GPAM* and *DGAT1* were upregulated only at 2d prepartum [FC = 1.91 and 1.31]. In the case of *DGAT1*, this result is not in agreement with studies in bovine and human, where its upregulation occurred postpartum (Bionaz and Loor, 2008a; Mohammad and Haymond, 2013). This may suggest that *DGAT1*, compared with other genes involved in TAG synthesis, is of minor importance in the overall process of milk fat synthesis in the pig. In a recent study, however, a western blot analysis of DGAT1 and other proteins in porcine mammary tissue confirmed its increase during lactation compared with late-pregnancy (Lv et al., 2015). Because our time frame of interest was around colostrogenesis, further protein expression and functional studies during these times would have to be conducted to clarify the importance of *DGAT* in colostrogenesis. Thus, we do not believe our findings contradict previous functional studies (Grisart et al., 2004) demonstrating a pivotal role for *DGAT1* in milk TAG synthesis. The fact remains that DGAT1 is one of many proteins composing the TAG synthesis pathway (Coleman and Lee, 2004).

Because they act in an interdependent manner to optimize lipid homeostasis in various tissues, it is currently believed that LPIN protein function and its role in glycerolipid synthesis are influenced by intricate functional interactions among the various LPIN family members (Csaki et al., 2014; Dwyer et al., 2012). Lv et al. (Lv et al., 2015) argued for a major role of *LPIN1* in TAG synthesis in the porcine mammary gland during lactation. In the present study, the upregulation of *LPIN1* at 1d postpartum [FC = 2.26] seems to confirm this argument. The LPIN1 isoform was reported to be the most abundant among LPIN isoforms in human (Mohammad and Haymond, 2013), mouse (Han et al., 2010), and bovine mammary tissue (Bionaz and Loor, 2008b), with a marked upregulation during lactation.

Lipid droplet formation in milk. Milk fat globules are formed in the ER membrane via incorporation of newly-formed TAG, transported to the apical membrane, and eventually released during milk secretion (Keenan and Mather, 2006). Well-defined proteins involved in

these processes in mammary gland include butyrophilin (BTN1A1) and xanthine dehydrogenase (XDH) (Bionaz and Loor, 2008a), having a function as a structural protein in milk fat droplets in the lactating mammary gland (Murakami et al., 2014). Furthermore, it is known that perilipins are a family of proteins localized in the periphery of intracellular lipid droplets that are essential for droplet formation (Tansey et al., 2004). Our data is consistent with this evidence and support a significant role of all *BTN1A1*, *XDH* and *PLIN5* genes in swine mammary lipid droplet formation. In fact, in the last 2 time comparisons we detected a strong upregulation of *BTN1A1* [FC = 90.49 and 202.51] and at 1d postpartum we detected a marked upregulation of *XDH* [FC = 17.24] and *PLIN5* [FC = 21.42]. This is in agreement with a recent study showing that fat is taken up in substantial amounts by sow mammary glands in late gestation (Krogh et al., 2017b).

***Cholesterol synthesis genes*.** The shift of nutrients from body stores towards the mammary gland for milk production requires not only the adaptation of glucose and lipid metabolism to the lactating state, but also cholesterol metabolism in particular during early lactation (Kessler et al., 2014). In our results the upregulation of *HMGCS1*, *FAXDC2*, *NSDHL* at 1d postpartum [FC = 7.63; 4.69 and; 1.50] seemed to confirm this evidence also in swine mammary gland. In particular *HMGCS1*, which is important for the regulation of cholesterol synthesis (Rikitake et al., 2001), was markedly upregulated during early lactation compared with late pregnancy in the bovine mammary gland (Kessler et al., 2014).

***Utilization of ketone bodies*.** On day -2 and day 1, we detected moderate upregulation of *BDH* (3-hydroxybutyrate dehydrogenase), encoding a protein catalyzing the initial steps of BHBA utilization in mitochondria (Robinson and Williamson, 1980). In humans, cytosolic type BDH2 is involved in the cytosolic utilization of ketone bodies, which can subsequently enter mitochondria and the tricarboxylic acid cycle (Yang et al., 2013). In ruminants, previous studies showed that the mammary gland takes up large amounts of BHBA and concluded that the use of BHBA (as 4-carbon units) by mammary cells is primarily for *de novo* FA synthesis (Palmquist et al., 1969; Bionaz and Loor, 2008a). The moderate upregulation of *BDH2* in the last 2 time comparisons [FC = 1.68; 1.93] suggested that ketone bodies likely are an energy source also for the sow mammary gland.

***Ceramide-synthesis genes in mammary gland*.** There was moderate upregulation of *SGMS1* (sphingomyelin synthase 1) in the prepartum period [FC = 1.42] and higher upregulation of *SGMS2* (sphingomyelin synthase 2) in the postpartum period [FC = 4.37], and we found a

concomitant downregulation of *SGMS1* [FC = -1.35]. Sphingomyelin synthases synthesize sphingomyelin through transfer of the phosphatidyl head group in phosphatidylcholine to the primary hydroxyl group of ceramide. Ceramide, which is involved in cell signaling, cell cycle, and regulation of protein transport from ER to Golgi, is one of the most-studied sphingolipids in nature (Jazwinski and Conzelmann, 2002). Sphingomyelin synthesis from ceramide is considered an important step because sphingomyelin constitutes about 25% of the total phospholipids in dairy products, having highly bioactive properties and is considered to be functional in food (Palmquist, 2006b). The upregulation of *SMPD1* (sphingomyelin phosphodiesterase 1) [FC = 1.67] and the simultaneously downregulation of *CERS1* (Ceramide Synthase 1) [FC = -2.12] at 1d postpartum appears to have a biologic role in the overall process of sphingomyelin metabolism. SMPD1 is involved in the conversion of sphingomyelin to ceramide, whereas CERS1 catalyzes the synthesis of ceramide. Further protein expression and functional studies during the entire lactation should be conducted to clarify the role of sphingolipids with signaling roles and the role of ceramide in swine mammary gland.

The marked upregulation of *CYP4A21* [FC = 568.37], a member of the CYP4A subfamily discovered in pig (Lundell et al., 2001), is noteworthy because the protein possesses taurochenodeoxycholic acid 6α-hydroxylase activity but does not metabolise lauric acid, a common substrate for other CYP4As (Lundell et al., 2001). The function of CYP4A *in vivo* is not well understood but CYP4As are known for hydroxylating of a series of fatty acids, eicosanoids and prostaglandins (PG) (Simpson, 1997; Capdevila et al., 1999; Omura, 1999). The activity of CYP4A21 is still uncharacterized in mammary gland. CYP4A21 is believed to be responsible for formation of hyocholic acid, a bile acid typically found in porcine (Lundell, 2004). Further analysis is required to investigate the role of this gene during the onset of lactation in swine mammary gland.

In summary, our data showed an abrupt increase of all pathways involved in the synthesis of main milk components. Hence, the upregulation at 1d postpartum of *OXTR* (oxytocin receptor) [FC = 3.17] (a G protein-coupled receptor) would help guarantee ejection of these components from the mammary gland (Kimura et al., 1992).

III - 5.4. Transcription factors

The first step of gene expression and the primary step at which gene expression is controlled is transcription. This is accomplished through the recruitment of several transcription factors,

which have the ability to bind to certain target-sequences primarily located in the 5' upstream regulatory region of the genes, and promote or suppress gene transcription according to the stimuli (Laliotis et al., 2010). Considering those transcription regulators in IPA findings that overlap in the last 2 time comparisons (p-value cutoff ≤ 0.01 and activation z-score ≥ ± 2) the results supported the suggestion that SREBP1 and XBP1 are pivotal in the transition from colostrogenesis to lactogenesis in swine mammary gland. They likely act on regulation of lipid synthesis (Anderson et al., 2007) and morphological mammary development (Davis et al., 2016), respectively.

Regulation of lipid biosynthesis. The function of SREBP1 (sterol regulatory element-binding protein 1) is well-established and it is a TF that plays a central role in the regulation of hepatic cholesterol biosynthesis and FA metabolism, particularly the biosynthesis of fat (Desvergne et al., 2006; Goldstein et al., 2006). Our results show that SREBP1 is also important in the mammary gland for cholesterol biosynthesis and this is consistent with real-time PCR measurements that confirmed the upregulation of SREBP1 during the transition from pregnancy to lactation in murine mammary gland (Rudolph et al., 2007). In non-ruminants, SREBP1 resides as an inactive precursor on the endoplasmic reticulum membrane and is transported to the Golgi for proteolytic cleavage (i.e., activation) prior to entering the nucleus where it activates expression of sterol response element (SRE)-containing genes (Bionaz and Loor, 2008a). The transport step to the Golgi is blocked by sterols via the sterol-sensing protein SCAP (SREBP cleavage activating protein), and its expression was modestly upregulated in the last comparison [FC = 1.26]. Release of SREBP1-SCAP is essential for the movement from the ER to the Golgi, acting as gate keeper for movement of inactive SREBP1.

Insulin induced gene (INSIG) 1 and 2 are proteins that interact with SCAP in an oxysterol-dependent and independent fashion (in non-ruminants) and regulate the responsiveness of SREBP1 and 2 processing via SCAP, thus, altering rates of lipogenesis and cholesterogenesis. The precise role of INSIG1, strongly upregulated in our last comparison [FC = 13.22], is controversial. In fact, decreased SREBP activity as a consequence of increased *INSIG1* has been observed in liver when *INSIG1* is overexpressed (Engelking et al., 2004), but upregulation of *INSIG1* was detected during lactation and positively correlated with the ratio of synthesized/imported FA in bovine mammary gland (Bionaz and Loor, 2008a).

Our data support a need of *INSIG1* in controlling the induction of gene expression by SREBP. Therefore, INSIG1 could play a central role in orchestrating lipid metabolism also in swine

mammary tissue during lactation. In this regard, the predicted upregulation of PPARG expression at 1d postpartum (known to be involved in regulation of lipid synthesis in goat and bovine mammary cells (Shi et al., 2016c; Shi et al., 2013; Kadegowda et al., 2009; Bionaz and Loor, 2008a) is noteworthy [PPARGC1B z-score = 2.06]. A potential role of this nuclear receptor in milk fat synthesis was already postulated in particular in bovine mammary gland, where *INSIG1* was demonstrated to be a PPARG responsive gene, suggesting that PPARG in mammary tissue could serve as regulator of SREBP activity (Bionaz and Loor, 2008a). PPARG could represent an important control point of milk fat synthesis, in particular in triacylglycerol synthesis and milk secretion in pig as well as in goat and bovine (Shi et al., 2013; Kadegowda et al., 2009), acting indirectly on SREBP1 protein activity through regulation of the expression of insulin-induced gene 1 (*INSIG1*) and directly on SREBP1.

Regulation of morphological mammary development. Colostrum and milk synthesis occurs in alveolar structures composed of a single layer of MEC encircling a lumen where milk is secreted (Anderson et al., 2007). In order to become fully functional, MEC acquire a number of cellular characteristics during late pregnancy including the development of an elaborate endoplasmic reticulum (ER) system (Akers et al., 1981), which is required for the synthesis of secreted proteins but is also the site where fatty acids are assembled into TAG and phospholipids (Fagone and Jackowski, 2009). The coordination of synthesis and export of products in murine mammary epithelial cells is orchestrated in part by the transcription factor X-box binding protein 1 (XBP1) (Kim et al., 2016), which has multiple functions. Briefly, it promotes ER biogenesis (Sriburi et al., 2007) and is a component of a highly-conserved signaling cascade responsible for restoring homeostasis when the ER is confronted with various stresses, including increased protein synthesis and secretion (Hetz, 2012; Moore and Hollien, 2012). XBP1 is also implicated as a positive regulator of both lipogenesis and VLDL (very low density lipoprotein) secretion in hepatocytes (Lee et al., 2008; Wang et al., 2012a). Recently, in murine, it was shown that XBP1 is required for MEC population expansion during lactation and its ability to develop an elaborate endoplasmic reticulum compartment (Davis et al., 2016). All the above evidence is consistent with the suggestion that XBP1 may be indispensable for morphologically mammary development, colostrum and milk synthesis and secretion during late-pregnancy and the onset of lactation in pig. In particular, focusing on the upregulated genes involved in protein processing in endoplasmic reticulum that were detected in the last two comparisons, the significant upregulation of *PDIA4* [FC = 2.42 and 3.11], *PDIA3* [FC = 1.60

and 2.15], *PDIA6* [FC = 1.49 and 2.05] and *CALR* (calreticulin) [FC = 1.63 and 2.11] is noteworthy. *PDIA4*, *PDIA3* and *PDIA6* are genes that encode for specific members of the disulfide isomerase (PDI) family of endoplasmic reticulum (ER) proteins that catalyze protein folding and thiol-disulfide interchange reactions. Calreticulin is a multifunctional protein that acts as a major Ca(2+)-binding (storage) protein in the lumen of the endoplasmic reticulum. In MEC, the role of ER-resident proteins on the folding and the retention of milk proteins is not well defined. However, calreticulin and PDI have been detected in rat and goat lactating MEC, leading the authors to suggest that these proteins could be involved in the formation of lipid droplets, raising questions about a possible link between the enzymes involved in protein and lipid synthesis (Ghosal et al., 1994).

Other upregulated Transcription Regulators. At 1d postpartum, we detected a marked upregulation of IRF7 [z-score = 4.69], TP53 [z-score = 4.25], NUPR1 [z-score = 4.10] and NFATC2 [z-score = 4.09] together with XBP1 and SREBP1 which had the highest z-score values. IRF7 encodes interferon regulatory factor 7, a member of the interferon regulatory transcription factor (IRF) family, it is a key transcriptional regulator of type I interferon (IFN) dependent immune responses and plays a critical role in the innate immune response against DNA and RNA viruses (Ning et al., 2011). It regulates the transcription of type I IFN genes (*IFN-α* and *IFN-β*) and IFN-stimulated genes (*ISG*), which are markedly upregulated [*ISG15*, FC = 2.95], by binding to an interferon-stimulated response element (ISRE) in their promoters. TP53 (Tumor Protein P53) is a tumor suppressor implicated in several types of human tumors, and it functions both as a gene-specific transcription factor as well as a specific inhibitor of the transcription of certain genes (Shaw, 1996). Its tumor suppressor activity is typically ascribed to its role as a transcription factor regulating expression of genes involved in control of cell cycle, cellular senescence, and apoptosis (Vousden and Prives, 2009) but recently Munne et al. (Munne et al., 2014) suggested and demonstrated a role for TP53 in the epithelial-to-mesenchymal transition (EMT) and differentiation of mammary epithelia. NUPR1 is a nuclear protein transcriptional regulator. Together with other TR, Zhou et al. (Zhou et al., 2014) reported a high expression of NUPR1 during lactation compared with pregnancy. NUPR1 is involved in negative regulation of the cell cycle (Sambasivan et al., 2009), which could explain why cell cycle-related genes are more active in pregnancy. NFATC2 (nuclear factor of activated t-cells 2) is a member of the nuclear factor of activated T cells (NFAT) family. Most of the work on NFAT proteins has been related to immune cell activation and its mediators, such as

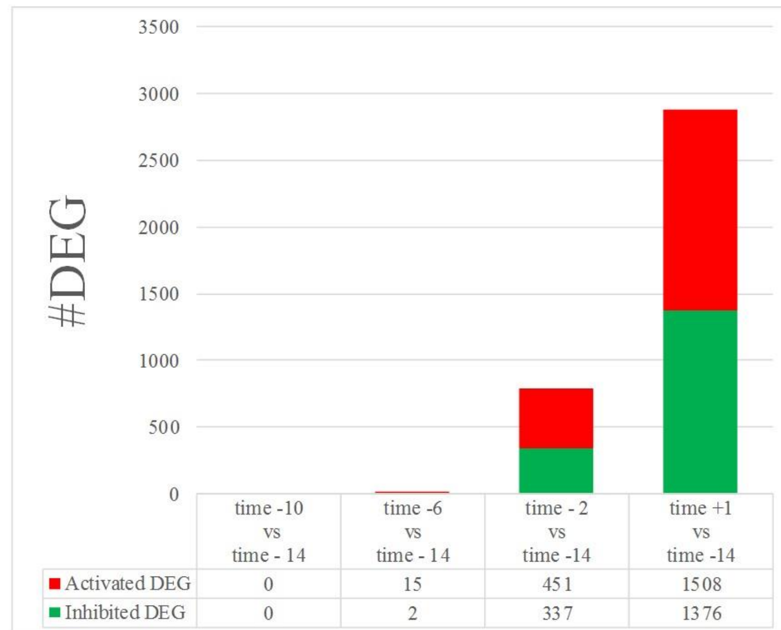
cytokines (Rao et al., 1997). The product of this gene is a DNA-binding protein with a REL-homology region (RHR) and an NFAT-homology region (NHR). This protein is present in the cytosol and only translocated to the nucleus upon T cell receptor (TCR) stimulation, where it becomes a member of the nuclear factors of activated T cells transcription complex. This complex plays a central role in inducing gene transcription during the immune response (Kuklina and Shirshev, 2001).

III - 6. Conclusion

The transcriptome changes greatly between 6 and 2 days prepartum and these changes are highly likely to be involved in coordinating the synthesis of colostrum and main milk components (i.e. protein, fat, lactose and antimicrobial factors) as revealed by influenced pathways. The lipid metabolism pathway changes greatly and some of those adaptations are controlled at least in part via SREBP1 and XBP1, acting on regulation of lipid synthesis and morphological development of the mammary gland. Other transcription regulators including IRF7, TR53, NUPR1 and NFATC2 acting across a wide number of pathways become important at the onset of lactation. Further research will help confirm the functional relevance of the pathways uncovered, and how they influence the transition from colostrum to mature milk during a stage when slight abnormalities may potentially threaten piglet survival. Clearly, milk synthesis requires a complexity of factors beyond transcription of the major proteins involved in the synthesis and secretion of protein, fat, and lactose. Holistically, milk synthesis is the product of complex interactions among several tissues and organs that only an integrative systems-biology approach may help elucidate.

III - 7. Figures and tables

Figure III-1. Total number of DEGs due to time resulting from DE analysis of RNASeq data on swine mammary gland from late pregnancy to farrowing (FDR and p-value ≤ 0.05).



	time -10 vs time - 14	time -6 vs time - 14	time - 2 vs time -14	time +1 vs time -14
Activated DEG	0	15	451	1508
Inhibited DEG	0	2	337	1376

Figure III-2. KEGG main categories resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data on swine mammary gland from late pregnancy to farrowing (FDR and p-value ≤ 0.05). For each time comparison, the columns represent the effect (impact) and flux responses. The blue bars represent the effect value (0 to 150), and the flux columns represent negative (−) and positive (+) flux (−150 to +150) based on the direction of the effect. The negative flux (green bars) indicates a downregulation, while the positive flux (red bars) indicates an upregulation.

Figure III-3. KEGG 'Lipid Metabolism' pathways resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data on swine mammary gland from late pregnancy to farrowing (FDR and p-value ≤ 0.05). For each time comparison, the columns represent the effect (impact) and flux responses. The blue bars represent the effect value (0 to 300), and the flux columns represent negative (−) and positive (+) flux (−300 to +300) based on the direction of the effect. The negative flux (green bars) indicates a downregulation, while the positive flux (red bars) indicates an upregulation.

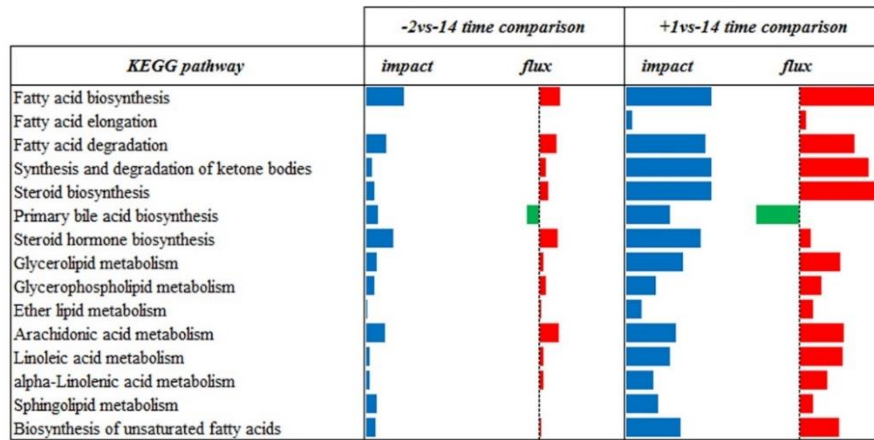


Figure III-4. KEGG ’Endocrine system’ pathways resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data on swine mammary gland from late pregnancy to farrowing (FDR and p-value ≤ 0.05). For each time comparison, the columns represent the effect (impact) and flux responses. The blue bars represent the effect value (0 to 200), and the flux columns represent negative (−) and positive (+) flux (−200 to +200) based on the direction of the effect. The negative flux (green bars) indicates a downregulation, while the positive flux (red bars) indicates an upregulation.

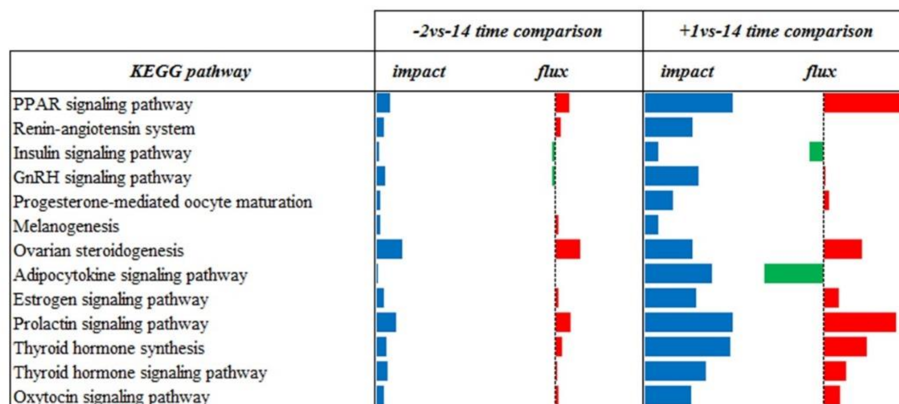


Figure III-5. Top 10 upregulated KEGG pathways in -2d vs -14 d comparison resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data on swine mammary gland from late pregnancy to farrowing (FDR and p-value ≤ 0.05). The columns represent the effect (impact) and flux responses. The blue bars represent the effect value (0 to 150) and red the bars represent the flux (the direction of the effect).

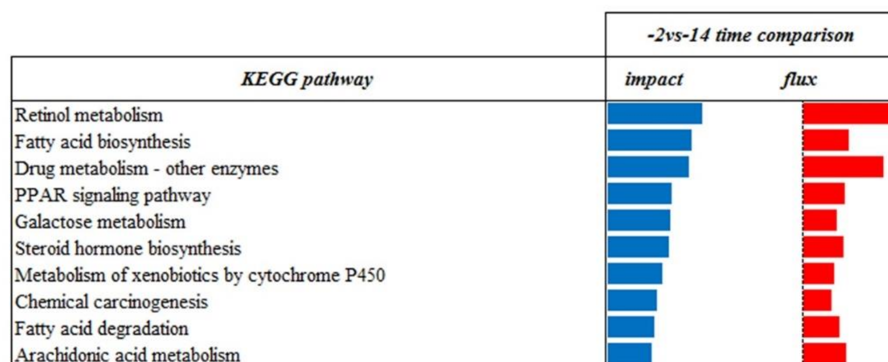


Figure III-6. Top 10 upregulated KEGG pathways in +1d vs -14 d comparison resulting from the DIA analysis on DEGs obtained by DE analysis of RNASeq data on swine mammary gland from late pregnancy to farrowing (FDR and p-value ≤ 0.05). The columns represent the effect (impact) and flux responses. The blue bars represent the effect value (0 to 150) and red the bars represent the flux (the direction of the effect).

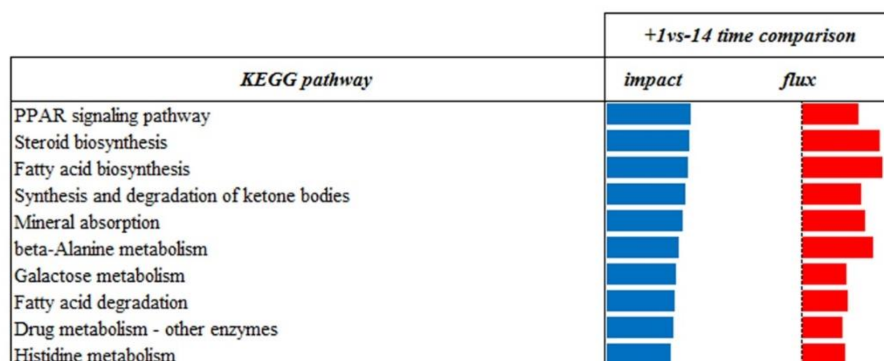


Table III-1. Summary of RNA extraction and quality check for all the samples.

Sample ID	Time days to parturition	Mass (mg)	RNA Concentration (ng/ul)	RIN
1	-14	9.3	473.53	6.7
2	-14	17.7	757.92	6.7
3	-14	12.2	85.72	6.3
1	-10	19.6	977.97	8.6
2	-10	13.8	1007.79	7.7
3	-10	15.2	409.14	6.7
1	-6	26.6	979.99	8.2
2	-6	27.3	1750.72	8.2
3	-6	29.3	852.2	8.1
1	-2	18.4	829.75	8.7
2	-2	15.7	784.09	8.7
3	-2	28.2	830.40	9.3
1	1	17.0	1056.95	9.4
2	1	23.6	1207.2	9.0
3	1	31.7	809.31	9.1

Table III-2. Quantitative real time PCR (qPCR) validation of sequencing (Seq) results. For each gene the overall false discovery rate (FDR) was reported, together with the comparison specific fold-change (FC) and P-value, generated applying the same statistical model to either qPCR or Seq data.

		-14 vs -2 time comparison		-14 vs +1 time comparison	
Target	**FDR**	**FC**	***P*-value**	**FC**	***P*-value**
CSN2					
qPCR	<.0001	51.9	<.0001	254.8	<.0001
Seq	0.001	31.0	0.003	33.5	<.0001
LALBA					
qPCR	<.0001	156.1	<.0001	8321.9	<.0001
Seq	0.003	128.5	0.03	1275.5	0.001
LTF					
qPCR	<.0001	1.5	0.02	3.4	<.0001
Seq	0.001	1.7	0.10	3.9	<.0001
PAEP					
qPCR	0.0001	2.2	0.0005	3.4	<.0001
Seq	0.004	1.8	0.01	1.9	0.001

Table III-3. RNA sequencing and alignment for all the samples.

Sample ID	Time days to parturition	Total Reads	Total Mapped Reads	Percent Mapped	Uniquely mapped Reads	Percent uniquely mapped	Reads mapped to annotated exons
1	-14	30977882	28530182	92.1%	26049171	91.3%	18924790
2	-14	31749008	29241577	92.1%	26662181	91.2%	19292238
3	-14	28119604	25931870	92.2%	23632716	91.1%	16862026
1	-10	31167527	28730778	92.2%	26202356	91.2%	19238556
2	-10	27627300	25520346	92.4%	23356330	91.5%	17280281
3	-10	27913828	25648864	91.9%	23372282	91.1%	17048788
1	-6	28389469	26134316	92.1%	23946131	91.6%	17726719
2	-6	28896592	26677836	92.3%	24500842	91.8%	18261982
3	-6	29503599	27221174	92.3%	24889449	91.4%	18274050
1	-2	29505106	27138019	92.0%	25011949	92.2%	19222359
2	-2	30207980	27794713	92.0%	25783333	92.8%	19952918
3	-2	21095245	19535089	92.6%	17944543	91.9%	13459054
1	1	15360190	13983211	91.0%	13130206	93.9%	10782343
2	1	17917675	16454059	91.8%	15185453	92.3%	11416874
3	1	24317858	22265447	91.6%	20539061	92.2%	15744454

Table III-4. Top 10 upregulated genes in both and specific time comparisons (FDR and p-value ≤ 0.05).

Status	**+1vs-14 time comparison**	**-2vs-14 time comparison**	**both time comparison**
upregulated	*WAP, SAA2, LBP, CP*	*CYP1A1, LOC100524679, CSN2, ACSL6*	*HP, LALBA, CSN1S2, BTN1A1, BTN1A1-like LOC100522145*

Table III-5. Top ten upregulated genes in -2vs-14 comparison.

Gene symbol	**Entrez gene ID**	**FDR**	**FC**	**P-value**
CSN1S2	445515	0.000	629.572	0.002
LALBA	397647	0.003	128.511	0.032
taurochenodeoxycholic 6 alpha-hydroxylase-like	100522145	0.002	95.634	0.019
butyrophilin subfamily 1 member A1	100153328	0.007	90.488	0.023
butyrophilin subfamily 1 member A1-like	100626139	0.007	90.488	0.023
CYP1A1	403103	0.005	59.108	0.011
ovostatin homolog 2-like	100524679	0.006	59.017	0.015
CSN2	404088	0.001	31.014	0.003
HP	397061	0.000	29.272	0.001
ACSL6	100522126	0.012	26.040	0.031

Table III-6. Summary of upregulated genes of most recurrent subcategory in both and specific comparisons.

KEGG Category	Status	Genes in +1vs-14 comparison	Genes in -2vs-14 comparison	Genes in both comparison
Lipid metabolism	upregulated	ALDH2, HMGCS1, HSD11B1, SCD, ACSL3, FAXDC2, SGMS2, SC5D, 403334, ACACB, SQLE, MSMO1, CYP2J2, CERS4, CYP2J34, LCLAT1, DHCR24, 100233182, PLPP3, FADS2, 100170845, 100517533, ACAT2, MGLL, CYP2D25, TM7SF2, ACADM, ARSA, PNPLA2, GLA, SMPD1, PTGS1, LPCAT3, GPCPD1, PAFAH2, ACADS, NSDHL, PLA2G12A, KDSR, ECI2, GBA2, COMT, 100515577	CYP1A1, 100157065, GPAM, GPD2, SGMS1, CDIPT, DGAT1	GGT1, 397097, FADS1, AGPAT1, NEU1, GPAT4, CEPT1, HSD17B7, CDS2, PLA2G16, 100522126, 100522145, 100522692, GPAT3, 100625138, 100625332, 100738292
Endocrine system	upregulated	CYP1A1, PYGB	396835, CD14, FXYD2, FOS, HK2, RCAN1, SH2B2, PFKFB2, STAT3, ACACB, PHKG1, ITGB3, HSPA5, OXTR, CYP2J2, PLN, PLCD3, SEC11C, CYP2J34, HSP90B1, MYL9, CPEB4, 100514493, ITGA11, RRAS, CREB3L2, CPEB3, CFL2, NFATC2, CTSB, SEC61G, SRP54, SPCS3, PRKCI, ITGA2, PLCD1, JUN, DIAPH1, SEC63, PRKAB2, SP1, BCAR1, B2M, 106504143, PRKAB1, PIKFYVE, KAT2B, CREB3, PRKCD, STAT1, RYR2,	CTSV, KCNJ2, NOS3, PCK2, CSN2, FOXO3, MTOR, NCOA2, SPCS1, SRC, SEC61A1, HSD17B7, PDIA3, MAPK14, CALR, SEC61B, RPTOR, SRPRA, CREB3L1, CANX, VAV1, 100522176, PDIA4, GNA13, 100523015, 100523202, EEF2K, MRAS

Table III-7. Summary Transcription Regulators (TRs) in both and specific comparisons.

Status	**TRs in +1vs-14 comparison**	**TRs in -2vs-14 comparison**	**TRs in both comparison**
activated	ATF4, CDKN2A, CEBPA, CREB1, CREM, E2F6, ECSIT, EPAS1, FOXO3, GATA1, HIF1A, ID3, IRF1, IRF3, IRF5, IRF7, KDM5B, MEF2D, MXI1, NFATC2, NFKB1, NFKBIA, NUPR1, PDX1, PPARGC1B, RB1, RBL1, RELA, SMARCA4, SMARCB1, SREBF2, STAT1, STAT2, TCF3, TCF7L2, TOB1, TP53	ATF6	SREBF1 XBP1

III - 8. References

Abu-Elheiga, L., W.R. Brinkley, L. Zhong, S.S. Chirala, G. Woldegiorgis, and S.J. Wakil. 2000. The subcellular localization of acetyl-CoA carboxylase 2. Proc. Natl. Acad. Sci. U. S. A. 97:1444–1449.

Adiconis, X., D. Borges-Rivera, R. Satija, D.S. DeLuca, M.A. Busby, A.M. Berlin, A. Sivachenko, D.A. Thompson, A. Wysoker, T. Fennell, A. Gnirke, N. Pochet, A. Regev, and J.Z. Levin. 2013. Comprehensive comparative analysis of RNA sequencing methods for degraded or low input samples. Nat. Methods 10:623–629. doi:10.1038/nmeth.2483.

Akers, R.M., D.E. Bauman, G.T. Goodman, A.V. Capuco, and H.A. Tucker. 1981. Prolactin regulation of cytological differentiation of mammary epithelial cells in periparturient cows. Endocrinology 109:31–40. doi:10.1210/endo-109-1-31.

Alain, K., N.A. Karrow, C. Thibault, J. St-Pierre, M. Lessard, and N. Bissonnette. 2009. Osteopontin: an early innate immune marker of Escherichia coli mastitis harbors genetic polymorphisms with possible links with resistance to mastitis. BMC Genomics 10:444. doi:10.1186/1471-2164-10-444.

Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106. doi:10.1186/gb-2010-11-10-r106.

Anders, S., D.J. McCarthy, Y. Chen, M. Okoniewski, G.K. Smyth, W. Huber, and M.D. Robinson. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat. Protoc. 8:1765–1786. doi:10.1038/nprot.2013.099.

Anderson, S.M., M.C. Rudolph, J.L. McManaman, and M.C. Neville. 2007. Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis!. Breast Cancer Res. 9:204. doi:10.1186/bcr1653.

Balzani, A., H.J. Cordell, E. Sutcliffe, and S.A. Edwards. 2016. Heritability of udder morphology and colostrum quality traits in swine. J. Anim. Sci. 94:3636–3644. doi:10.2527/jas.2016-0458.

Barrington, G.M., T.E. Besser, C.C. Gay, W.C. Davis, J.J. Reeves, T.B. McFadden, and R.M. Akers. 1999. Regulation of the immunoglobulin G1 receptor: effect of prolactin on in vivo expression of the bovine mammary immunoglobulin G1 receptor. J. Endocrinol. 163:25–31.

Bentley, D.R., S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, J.M. Boutell, J. Bryant, R.J. Carter, R.K. Cheetham, A.J. Cox, D.J. Ellis, M.R. Flatbush, N.A. Gormley, S.J. Humphray, L.J. Irving, M.S. Karbelashvili, S.M. Kirk, H. Li, X. Liu, K.S. Maisinger, L.J. Murray, B. Obradovic, T. Ost, M.L. Parkinson, M.R. Pratt, I.M.J. Rasolonjatovo, M.T. Reed, R. Rigatti, C. Rodighiero, M.T. Ross, A. Sabot, S.V. Sankar, A. Scally, G.P. Schroth, M.E. Smith, V.P. Smith, A. Spiridou, P.E. Torrance, S.S. Tzonev, E.H. Vermaas, K. Walter, X. Wu, L. Zhang, M.D. Alam, C. Anastasi, I.C. Aniebo, D.M.D. Bailey, I.R. Bancarz, S. Banerjee, S.G. Barbour, P.A. Baybayan, V.A. Benoit, K.F. Benson, C. Bevis, P.J. Black, A. Boodhun, J.S. Brennan, J.A. Bridgham, R.C. Brown, A.A. Brown, D.H. Buermann, A.A. Bundu, J.C. Burrows, N.P. Carter, N. Castillo, M.C.E. Catenazzi, S. Chang, R.N. Cooley, N.R. Crake, O.O. Dada, K.D. Diakoumakos, B. Dominguez-Fernandez, D.J. Earnshaw, U.C. Egbujor, D.W. Elmore, S.S. Etchin, M.R. Ewan, M. Fedurco, L.J. Fraser, K.V.F. Fajardo, W.S. Furey, D. George, K.J. Gietzen, C.P. Goddard, G.S. Golda, P.A. Granieri, D.E. Green, D.L. Gustafson, N.F. Hansen, K. Harnish, C.D. Haudenschild, N.I. Heyer, M.M. Hims, J.T. Ho, A.M. Horgan, K. Hoschler, S. Hurwitz, D.V. Ivanov, M.Q. Johnson, T. James, T.A.H. Jones, G.-D. Kang, T.H. Kerelska, A.D. Kersey, I. Khrebtukova, A.P. Kindwall, Z. Kingsbury, P.I. Kokko-Gonzales, A. Kumar, M.A. Laurent, C.T. Lawley, S.E. Lee, X. Lee, A.K. Liao, J.A. Loch, M. Lok, S. Luo, R.M. Mammen, J.W. Martin, P.G. McCauley, P. McNitt, P. Mehta, K.W. Moon, J.W. Mullens, T. Newington, Z. Ning, B.L. Ng, S.M. Novo, M.J. O'Neill, M.A. Osborne, A. Osnowski, O. Ostadan, L.L. Paraschos, L. Pickering, A.C. Pike, A.C. Pike, D.C. Pinkard, D.P. Pliskin, J. Podhasky, V.J. Quijano, C. Raczy, V.H. Rae, S.R. Rawlings, A.C. Rodriguez, P.M. Roe, J. Rogers, M.C. Rogert Bacigalupo, N. Romanov, A. Romieu, R.K. Roth, N.J. Rourke, S.T. Ruediger, E. Rusman, R.M. Sanches-Kuiper, M.R. Schenker, J.M. Seoane, R.J. Shaw, M.K. Shiver, S.W.

Short, N.L. Sizto, J.P. Sluis, M.A. Smith, J.E.S. Sohna, E.J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C.L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovsky, S.M. Virk, S. Wakelin, G.C. Walcott, J. Wang, G.J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J.C. Mullikin, M.E. Hurles, N.J. McCooke, J.S. West, F.L. Oaks, P.L. Lundberg, D. Klenerman, R. Durbin, and A.J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59. doi:10.1038/nature07517.

Bionaz, M., and J.J. Loor. 2008a. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366. doi:10.1186/1471-2164-9-366.

Bionaz, M., and J.J. Loor. 2008b. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 138:1019–1024.

Bionaz, M., W. Hurley, and J. Loor. 2012a. Milk protein synthesis in the lactating mammary gland: insights from transcriptomics analyses. InTech 11:10.5772/46054. doi:10.5772/46054.

Bionaz, M., K. Periasamy, S.L. Rodriguez-Zas, W.L. Hurley, and J.J. Loor. 2012b. A novel dynamic impact approach (DIA) for functional analysis of time-course omics studies: validation using the bovine mammary transcriptome. PLos One 7:e32455. doi:10.1371/journal.pone.0032455.

Bjornvad, C.R., T. Thymann, A.Z. Budek, D.H. Nielsen, C. Mølgaard, K.F. Michaelsen, and P.T. Sangild. 2007. Gastrointestinal and body growth in colostrum-deprived piglets in response to whey, casein or soy protein diets. Livest. Sci. 109:30–33. doi:10.1016/j.livsci.2007.01.035.

Bourne, F.J., and J. Curtis. 1973. The transfer of immunoglobulins IgG, IgA and IgM from serum to colostrum and milk in the sow. Immunology 24:157.

Bullard, J.H., E. Purdom, K.D. Hansen, and S. Dudoit. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11:94. doi:10.1186/1471-2105-11-94.

Bussmann, L.E., I.M. Bussmann, and E.H. Charreau. 1996. Role of receptors for epidermal growth factor and insulin-like growth factors I and II in the differentiation of rat mammary glands from lactogenesis I to lactogenesis II. J. Reprod. Fertil. 107:307–314. doi:10.1530/jrf.0.1070307.

Capdevila, J.H., V. Holla, C. Helvig, and J.R. Falck. 1999. Microsomal cytochrome P450 and eicosanoid metabolism. Mol. Aspects Med. 20:42–55, 56–137.

Cerveza, P.J., F. Mehrbod, S.J. Cotton, N. Lomeli, M.C. Linder, E.G. Fonda, and S.J. Wickler. 2000. Milk ceruloplasmin and its expression by mammary gland and liver in pigs. Arch. Biochem. Biophys. 373:451–461. doi:10.1006/abbi.1999.1572.

Chandan, R.C., R.M. Parry, and K.M. Shahani. 1968. Lysozyme, lipase, and ribonuclease in milk of various species. J. Dairy Sci. 51:606–607. doi:10.3168/jds.S0022-0302(68)87036-5.

Coleman, R.A., and D.P. Lee. 2004. Enzymes of triacylglycerol synthesis and their regulation. Prog. Lipid Res. 43:134–176.

Cook, N.R., and J.H. Ware. 1983. Design and Analysis Methods for Longitudinal Research. Annu. Rev. Public Health 4:1–23. doi:10.1146/annurev.pu.04.050183.000245.

Costa-Silva, J., D. Domingues, and F.M. Lopes. 2017. RNA-Seq differential expression analysis: An extended review and a software tool. PLOS ONE 12:e0190152. doi:10.1371/journal.pone.0190152.

Crisà, A., F. Ferrè, G. Chillemi, and B. Moioli. 2016. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. BMC Vet. Res. 12:264. doi:10.1186/s12917-016-0881-7.

Csaki, L.S., J.R. Dwyer, X. Li, M.H.K. Nguyen, J. Dewald, D.N. Brindley, A.J. Lusis, Y. Yoshinaga, P. de Jong, L. Fong, S.G. Young, and K. Reue. 2014. Lipin-1 and lipin-3 together determine adiposity in vivo. Mol. Metab. 3:145–154. doi:10.1016/j.molmet.2013.11.008.

Davis, K.R., S.L. Giesy, Q. Long, C.S. Krumm, K.J. Harvatine, and Y.R. Boisclair. 2016. XBP1 regulates the biosynthetic capacity of the mammary gland during lactation by controlling epithelial expansion and endoplasmic reticulum formation. Endocrinology 157:417–428. doi:10.1210/en.2015-1676.

Desvergne, B., L. Michalik, and W. Wahli. 2006. Transcriptional regulation of metabolism. Physiol. Rev. 86:465–514. doi:10.1152/physrev.00025.2005.

Devillers, N., C. Farmer, A.-M. Mounier, J. Le Dividich, and A. Prunier. 2004a. Hormones, IgG and lactose changes around parturition in plasma, and colostrum or saliva of multiparous sows. Reprod. Nutr. Dev. 44:381–396.

Devillers, N., J. van Milgen, A. Prunier, and J.L. Dividich. 2004b. Estimation of colostrum intake in the neonatal pig. Anim. Sci. 78:305–313. doi:10.1017/S1357729800054096.

Dividich, J.L., J.A. Rooke, and P. Herpin. 2005. Nutritional and immunological importance of colostrum for the new-born pig. J. Agric. Sci. 143:469–485. doi:10.1017/S0021859605005642.

Doege, H., and A. Stahl. 2006. Protein-mediated fatty acid uptake: novel insights from in vivo models. Physiol. Bethesda 21:259–268. doi:10.1152/physiol.00014.2006.

Dudemaine, P.L., C. Thibault, K. Alain, and N. Bissonnette. 2014. Genetic variations in the SPP1 promoter affect gene expression and the level of osteopontin secretion into bovine milk. Anim. Genet. 45:629–640. doi:10.1111/age.12176.

Dwyer, J.R., J. Donkor, P. Zhang, L.S. Csaki, L. Vergnes, J.M. Lee, J. Dewald, D.N. Brindley, E. Atti, S. Tetradis, Y. Yoshinaga, P.J. De Jong, L.G. Fong, S.G. Young, and K. Reue. 2012. Mouse lipin-1 and lipin-2 cooperate to maintain glycerolipid homeostasis in liver and aging cerebellum. Proc. Natl. Acad. Sci. U. S. A. 109:E2486-2495. doi:10.1073/pnas.1205221109.

Engelking, L.J., H. Kuriyama, R.E. Hammer, J.D. Horton, M.S. Brown, J.L. Goldstein, and G. Liang. 2004. Overexpression of Insig-1 in the livers of transgenic mice inhibits SREBP

processing and reduces insulin-stimulated lipogenesis. J. Clin. Invest. 113:1168–1175. doi:10.1172/JCI20978.

Fagone, P., and S. Jackowski. 2009. Membrane phospholipid synthesis and endoplasmic reticulum function. J. Lipid Res. 50 Suppl:S311-316. doi:10.1194/jlr.R800049-JLR200.

Farmer, C., D. Petitclerc, M.T. Sorensen, M. Vignola, and J.Y. Dourmad. 2004. Impacts of dietary protein level and feed restriction during prepuberty on mammogenesis in gilts. J. Anim. Sci. 82:2343–2351. doi:10.2527/2004.8282343x.

Farmer, C. 2006. Colostrum production in swine: from the mammary glands to the piglets.. CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour. 1. doi:10.1079/PAVSNNR20061003.

Farmer, C. 2012. Review: Mammary development in swine: effects of hormonal status, nutrition and management. Can. J. Anim. Sci. 93:1–7. doi:10.4141/cjas2012-066.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk--sixth revision. J. Dairy Sci. 87:1641–1674. doi:10.3168/jds.S0022-0302(04)73319-6.

Filipp, D., K. Alizadeh-Khiavi, C. Richardson, A. Palma, N. Paredes, O. Takeuchi, S. Akira, and M. Julius. 2001. Soluble CD14 enriched in colostrum and milk induces B cell growth and differentiation. Proc. Natl. Acad. Sci. U. S. A. 98:603. doi:10.1073/pnas.98.2.603.

Fontanesi, L., G. Galimberti, D.G. Calò, M. Colombo, A. Astolfi, S. Formica, and V. Russo. 2011. Microarray gene expression analysis of porcine skeletal muscle sampled at several post mortem time points. Meat Sci. 88:604–609. doi:10.1016/j.meatsci.2011.02.001.
Fujino, T., M.J. Kang, H. Suzuki, H. Iijima, and T. Yamamoto. 1996. Molecular characterization and expression of rat acyl-CoA synthetase 3. J. Biol. Chem. 271:16748–16752.

Gao, H., H. Hu, N. Zheng, and J. Wang. 2015. Leucine and histidine independently regulate milk protein synthesis in bovine mammary epithelial cells via mTOR signaling pathway. J. Zhejiang Univ. Sci. B 16:560–572. doi:10.1631/jzus.B1400337.

Gao, Y., X. Lin, K. Shi, Z. Yan, and Z. Wang. 2013. Bovine mammary gene expression profiling during the onset of lactation. PLoS One 8:e70393. doi:10.1371/journal.pone.0070393.

Ghosal, D., N.W. Shappell, and T.W. Keenan. 1994. Endoplasmic reticulum lumenal proteins of rat mammary gland. Potential involvement in lipid droplet assembly during lactation. Biochim. Biophys. Acta 1200:175–181.

Goldman, A.S. 1977. Human milk, leukocytes, and immunity. J. Pediatr. 90:167–168.

Goldstein, J.L., R.A. DeBose-Boyd, and M.S. Brown. 2006. Protein sensors for membrane sterols. Cell 124:35–46. doi:10.1016/j.cell.2005.12.022.

Gonzalez-Baró, M.R., T.M. Lewin, and R.A. Coleman. 2007. Regulation of Triglyceride Metabolism II. Function of mitochondrial GPAT1 in the regulation of triacylglycerol biosynthesis and insulin action. Am. J. Physiol. Gastrointest. Liver Physiol. 292:G1195. doi:10.1152/ajpgi.00553.2006.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.-J. Kim, A. Kvasz, M. Mni, P. Simon, J.-M. Frère, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. U. S. A. 101:2398–2403.

Han, L.Q., H.J. Li, Y.Y. Wang, H.S. Zhu, L.F. Wang, Y.J. Guo, W.F. Lu, Y.L. Wang, and G.Y. Yang. 2010. mRNA abundance and expression of SLC27A, ACC, SCD, FADS, LPIN, INSIG, and PPARGC1 gene isoforms in mouse mammary glands during the lactation cycle. Genet. Mol. Res. GMR 9:1250–1257. doi:10.4238/vol9-2gmr814.

Han, Y., S. Gao, K. Muegge, W. Zhang, and B. Zhou. 2015. Advanced Applications of RNA Sequencing and Challenges. Bioinforma. Biol. Insights 9:29–46. doi:10.4137/BBI.S28991.

Hardcastle, T.J., and K.A. Kelly. 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422. doi:10.1186/1471-2105-11-422.

Hartmann, P.E., N.A. Smith, M.J. Thompson, C.M. Wakeford, and P.G. Arthur. 1997. The lactation cycle in the sow: physiological and management contradictions. Livest. Prod. Sci. 50:75–87. doi:10.1016/S0301-6226(97)00076-6.

Harvatine, K.J., Y.R. Boisclair, and D.E. Bauman. 2009. Recent advances in the regulation of milk fat synthesis. Animal 3:40–54. doi:10.1017/S1751731108003133.

He, Y., N.T. Lawlor, and D.S. Newburg. 2016. Human milk components modulate toll-like receptor–mediated inflammation. Adv. Nutr. 7:102. doi:10.3945/an.115.010090.

Hetz, C. 2012. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. Nat. Rev. Mol. Cell Biol. 13:89–102. doi:10.1038/nrm3270.

Hiss-Pesch, S., F. Daniel, S. Dunkelberg-Denk, M. Mielenz, and H. Sauerwein. 2011. Transfer of maternal haptoglobin to suckling piglets. Vet. Immunol. Immunopathol. 144:104–110. doi:10.1016/j.vetimm.2011.07.015.

Hrdlickova, R., M. Toloue, and B. Tian. 2017. RNA-Seq methods for transcriptome analysis. Wiley Interdiscip. Rev. RNA 8. doi:10.1002/wrna.1364.

Hurley, W.L., and P.K. Theil. 2011. Perspectives on immunoglobulins in colostrum and milk. Nutrients 3:442–474. doi:10.3390/nu3040442.

Hurley, W.L., and P.K. Theil. 2013. Immunoglobulins in mammary secretions. Adv. Dairy Chem. 275–294. doi:10.1007/978-1-4614-4714-6_9.

Hurley, W.L. 2014. Composition of sow colostrum and milk. Wageningen Academic Publishers.

Janeway CA Jr, P. Travers, M. Walport, and M.J. Shlomchik. 2001. The complement system and innate immunity. N. Y. Garland Sci. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27100/.

Jazwinski, S.M., and A. Conzelmann. 2002. LAG1 puts the focus on ceramide signaling. Int. J. Biochem. Cell Biol. 34:1491–1495.

Jean, K.-B., and S.-H. Chiang. 1999. Increased survival of neonatal pigs by supplementing medium-chain triglycerides in late-gestating sow diets. Anim. Feed Sci. Technol. 76:241–250. doi:10.1016/S0377-8401(98)00224-7.

Ji, F., W.L. Hurley, and S.W. Kim. 2006. Characterization of mammary gland development in pregnant gilts. J. Anim. Sci. 84:579–587.

Kadegowda, A.K.G., M. Bionaz, L.S. Piperova, R.A. Erdman, and J.J. Loor. 2009. Peroxisome proliferator-activated receptor-γ activation and long-chain fatty acids alter lipogenic gene networks in bovine mammary epithelial cells to various extents. J. Dairy Sci. 92:4276–4289. doi:10.3168/jds.2008-1932.

Kaselonis, G.L., E.R. McCabe, and S.M. Gray. 1999. Expression of hexokinase 1 and hexokinase 2 in mammary tissue of nonlactating and lactating rats: evaluation by RT-PCR. Mol. Genet. Metab. 68:371–374. doi:10.1006/mgme.1999.2923.

Keenan, T.W., and I.H. Mather. 2006. Intracellular origin of milk fat globules and the nature of the milk fat globule membrane. Adv. Dairy Chem. 2 Lipids:137–171. doi:10.1007/0-387-28813-9_4.

Kensinger, R., R. Collier, F. Bazer, C. Ducsay, and H. Becker. 1982. Nucleic acid, metabolic and histological changes in gilt mammary tissue during pregnancy and lactogenesis. J. Anim. Sci. 54:1297–1308.

Kessler, E.C., J.J. Gross, R.M. Bruckmaier, and C. Albrecht. 2014. Cholesterol metabolism, transport, and hepatic regulation in dairy cows during transition and early lactation. J. Dairy Sci. 97:5481–5490. doi:10.3168/jds.2014-7926.

Kim, S.W., R.A. Easter, and W.L. Hurley. 2001. The regression of unsuckled mammary glands during lactation in sows: the influence of lactation stage, dietary nutrients, and litter size. J. Anim. Sci. 79:2659–2668.

Kim, R.S., D. Hasegawa, N. Goossens, T. Tsuchida, V. Athwal, X. Sun, C.L. Robinson, D. Bhattacharya, H.-I. Chou, D.Y. Zhang, B.C. Fuchs, Y. Lee, Y. Hoshida, and S.L. Friedman. 2016. The XBP1 arm of the unfolded protein response induces fibrogenic activity in hepatic stellate cells through autophagy. Sci. Rep. 6:39342. doi:10.1038/srep39342.

Kimura, T., O. Tanizawa, K. Mori, M.J. Brownstein, and H. Okayama. 1992. Structure and expression of a human oxytocin receptor. Nature 356:526–529. doi:10.1038/356526a0.

Kinsella, J.E. 1972. Stearyl CoA as a precursor of oleic acid and glycerolipids in mammary microsomes from lactating bovine: possible regulatory step in milk triglyceride synthesis. Lipids 7:349–355. doi:10.1007/BF02532654.

Krakowski, L., J. Krzyzanowski, Z. Wrona, K. Kostro, and A.K. Siwicki. 2002. The influence of nonspecific immunostimulation of pregnant sows on the immunological value of colostrum. Vet. Immunol. Immunopathol. 87:89–95.

Krogh, U., T.S. Bruun, J. Poulsen, and P.K. Theil. 2017a. Impact of fat source and dietary fibers on feed intake, plasma metabolites, litter gain and the yield and composition of milk in sows. Animal 11:975–983. doi:10.1017/S1751731116002585.

Krogh, U., N. Oksbjerg, A.C. Storm, T. Feyera, and P.K. Theil. 2017b. Mammary nutrient uptake in multiparous sows fed supplementary arginine during gestation and lactation. J. Anim. Sci. 95:2517–2532. doi:10.2527/jas.2016.1291.

Kruse, P.E. 1983. The importance of colostral immunoglobulins and their absorption from the intestine of the newborn animals. Ann. Rech. Veterinaires Ann. Vet. Res. 14:349–353.

Kuklina, E.M., and S.V. Shirshev. 2001. Role of transcription factor NFAT in the immune response. Biochem. Biokhimiia 66:467–475.

Kukurba, K.R., and S.B. Montgomery. 2015. RNA Sequencing and Analysis. Cold Spring Harb. Protoc. 2015:951–969. doi:10.1101/pdb.top084970.

Lai, I.-H., J.H. Tsao, Y.P. Lu, J.W. Lee, X. Zhao, F.L. Chien, and S.J.T. Mao. 2009. Neutrophils as one of the major haptoglobin sources in mastitis affected milk. Vet. Res. 40:17. doi:10.1051/vetres:2008055.

Laliotis, G.P., I. Bizelis, and E. Rogdakis. 2010. Comparative approach of the de novo fatty acid synthesis (lipogenesis) between ruminant and non ruminant mammalian species: from biochemical level to the main regulatory lipogenic genes. Curr. Genomics 11:168. doi:10.2174/138920210791110960.

Larson, B.L., H.L. Heary, and J.E. Devery. 1980. Immunoglobulin production and transport by the mammary gland. J. Dairy Sci. 63:665–671. doi:10.3168/jds.S0022-0302(80)82988-2.

Law, C.W., Y. Chen, W. Shi, and G.K. Smyth. 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15:R29. doi:10.1186/gb-2014-15-2-r29.

LeBouder, E., J.E. Rey-Nores, N.K. Rushmere, M. Grigorov, S.D. Lawn, M. Affolter, G.E. Griffin, P. Ferrara, E.J. Schiffrin, B.P. Morgan, and M.O. Labéta. 2003. Soluble forms of Toll-

like receptor (TLR)2 capable of modulating TLR2 signaling are present in human plasma and breast milk. J. Immunol. 171:6680–6689. doi:10.4049/jimmunol.171.12.6680.

Lee, A.-H., E.F. Scapa, D.E. Cohen, and L.H. Glimcher. 2008. Regulation of hepatic lipogenesis by the transcription factor XBP1. Science 320:1492–1496. doi:10.1126/science.1158042.

Lee, J.-W., M.J. Paape, T.H. Elsasser, and X. Zhao. 2003. Elevated milk soluble CD14 in bovine mammary glands challenged with Escherichia coli lipopolysaccharide. J. Dairy Sci. 86:2382–2389. doi:10.3168/jds.S0022-0302(03)73832-6.

Leng, N., J.A. Dawson, J.A. Thomson, V. Ruotti, A.I. Rissman, B.M.G. Smits, J.D. Haag, M.N. Gould, R.M. Stewart, and C. Kendziorski. 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinforma. Oxf. Engl. 29:1035–1043. doi:10.1093/bioinformatics/btt087.

Leung, D.W. 2001. The structure and functions of human lysophosphatidic acid acyltransferases. Front. Biosci. 6:D944-953.

Li, J., and R. Tibshirani. 2013. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. 22:519–536. doi:10.1177/0962280211428386.

Li, S., S.W. Tighe, C.M. Nicolet, D. Grove, S. Levy, W. Farmerie, A. Viale, C. Wright, P.A. Schweitzer, Y. Gao, D. Kim, J. Boland, B. Hicks, R. Kim, S. Chhangawala, N. Jafari, N. Raghavachari, J. Gandara, N. Garcia-Reyero, C. Hendrickson, D. Roberson, J. Rosenfeld, T. Smith, J.G. Underwood, M. Wang, P. Zumbo, D.A. Baldwin, G.S. Grills, and C.E. Mason. 2014. Multi-platform and cross-methodological reproducibility of transcriptome profiling by RNA-seq in the ABRF Next-Generation Sequencing Study. Nat. Biotechnol. 32:915–925. doi:10.1038/nbt.2972.

Lin, C.Y., S. Abraham, and S. Smith. 1976. Acyl specificity in triglyceride synthesis by lactating rat mammary gland. J. Lipid Res. 17:647–656.

Loisel, F., C. Farmer, P. Ramaekers, and H. Quesnel. 2013. Effects of high fiber intake during late pregnancy on sow physiology, colostrum production, and piglet performance. J. Anim. Sci. 91:5269–5279. doi:10.2527/jas.2013-6526.

Lu, W., Z. Zhao, Y. Zhao, S. Yu, Y. Zhao, B. Fan, I. Kacskovics, L. Hammarström, and N. Li. 2007. Over-expression of the bovine FcRn in the mammary gland results in increased IgG levels in both milk and serum of transgenic mice. Immunology 122:401–408. doi:10.1111/j.1365-2567.2007.02654.x.

Lundell, K. 2004. The porcine taurochenodeoxycholic acid 6alpha-hydroxylase (CYP4A21) gene: evolution by gene duplication and gene conversion.. Biochem. J. 378:1053. doi:10.1042/BJ20031657.

Lundell, K., R. Hansson, and K. Wikvall. 2001. Cloning and expression of a pig liver taurochenodeoxycholic acid 6alpha-hydroxylase (CYP4A21): a novel member of the CYP4A subfamily. J. Biol. Chem. 276:9606–9612. doi:10.1074/jbc.M006584200.

Lv, Y., W. Guan, H. Qiao, C. Wang, F. Chen, Y. Zhang, and Z. Liao. 2015. Veterinary Medicine and Omics (Veterinomics): Metabolic Transition of Milk Triacylglycerol Synthesis in Sows from Late Pregnancy to Lactation. OMICS 19:602–616. doi:10.1089/omi.2015.0102.

Manjarin, R., J.P. Steibel, V. Zamora, N. Am-in, R.N. Kirkwood, C.W. Ernst, P.S. Weber, N.P. Taylor, and N.L. Trottier. 2011. Transcript abundance of amino acid transporters, β-casein, and α-lactalbumin in mammary tissue of periparturient, lactating, and postweaned sows. J. Dairy Sci. 94:3467–3476. doi:10.3168/jds.2011-4163.

Marioni, J.C., C.E. Mason, S.M. Mane, M. Stephens, and Y. Gilad. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18:1509–1517. doi:10.1101/gr.079558.108.

Mayer, B., A. Zolnai, L.V. Frenyó, V. Jancsik, Z. Szentirmay, L. Hammarström, and I. Kacskovics. 2002. Localization of the sheep FcRn in the mammary gland. Vet. Immunol. Immunopathol. 87:327–330.

Mayer, B., M. Doleschall, B. Bender, J. Bartyik, Z. Bosze, L.V. Frenyó, and I. Kacskovics. 2005. Expression of the neonatal Fc receptor (FcRn) in the bovine mammary gland. J. Dairy Res. 72 Spec No:107–112.

Mei, J., Y. Zhang, T. Wang, P.T. Sangild, and R.-J. Xu. 2006. Oral ingestion of colostrum alters intestinal transforming growth factor-beta receptor intensity in newborn pigs. Livest. Sci. 105:214–222. doi:10.1016/j.livsci.2006.06.017.

Messer, M., and C. Elliott. 1987. Changes in alpha-lactalbumin, total lactose, UDP-galactose hydrolase and other factors in tammar wallaby (Macropus eugenii) milk during lactation. Aust. J. Biol. Sci. 40:37–46.

Metzker, M.L. 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11:31–46. doi:10.1038/nrg2626.

Mohammad, M.A., D.L. Hadsell, and M.W. Haymond. 2012. Gene regulation of UDP-galactose synthesis and transport: potential rate-limiting processes in initiation of milk production in humans. Am. J. Physiol. Endocrinol. Metab. 303:E365-376. doi:10.1152/ajpendo.00175.2012.

Mohammad, M.A., and M.W. Haymond. 2013. Regulation of lipid synthesis genes and milk fat production in human mammary epithelial cells during secretory activation. Am. J. Physiol. Endocrinol. Metab. 305:E700–E716. doi:10.1152/ajpendo.00052.2013.

Moore, K.A., and J. Hollien. 2012. The unfolded protein response in secretory cell function. Annu. Rev. Genet. 46:165–183. doi:10.1146/annurev-genet-110711-155644.

Morrison, B.L., C.C. Jose, and M.L. Cutler. 2010. Connective Tissue Growth Factor (CTGF/CCN2) enhances lactogenic differentiation of mammary epithelial cells via integrin-mediated cell adhesion. BMC Cell Biol. 11:35. doi:10.1186/1471-2121-11-35.

Munne, P.M., Y. Gu, M. Tumiati, P. Gao, S. Koopal, S. Uusivirta, J. Sawicki, G.-H. Wei, and S.G. Kuznetsov. 2014. TP53 supports basal-like differentiation of mammary epithelial cells by preventing translocation of deltaNp63 into nucleoli. Sci. Rep. 4:4663. doi:10.1038/srep04663.

Murakami, N., T. Ohtsubo, Y. Kansui, K. Goto, H. Noguchi, Y. Haga, Y. Nakabeppu, K. Matsumura, and T. Kitazono. 2014. Mice heterozygous for the xanthine oxidoreductase gene facilitate lipid accumulation in adipocytes. Arterioscler. Thromb. Vasc. Biol. 34:44–51. doi:10.1161/ATVBAHA.113.302214.

Nagle, C.A., L. Vergnes, H. DeJong, S. Wang, T.M. Lewin, K. Reue, and R.A. Coleman. 2008. Identification of a novel sn-glycerol-3-phosphate acyltransferase isoform, GPAT4, as the enzyme deficient in Agpat6−/− mice. J. Lipid Res. 49:823. doi:10.1194/jlr.M700592-JLR200.

Nathan, C. 2006. Neutrophils and immunity: challenges and opportunities. Nat. Rev. Immunol. 6:173–182. doi:10.1038/nri1785.

Nemir, M., D. Bhattacharyya, X. Li, K. Singh, A.B. Mukherjee, and B.B. Mukherjee. 2000. Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency. J. Biol. Chem. 275:969–976.

Newcomb, M.D., D.L. Harmon, J.L. Nelssen, A.J. Thulin, and G.L. Allee. 1991. Effect of energy source fed to sows during late gestation on neonatal blood metabolite homeostasis, energy stores and composition. J. Anim. Sci. 69:230–236.

Nguyen, T.V., L. Yuan, M.S.P. Azevedo, K. Jeong, A.-M. Gonzalez, and L.J. Saif. 2007. Transfer of maternal cytokines to suckling piglets: in vivo and in vitro models with implications for immunomodulation of neonatal immunity. Vet. Immunol. Immunopathol. 117:236. doi:10.1016/j.vetimm.2007.02.013.

Ning, S., J.S. Pagano, and G.N. Barber. 2011. IRF7: activation, regulation, modification and function. Genes Immun. 12:399–414. doi:10.1038/gene.2011.21.

Nissen, A., E. Bendixen, K.L. Ingvartsen, and C.M. Røntved. 2012. In-depth analysis of low abundant proteins in bovine colostrum using different fractionation techniques. Proteomics 12:2866–2878. doi:10.1002/pmic.201200231.

Nissen, S., T.D. Faidley, D.R. Zimmerman, R. Izard, and C.T. Fisher. 1994. Colostral milk fat percentage and pig performance are enhanced by feeding the leucine metabolite beta-hydroxy-beta-methyl butyrate to sows. J. Anim. Sci. 72:2331–2337.

Nookaew, I., M. Papini, N. Pornputtapong, G. Scalcinati, L. Fagerberg, M. Uhlén, and J. Nielsen. 2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. Nucleic Acids Res. 40:10084–10097. doi:10.1093/nar/gks804.

Omura, T. 1999. Forty years of cytochrome P450. Biochem. Biophys. Res. Commun. 266:690–698. doi:10.1006/bbrc.1999.1887.

Palmquist, D.L. 2006. Milk Ffat: origin of fatty acids and influence of nutritional factors thereon. Adv. Dairy Chem. 2 Lipids:43–92. doi:10.1007/0-387-28813-9_2.

Palmquist, D.L., C.L. Davis, R.E. Brown, and D.S. Sachan. 1969. Availability and metabolism of various substrates in ruminants. V. Entry rate into the body and Incorporation into milk fat of d(−)β-hydroxybutyrate1. J. Dairy Sci. 52:633–638. doi:10.3168/jds.S0022-0302(69)86620-8.

Pastorelli, G., M. Neil, and I. Wigren. 2009. Body composition and muscle glycogen contents of piglets of sows fed diets differing in fatty acids profile and contents. Livest. Sci. 123:329–334. doi:10.1016/j.livsci.2008.11.023.

van der Poll, T., and H. Herwald. 2014. The coagulation system and its function in early immune defense. Thromb. Haemost. 112:640–648. doi:10.1160/TH14-01-0053.

Poltorak, A., X. He, I. Smirnova, M.Y. Liu, C. Van Huffel, X. Du, D. Birdwell, E. Alejos, M. Silva, C. Galanos, M. Freudenberg, P. Ricciardi-Castagnoli, B. Layton, and B. Beutler. 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science 282:2085–2088.

Quesnel, H., C. Farmer, and N. Devillers. 2012. Colostrum intake: Influence on piglet performance and factors of variation. Livest. Sci. 146:105–114. doi:10.1016/j.livsci.2012.03.010.

Rainard, P., and C. Riollet. 2006. Innate immunity of the bovine mammary gland. Vet. Res. 37:369–400. doi:10.1051/vetres:2006007.

Rainard, P., C. Riollet, P. Berthon, P. Cunha, A. Fromageau, C. Rossignol, and F.B. Gilbert. 2008. The chemokine CXCL3 is responsible for the constitutive chemotactic activity of bovine milk for neutrophils. Mol. Immunol. 45:4020–4027. doi:10.1016/j.molimm.2008.06.010.

Ramakrishnan, B., P.S. Shah, and P.K. Qasba. 2001. alpha-Lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase I (beta 4Gal-T1) to transfer glucose from UDP-glucose to N-acetylglucosamine. Crystal structure of beta 4Gal-T1 x LA complex with UDP-Glc. J. Biol. Chem. 276:37665–37671. doi:10.1074/jbc.M102458200.

Rao, A., C. Luo, and P.G. Hogan. 1997. Transcription factors of the NFAT family: regulation and function. Annu. Rev. Immunol. 15:707–747. doi:10.1146/annurev.immunol.15.1.707.

Rezaei, R., Z. Wu, Y. Hou, F.W. Bazer, and G. Wu. 2016. Amino acids and mammary gland development: nutritional implications for milk production and neonatal growth. J. Anim. Sci. Biotechnol. 7:20. doi:10.1186/s40104-016-0078-8.

Rikitake, Y., S. Kawashima, S. Takeshita, T. Yamashita, H. Azumi, M. Yasuhara, H. Nishi, N. Inoue, and M. Yokoyama. 2001. Anti-oxidative properties of fluvastatin, an HMG-CoA reductase inhibitor, contribute to prevention of atherosclerosis in cholesterol-fed rabbits. Atherosclerosis 154:87–96.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43:e47. doi:10.1093/nar/gkv007.

Robinson, A.M., and D.H. Williamson. 1980. Physiological roles of ketone bodies as substrates and signals in mammalian tissues.. Physiol. Rev. 60:143–187.

Robinson, G.W., R.A. McKnight, G.H. Smith, and L. Hennighausen. 1995. Mammary epithelial cells undergo secretory differentiation in cycling virgins but require pregnancy for the establishment of terminal differentiation. Development 121:2079–2090.

Robinson, M.D., and G.K. Smyth. 2007. Moderated statistical tests for assessing differences in tag abundance. Bioinforma. Oxf. Engl. 23:2881–2887. doi:10.1093/bioinformatics/btm453.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. 26:139–140. doi:10.1093/bioinformatics/btp616.

Robinson, M.D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25. doi:10.1186/gb-2010-11-3-r25.

Rodriguez, B. de J., C. Chevaleyre, G. Henry, D. Mollé, I. Virlogeux-Payant, M. Berri, F. Boulay, J. Léonil, F. Meurens, and H. Salmon. 2009. Identification in milk of a serum amyloid A peptide chemoattractant for B lymphoblasts. BMC Immunol. 10:4. doi:10.1186/1471-2172-10-4.

Rodriguez-Cruz, M., A.R. Tovar, B. Palacios-González, M. Del Prado, and N. Torres. 2006. Synthesis of long-chain polyunsaturated fatty acids in lactating mammary gland: role of Delta5 and Delta6 desaturases, SREBP-1, PPARalpha, and PGC-1. J. Lipid Res. 47:553–560. doi:10.1194/jlr.M500407-JLR200.

Rodriguez-Cruz, M., R. Sánchez, A.M. Sánchez, S.L. Kelleher, F. Sánchez-Muñoz, J. Maldonado, and M. López-Alarcón. 2011. Participation of mammary gland in long-chain polyunsaturated fatty acid synthesis during pregnancy and lactation in rats. Biochim. Biophys. Acta 1811:284–293. doi:10.1016/j.bbalip.2011.01.007.

Rudolph, M.C., J.L. McManaman, T. Phang, T. Russell, D.J. Kominsky, N.J. Serkova, T. Stein, S.M. Anderson, and M.C. Neville. 2007. Metabolic regulation in the lactating mammary gland: a lipid synthesizing machine. Physiol. Genomics 28:323–336. doi:10.1152/physiolgenomics.00020.2006.

Salmon, H. 2000. Mammary gland immunology and neonate protection in pigs. Homing of lymphocytes into the MG. Adv. Exp. Med. Biol. 480:279–286. doi:10.1007/0-306-46832-8_32.

Sambasivan, R., S. Cheedipudi, N. Pasupuleti, A. Saleh, G.K. Pavlath, and J. Dhawan. 2009. The small chromatin-binding protein p8 coordinates the association of anti-proliferative and pro-myogenic proteins at the myogenin promoter. J. Cell Sci. 122:3481–3491. doi:10.1242/jcs.048678.

Sanchez, L., L. Lujan, R. Oria, H. Castillo, D. Perez, J.M. Ena, and M. Calvo. 1992. Synthesis of lactoferrin and transport of transferrin in the lactating mammary gland of sheep. J. Dairy Sci. 75:1257–1262. doi:10.3168/jds.S0022-0302(92)77875-8.

Seerley, R.W., T.A. Pace, C.W. Foley, and R.D. Scarth. 1974. Effect of energy Intake prior to parturition on milk lipids and survival rate, thermostability and carcass composition of piglets. J. Anim. Sci. 38:64–70. doi:10.2527/jas1974.38164x.

Shaw, P.H. 1996. The role of p53 in cell cycle regulation. Pathol. Res. Pract. 192:669–675. doi:10.1016/S0344-0338(96)80088-4.

Sheehy, P.A., L.G. Riley, H.W. Raadsma, P. Williamson, and P.C. Wynn. 2009. A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. Anim. Genet. 40:492–498. doi:10.1111/j.1365-2052.2009.01862.x.

Shennan, D.B., D.T. Calvert, M.T. Travers, Y. Kudo, and C. a. R. Boyd. 2002. A study of L-leucine, L-phenylalanine and L-alanine transport in the perfused rat mammary gland: possible involvement of LAT1 and LAT2. Biochim. Biophys. Acta 1564:133–139.

Shennan, D.B., and C. a. R. Boyd. 2014. The functional and molecular entities underlying amino acid and peptide transport by the mammary gland under different physiological and pathological conditions. J. Mammary Gland Biol. Neoplasia 19:19–33. doi:10.1007/s10911-013-9305-5.

Shi, H., J. Luo, J. Zhu, J. Li, Y. Sun, X. Lin, L. Zhang, D. Yao, and H. Shi. 2013. PPARγ regulates genes involved in triacylglycerol synthesis and secretion in mammary gland epithelial cells of dairy goats. PPAR Res. 2013:310948. doi:10.1155/2013/310948.

Shi, H., W. Zhao, C. Zhang, K. Shahzad, J. Luo, and J.J. Loor. 2016. Transcriptome-wide analysis reveals the role of PPARγ controlling the lipid metabolism in goat mammary epithelial cells. PPAR Res. 2016:9195680. doi:10.1155/2016/9195680.

Simpson, A.E. 1997. The cytochrome P450 4 (CYP4) family. Gen. Pharmacol. 28:351–359.

Simpson, K.J., P. Bird, D. Shaw, and K. Nicholas. 1998. Molecular characterisation and hormone-dependent expression of the porcine whey acidic protein gene. J. Mol. Endocrinol. 20:27–35.

Sriburi, R., H. Bommiasamy, G.L. Buldak, G.R. Robbins, M. Frank, S. Jackowski, and J.W. Brewer. 2007. Coordinate regulation of phospholipid biosynthesis and secretory pathway gene

expression in XBP-1(S)-induced endoplasmic reticulum biogenesis. J. Biol. Chem. 282:7024–7034. doi:10.1074/jbc.M609490200.

Stein, T., J.S. Morris, C.R. Davies, S.J. Weber-Hall, M.-A. Duffy, V.J. Heath, A.K. Bell, R.K. Ferrier, G.P. Sandilands, and B.A. Gusterson. 2004. Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. Breast Cancer Res. 6:R75-91. doi:10.1186/bcr753.

Stelwagen, K., E. Carpenter, B. Haigh, A. Hodgkinson, and T.T. Wheeler. 2009. Immune components of bovine colostrum and milk. J. Anim. Sci. 87:3–9. doi:10.2527/jas.2008-1377.

Suárez-Vega, A., B. Gutiérrez-Gil, C. Klopp, G. Tosser-Klopp, and J.-J. Arranz. 2016. Comprehensive RNA-Seq profiling to evaluate lactating sheep mammary gland transcriptome. Sci. Data 3:160051. doi:10.1038/sdata.2016.51.

Takeuchi, K., and K. Reue. 2009. Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. Am. J. Physiol. Endocrinol. Metab. 296:E1195–E1209. doi:10.1152/ajpendo.90958.2008.

Tansey, J.T., C. Sztalryd, E.M. Hlavin, A.R. Kimmel, and C. Londos. 2004. The central role of perilipin a in lipid metabolism and adipocyte lipolysis. IUBMB Life 56:379–385. doi:10.1080/15216540400009968.

Tarazona, S., P. Furió-Tarí, D. Turrà, A.D. Pietro, M.J. Nueda, A. Ferrer, and A. Conesa. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. 43:e140–e140. doi:10.1093/nar/gkv711.

The Danish Ministry of Justice. 1995. Animal testing act, consolidation act no. 726 of September 9, 1993 (as amended by act no. 1081 of December 20, 1995). The Danish Ministry of Justice, Copenhagen, Denmark. 1995.

Theil, P.K., R. Labouriau, K. Sejrsen, B. Thomsen, and M.T. Sørensen. 2005. Expression of genes involved in regulation of cell turnover during milk stasis and lactation rescue in sow mammary glands. J. Anim. Sci. 83:2349–2356.

Theil, P.K., K. Sejrsen, W.L. Hurley, R. Labouriau, B. Thomsen, and M.T. Sørensen. 2006. Role of suckling in regulating cell turnover and onset and maintenance of lactation in individual mammary glands of sows. J. Anim. Sci. 84:1691–1698. doi:10.2527/jas.2005-518.

Theil, P.K., G. Cordero, P. Henckel, L. Puggaard, N. Oksbjerg, and M.T. Sørensen. 2011. Effects of gestation and transition diets, piglet birth weight, and fasting time on depletion of glycogen pools in liver and 3 muscles of newborn piglets. J. Anim. Sci. 89:1805–1816. doi:10.2527/jas.2010-2856.

Theil, P.K., C. Lauridsen, and H. Quesnel. 2014. Neonatal piglet survival: impact of sow nutrition around parturition on fetal glycogen deposition and production and composition of colostrum and transient milk. Animal 8:1021–1030. doi:10.1017/S1751731114000950.

Tomàs, A., J. Estellé, A. Clop, L. Gómez-Raya, J.L. Noguera, A. Sànchez, and M. Amills. 2003. Assignment of the mitochondrial glycerol-3-phosphate acyltransferase (GPAT) gene to porcine chromosome 14. Anim. Genet. 34:387.

Tramontana, S., M. Bionaz, A. Sharma, D.E. Graugnard, E.A. Cutler, P. Ajmone-Marsan, W.L. Hurley, and J.J. Loor. 2008. Internal controls for quantitative polymerase chain reaction of swine mammary glands during pregnancy and lactation. J. Dairy Sci. 91:3057–3066. doi:10.3168/jds.2008-1164.

Trapnell, C., D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, and L. Pachter. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31. doi:10.1038/nbt.2450.

Trégoat, V., P. Montagne, M.-L. Cuillière, M.-C. Béné, and G. Faure. 1999. C3/C4 concentration ratio reverses between colostrum and mature milk in human lactation. J. Clin. Immunol. 19:300–304. doi:10.1023/A:1020591508444.

Tsukamoto, H., K. Fukudome, S. Takao, N. Tsuneyoshi, and M. Kimoto. 2010. Lipopolysaccharide-binding protein-mediated Toll-like receptor 4 dimerization enables rapid signal transduction against lipopolysaccharide stimulation on membrane-associated CD14-expressing cells. Int. Immunol. 22:271–280. doi:10.1093/intimm/dxq005.

Vailati-Riboni, M., M. Kanwal, O. Bulgari, S. Meier, N.V. Priest, C.R. Burke, J.K. Kay, S. McDougall, M.D. Mitchell, C.G. Walker, M. Crookenden, A. Heiser, J.R. Roche, and J.J. Loor. 2016. Body condition score and plane of nutrition prepartum affect adipose tissue transcriptome regulators of metabolism and inflammation in grazing dairy cows during the transition period. J. Dairy Sci. 99:758–770. doi:10.3168/jds.2015-10046.

Vallet, J.L., A.K. McNeel, J.R. Miles, and B.A. Freking. 2014. Placental accommodations for transport and metabolism during intra-uterine crowding in pigs. J. Anim. Sci. Biotechnol. 5:55. doi:10.1186/2049-1891-5-55.

Vousden, K.H., and C. Prives. 2009. Blinded by the light: the growing complexity of p53. Cell 137:413–431. doi:10.1016/j.cell.2009.04.037.

Wagstrom, E.A., K.J. Yoon, and J.J. Zimmerman. 2000. Immune components in porcine mammary secretions. Viral Immunol. 13:383–397. doi:10.1089/08828240050144699.

Wang, S., Z. Chen, V. Lam, J. Han, J. Hassler, B.N. Finck, N.O. Davidson, and R.J. Kaufman. 2012. IRE1α-XBP1s induces PDI expression to increase MTP activity for hepatic VLDL assembly and lipid homeostasis. Cell Metab. 16:473–486. doi:10.1016/j.cmet.2012.09.003.

West, C.E., R. Bickerstaffe, E.F. Annison, and J.L. Linzell. 1972. Studies on the mode of uptake of blood triglycerides by the mammary gland of the lactating goat. The uptake and incorporation

into milk fat and mammary lymph of labelled glycerol, fatty acids and triglycerides. Biochem. J. 126:477–490.

Wheeler, T.T., A.J. Hodgkinson, C.G. Prosser, and S.R. Davis. 2007. Immune components of colostrum and milk--a historical perspective. J. Mammary Gland Biol. Neoplasia 12:237–247. doi:10.1007/s10911-007-9051-7.

Xie, L., and S.M. Innis. 2008. Genetic variants of the FADS1 FADS2 gene cluster are associated with altered (n-6) and (n-3) essential fatty acids in plasma and erythrocyte phospholipids in women during pregnancy and in breast milk during lactation. J. Nutr. 138:2222–2228. doi:10.3945/jn.108.096156.

Xyni, K., D. Rizos, G. Giannaki, A. Sarandakou, I. Phocas, and G. Creatsas. 2000. Soluble form of ICAM-1, VCAM-1, E- and L-selectin in human milk.. Mediators Inflamm. 9:133–140.

Yamashita, A., N. Kawagishi, T. Miyashita, T. Nagatsuka, T. Sugiura, K. Kume, T. Shimizu, and K. Waku. 2001. ATP-independent fatty acyl-coenzyme A synthesis from phospholipid: coenzyme A-dependent transacylation activity toward lysophosphatidic acid catalyzed by acyl-coenzyme A:lysophosphatidic acid acyltransferase. J. Biol. Chem. 276:26745–26752. doi:10.1074/jbc.M101795200.

Yang, W.-C., W.-C. Tsai, P.-M. Lin, M.-Y. Yang, Y.-C. Liu, C.-S. Chang, W.-H. Yu, and S.-F. Lin. 2013. Human BDH2, an anti-apoptosis factor, is a novel poor prognostic factor for de novo cytogenetically normal acute myeloid leukemia. J. Biomed. Sci. 20:58. doi:10.1186/1423-0127-20-58.

Zhang, L., S. Boeren, J.A. Hageman, T. van Hooijdonk, J. Vervoort, and K. Hettinga. 2015. Bovine milk proteome in the first 9 days: protein interactions in maturation of the immune and digestive system of the newborn. PLoS One 10:e0116710. doi:10.1371/journal.pone.0116710.

Zhang, Y., and R. Davis. 2014. Principal trend analysis for time-course data with applications in genomic medicine. doi:10.1214/13-AOAS659.

Zhang, Z.H., D.J. Jhaveri, V.M. Marshall, D.C. Bauer, J. Edson, R.K. Narayanan, G.J. Robinson, A.E. Lundberg, P.F. Bartlett, N.R. Wray, and Q.-Y. Zhao. 2014. A comparative study of techniques for differential expression analysis on RNA-Seq data. PloS One 9:e103207. doi:10.1371/journal.pone.0103207.

Zhao, W., K. Shahzad, M. Jiang, D.E. Graugnard, S.L. Rodriguez-Zas, J. Luo, J.J. Loor, and W.L. Hurley. 2013. Bioinformatics and gene network analyses of the swine mammary gland transcriptome during late gestation. Bioinforma. Biol. Insights 7:193–216. doi:10.4137/BBI.S12205.

Zhou, Y., W. Gong, J. Xiao, J. Wu, L. Pan, X. Li, X. Wang, W. Wang, S. Hu, and J. Yu. 2014. Transcriptomic analysis reveals key regulators of mammogenesis and the pregnancy-lactation cycle. Sci. China Life Sci. 57:340–355. doi:10.1007/s11427-013-4579-9.

Chapter IV - PIA (PATHWAYS INTERACTION ANALYSIS): AN R TOOL FOR ANALYSING AND INTERPRETING HIGH-THROUGHPUT DATA

Abstract

Increasing quantitative data generated from genomic and transcriptomic analysis requires integrative strategies to solve the challenge of data mining. The use of tools for pathway analysis or functional enrichment is *de facto* standard for the secondary analysis of high-throughput experiments. Nevertheless, the majority of these tools perform the analysis within a single pathway, not providing an integrated summary in terms of networks or interactions among more pathways of interest and related groups of genes.

Here we present Pathways Interaction Analysis (PIA), an R package that classifies functionally related genes taking into account a network of both upstream and downstream pathways in interaction. The network-based result helps to interpret functional profiles of cluster of genes underlying complex biological processes.

The suite has no species constraints, and is functional to analyse genomic or transcriptomic outcomes.

Index of chapter

IV - 1. Introduction.....158

IV - 1.1. Studying omic data using pathways.....158

IV - 1.2. Statistical methods for pathway enrichment analysis.....159

IV - 2. Aim of the study.....159

IV - 3. Methods.....159

IV - 3.2. Package installation.....160

IV - 3.3. Package functionality.....160

IV - 3.4. Data preparation.....160

IV - 3.4.1. Gene list dataset (post-genomic analysis dataset).....162

IV - 3.4.2. Gene expression dataset.....162

IV - 3.4.3. Pathways of interest.....163

IV - 3.4.4. Species code detection.....164

IV - 3.5. Data analyses.....165

IV - 3.5.1. Gene list dataset (post-genomic analysis dataset).....165

IV - 3.5.2. Gene expression dataset.....166

IV - 3.5.3. Enrichment analysis.....168

IV - 4. Validation.....169

IV - 4.1. post-GWAS dataset analysis.....170

IV - 4.2. Enrichment analysis.....172

IV - 4.3. Expression dataset analysis.....173

IV - 5. Conclusion.....175

IV - 6. Figures and tables.....176

IV - 7. References.....184

List of chapter figures and tables

Figure IV-1. The general architecture of the workflow of PIA package and schematic illustration of main functions.

Figure IV-2. An example of the '.html' file with the network-based visualization of PIA results.

Figure IV-3. An example of node selection of PIA network-based visualization result.

Figure IV-4. An example of the '.html' file with the network-based visualization of PIA result considering an expression dataset.

Figure IV-5. Node selection of PIA network-based visualization result obtained on expression dataset.

Figure IV-6. Network-based visualization of result obtained by PIA considering three interaction levels.

Figure IV-7. Network-based visualization of PIA result considering expression dataset (Levy et al., 2012; Qiu et al., 2014).

Table IV-1. Summary of node colour classification in network visualization obtained with `pia.exprscript()` function.

Table IV-2. Summary of PIA results for three interaction levels, considering 'Type I diabetes mellitus', 'Insulin resistance', and 'AGE-RAGE signaling pathway in diabetic complications' as FDI or 1DI pathways.

Table IV-3. Comparison between PIA and reference study results (Qiu et al., 2014).

In supplementary material section

Supplementary Table IV-S1. Summary of PIA enrichment analysis (adjusted p-value ≤ 0.05).

IV - 1. Introduction

Thanks to advancements in high-throughput techniques and simultaneous reduction in the associated costs, large scale 'omics' studies are now common, enabling the generation of a huge amount of biological data (Joyce and Palsson, 2006) and posing the challenge of data mining rather than data production to researchers. Typically, a key result of genomic or transcriptomic analysis (e.g. genome-wide association study, runs of homozygosity, selection signature, expression profile from microarray or RNA sequencing technologies etc…) is a long list of statistically significant genes that contribute to the phenotypes or physiological conditions of interest. The subsequent step is to extract meaning from this list in order to provide insights into the underlying biology of the state under study (Khatri et al., 2012b).

IV - 1.1. Studying omic data using pathways

To reduce the complexity of omic data mining, one common approach is to simplify the analysis by grouping long lists of individual genes into smaller sets of related ones sharing the same biological processes or molecular functions. This method, known as 'pathway analysis' (Curtis et al., 2005), has become popular during the last few years (Rk et al., 2005) and is *de facto* standard for the secondary analysis of high-throughput experiments (Khatri and Drăghici, 2005b).

This approach is driven by the increasing availability of free accessible repositories based on hierarchical and functional classification of terms (Ashburner et al., 2000). In this regard, many sources of pathway and functional information, which can be either generic or species-specific, are now available (Khatri and Drăghici, 2005a). These knowledge databases include Kyoto Encyclopedia of Genes and Genomes (KEGG) that represents a prominent reference repository constantly updated (Kanehisa et al., 2017). KEGG is a bioinformatics resource that annotates genes to specific pathways and helps the understanding of organism genome information (Kanehisa et al., 2017). At the same time, a large number of tools for pathway analysis have been developed (Berg et al., 2009b; Khatri et al., 2012a). Nevertheless, the majority of them perform the functional enrichment analysis within a single item (i.e. pathway) (Curtis et al., 2005) not providing integrated information in terms of networks or interactions among more pathways of interest and related groups of genes (Cirillo et al., 2017).

IV - 1.2. Statistical methods for pathway enrichment analysis

Particularly in the context of gene expression data mining, an interesting step is to ask if there are any pathways or classes that are significantly over-represented (over-representation analysis) (Khatri et al., 2012a). This involves comparing the list of identified genes to that of those from a specific pathway or classification with the aim to identify if there are more matches than would be expected by chance. Several statistical methods can be used for this purpose (Draghici et al., 2003; Beissbarth and Speed, 2004; Curtis et al., 2005), hypergeometric distribution test (one-sided Fisher's exact test) is one of the most common (Tavazoie et al., 1999; Curtis et al., 2005; Simoes and Emmert-Streib, 2012). The result is a p-value describing the likelihood of obtaining the observed result that can be corrected for multiple tests (Draghici et al., 2003).

IV - 2. Aim of the study

Here, PIA (Pathway Interaction Analysis) package is introduced. PIA is a suite of scripts built in R (Ihaka and Gentleman, 1996b), that starting from a list of significant genes and main pathways of interest, highlights networks of genes in interactions at multiple levels (upstream/downstream pathway levels) based on information available on KEGG (Kanehisa et al., 2017). The results help to interpret high-throughput data and identify candidate genes for function that can influence multiple and complex biological processes. We believe that PIA data visualization, in the form of interactive pathway diagrams and gene-pathway biological interactions such as genetic networks, enhances interpretation of scientific data, increases understanding the conclusions drawn, and promotes discussion and follow-up research questions.

IV - 3. Methods

The package is specifically designed for the data mining of post-genomic and transcriptomic analyses and can handle data obtained from many species. The analysis is completely based on information available on KEGG databases (Kanehisa et al., 2017).

The rationale of interaction analysis performed by PIA is to highlight candidate genes for function, taking into account multiple levels of upstream and downstream pathways connected to a set of main pathways of interest (first-degree interaction pathways – FDI or 1DI) known to be involved with the phenotype/condition under study. A list of genes is ordered into an

interaction network of multiple level pathways created by PIA based on KEGG information and starting from the FDI pathways chosen by the user to perform the investigation. More specifically, PIA uncovers the relations among the pathways at the FDI level, then it selects the upstream or downstream pathways in interaction (from 2 - second degree interaction, 2DI - to *n* degree of interaction) finally, once the backbone of pathways in interaction is created, PIA pin-points the genes in the multiple levels investigated and provides a pathways/genes-based network visualization.

IV - 3.2. Package installation

This thesis focuses on PIA v.1.1.1. This and further versions of the package can be easily installed in any R session using the `install.packages('PIA')` command. Once installed, the package is loaded in the R environment with the `library('PIA')` command. The tool requires several R libraries automatically uploaded along with the package.

IV - 3.3. Package functionality

PIA package functions could be divided in two different steps: data preparation and data analyses (Figure IV-1). The first step helps to obtain all parameters and information needed to run the analysis and to prepare a properly formatted list of genes and FDI pathways of interest. The second step provides enrichment analysis, pathway interaction analysis and results visualization.

Since PIA interrogates KEGG databases, an internet connection is required to run the functions. Trial datasets are downloadable along with the package and can be stored in the working directory using the command `pia.example()`command.

IV - 3.4. Data preparation

PIA package requires the entrez gene identifiers (ID) to work. In fact, only the genes with entrez ID have a corresponding item in KEGG databases. To enhance user experience, specific functions based on biomaRt package (Durinck et al., 2009) are provided to retrieve the correct entrez annotation: `pia.dataPreparation()` and `pia.exprdataPreparation()`. The two functions prepare the dataset with the correct format for PIA analysis on a simple gene list or as gene expression data. In particular, they allow us to get the PIA requested gene ID along with a gene symbol.

The biomaRt organism code is needed to run proper data preparation functions and is obtainable with `pia.biomartSpecies()` command.

```
# Look for a specific organism code for biomaRt annotation, matching a search
string
> list <- pia.biomartSpecies(string = "cow")
The list of available species matched your string was created!
Remember to use the correct organism code for relative PIA functions.
> head(list)
organism_code   description   version
1   btaurus_gene_ensembl   Cow genes (UMD3.1)   UMD3.1
> biomart.species.bos <- as.character(list[1,1]) #btaurus_gene_ensembl
> biomart.species.bos
[1] btaurus_gene_ensembl
# Example of gene list preparation from dataset with ensembl id
# Copy the example data file 'ensembl_genelist.txt' in your current working
directory
> pia.example()
> genelist <- read.table("ensembl_genelist.txt", header = FALSE)
> head(genelist)
V1
1   ENSBTAG00000000039
2   ENSBTAG00000000040
3   ENSBTAG00000000042
4   ENSBTAG00000000044
5   ENSBTAG00000001521
6   ENSBTAG00000001522
>      genelist.converted      <-      pia.dataPreparation(in.file      =
"ensembl_genelist.txt",    gene_id    =    "ensembl",    biomart.species    =
biomart.species.bos)
Input file imported!
BiomaRt species correct!
Gene id correct!
Convertion from ensembl ID to entrez ID ...
DONE
n. 36 of 47 genes have corresponding gene in KEGG database.
Gene list exported!
```

The use of data preparation functions is recommended but not mandatory. In fact, it must be emphasized how their correct performance depends on the availability of biomaRt data access for a specific species of interest and of its correct ID annotation. For this reason, we strongly suggest to double-check all-possible gene annotations, according to the data format guidelines, as shown below.

IV - 3.4.1. Gene list dataset (post-genomic analysis dataset)

PIA requires a specific input format dataset. The list of genes of interest must be provided as '.txt' file, containing three columns labelled (as shown in trial dataset), and stored in a working directory.

```
# Copy the example data file 'data.txt' in your current working directory
> pia.example()
> genelist <- read.table("data.txt", header=TRUE)
> head(genelist)
ensembl_gene_id   entrezgene   external_gene_name
1   ENSBTAG00000000039   505662   SIRT7
2   ENSBTAG00000000040   515219   MAFG
3   ENSBTAG00000000042   539606   PYCR1
4   ENSBTAG00000000044   617922   MYADML2
5   ENSBTAG00000001521   616871   UQCRB
6   ENSBTAG00000001522   526138   MTERF3
```

IV - 3.4.2. Gene expression dataset

Also for the expression dataset, PIA requires a specific input format. The list of differentially expressed genes (DEG) of interest, with the relative fold change (FC) and p-value, must be provided as '.txt' file, containing five columns (as shown in trial dataset), and stored in a working directory.

```
# Copy the example data file 'exprdata.txt' in your current working directory
> pia.example()
> expr.genelist <- read.table("exprdata.txt", header=TRUE)
> head(expr.genelist)
ensembl_gene_id   FC   pvalue    entrezgene    external_gene_name
1   ENSSSCG00000000002   -4.992506 0.00002740   NA    GTSE1
```

```
2   ENSSSCG00000000003   -1.230589   0.28308054    100518372    TTC38
3   ENSSSCG00000000005   -1.028827   0.88291084   100518729    CDPF1
4   ENSSSCG00000000006    3.737968    0.01211484    397239    PPARA
5   ENSSSCG00000000007    1.060006    0.64321216    100521087    TRMU
6   ENSSSCG00000000010    1.319015    0.31261992     100519144    FBLN1
```

IV - 3.4.3. Pathways of interest

The FDI pathways of interest, coded as KEGG path ID, are mandatory to run the core PIA functions.

The pathways are chosen by the user since are known or well-documented in literature to be involved in the phenotype/condition under study. When no previous biological assumptions are available or it is difficult to select pathways from literature information, `pia.stats.enrichement()` command can be useful to explore the list of interesting pathway candidates as FDI (see Data analyses section for further explanation).

The list of available pathways with the relative codes is obtainable by `pia.pathList()` command.

```
# Look for a specific pathway(s) for PIA, matching your search string
> list <- pia.pathList(string = "lipid")
The list of pathway(s), matched your string, was created!
Remember to use the correct path Id(s) for relative PIA functions.
> head(list)
path_description   path_ID
1   Glycerolipid metabolism   path:map00561
2   Glycerophospholipid metabolism   path:map00564
3   Ether lipid metabolism   path:map00565
4   Sphingolipid metabolism   path:map00600
5   Glycosphingolipid biosynthesis-lacto and neolacto series   path:map00601
6   Glycosphingolipid biosynthesis - globo and isoglobo series   path:map00603
```

For a genes list dataset analysis, the FDI must be provided as a vector of pathway identifiers (‘path_ID’).

```
# Create a vector of pathways of interest (FDI) for pia.script() function
```

```
> FDI.gene <- c("path:map00061", "path:map00062", "path:map00071", "path:map00072")
```

For a gene expression dataset analysis, the FDI must be provided as a '.txt' file, with the list of pathway identifiers and relative pathway expression estimated scores, and stored in a working directory. The pathway estimated score is obtainable by common gene set enrichment analysis or overrepresented approach analysis (Huang et al., 2009) (e.g. flux value (Bionaz et al., 2012c), as shown in trial data).

```
# Copy the example pathways list file 'expr_listPath.txt' for pia.exprscript() function in your current working directory
> pia.example()
> FDI.expr <- read.table("expr_listPath.txt", header=TRUE)
> head(FDI.expr)
path_ID   value
1   path:map00010   324.22879
2   path:map00020   -21.31287
3   path:map00071   385.73774
```

IV - 3.4.4. Species code detection

The KEGG organism code is needed to run proper data analysis functions. The list of codes is obtainable with `pia.speciesCode()` command.

```
# Look for the 'bos taurus' code
> list <- pia.speciesCode(string = "bos")
The list of available species matched your string was created!
> head(list)
species   pia_code
1   Bos taurus (cow)   bta
2   Bos mutus (wild yak)   bom
3   Bos indicus (zebu cattle)   biu
4   Malassezia globosa   mgl
5   Bosea sp. PAMC 26642   bop
6   Bosea sp. RAC05   bos
> KEGG.species.bos <- as.character(list[1,2]) #bta
> KEGG.species.bos
```

```
[1] bta
```

IV - 3.5. Data analyses

PIA analysis can be performed on a single gene list obtained from classical genomic analysis (Palombo et al., 2018) or on a gene expression dataset from transcriptomic study. In both cases, PIA helps to interpret functional profiles of genes, underlying complex biological processes, showing the genes falling inside a network of pathways in interaction. They then may be considered good functional candidates for the trait/condition under study.

IV - 3.5.1. Gene list dataset (post-genomic analysis dataset)

The `pia.script()` command allows PIA to perform the analysis on the gene list provided. The function requires (1) a properly formatted gene list, (2) a vector of FDI pathways, (3) the KEGG organism code and (4) the number of interaction levels. The interaction levels represent the number of upstream and downstream pathway levels (from 2 to *n*) required by the user for the investigation and connected to FDI pathways based on KEGG database information. Once the backbone of pathways in interaction is created, PIA highlights the genes falling inside the interaction network generated.

```
# Copy the example data file 'data.txt' in your current working directory
> pia.example()
# Perform PIA
> pia.script(in.file = "data.txt", out.file = "FA", species = KEGG.species.bos, FDI = FDI.gene, levels = 2)
Input file imported!
Gene list specified... and correct!
Species code specified... and correct!
Pathway(s) is specified... and correct!
Prerequisite check passed!
PIA is running ...
Please wait... It could be a while depending on the number of pathways and levels required!
PIA analysis completed and relative '.txt' files exported!
Preparing PIA diagram visualization!
Please wait... It could be a while depending on the number of pathways and levels required!
```

```
Well done! Diagram visualization was created and exported.
```

The function generates *n* '.txt' files, with *n* equal to the number of levels required for the investigation (in the above example *n* = 2), containing the highlighted genes and related pathways for each level of interaction.

```
# Summary of PIA results at FDI (1DI) level obtained with the example dataset
> genes.1DI <- read.table("PIA_RESULTS_FA/1DIgenes.txt", header = TRUE)
> genes.1DI
ensemblgene   entrezgene   gene_name   path_description   path_ID
1      ENSBTAG00000015980      281152      FASN      Fatty  acid  biosynthesis
path:bta00061
2   ENSBTAG00000015178   505355   ECI2   Fatty acid degradation   path:bta00071

# Summary of PIA results at 2DI level obtained with the example dataset
> genes.2DI <- read.table("PIA_RESULTS_FA/2DIgenes.txt", header = TRUE)
> genes.2DI
ensemblgene   entrezgene   gene_name   path_description   path_ID
1     ENSBTAG00000016253     369023     G6PC3     Glycolysis  /  Gluconeogenesis
path:bta00010
2    ENSBTAG00000001868    510274    PCYT2    Glycerophospholipid  metabolism
path:bta00564
```

Along with the tabular format, the function also provides the genes/pathways network visualization of PIA results, saved in a '.html' file. The diagram allows us to zoom in on all content for an optimal readability and it is interactive, enabling the selection of specific nodes (Figures IV-2 and -3).

IV - 3.5.2. Gene expression dataset

For dataset expression, PIA takes into account any possible interaction among the FDI pathways and the list of DEG. The dedicated function is `pia.exprscript()` and requires (1) a properly formatted DEG list, (2) a properly formatted FDI pathways list, (3) the KEGG organism code and (4) a p-value cut-off. The function generates a diagram visualization of an interaction network that helps to interpret the results obtained from gene expression experiments showing the nodes (i.e. genes and pathways) coloured according to their FCs and the provided pathway

expression estimated score (e.g. flux values (Bionaz et al., 2012c)), respectively. The node classification is a function of top FC or estimated score value, as shown below (Table IV-1).

```
# Copy the example data files "exprdata.txt" and 'expr_listPath.txt' in your
current working directory
> pia.example()
# Look for the 'sus scrofa' code
> list <- pia.speciesCode(string = "pig")
The list of available species matched your string was created!
Remember to use the correct organism code for relative PIA functions.
> list
species   pia_code
1   Sus scrofa (pig)   ssc
2   Columba livia (rock pigeon)   clv
3   Cajanus cajan (pigeon pea)   ccaj
4   Desulfovibrio piger   dpg
5   Salipiger profundus   tpro
6   Halopiger xanaduensis   hxa
> KEGG.species.sus <- as.character(list[1,2]) #ssc
> KEGG.species.sus
[1] ssc

# Perform PIA on gene expression dataset
> pia.exprscript(in.file = "exprdata.txt", path.file = "expr_listPath.txt",
out.file = "expression_data", species = KEGG.species.sus, pvalue = 0.05)
Input file imported!
Pathway input file imported!
Gene list specified... and correct!
Your path list colnames are correct!
Species code specified... and correct!
Pathway(s) is specified... and correct!
Prerequisite check passed!
PIA is running ...
Please wait... It could be a while depending on the number of pathways
required! n. 3151 of 7934 genes passed the p-value filtering.
Well done! Diagram visualization was created and exported.
```

The diagram is interactive allowing us to zoom in on all content for an optimal readability and select specific nodes (Figures IV-4 and IV-5).

IV - 3.5.3. Enrichment analysis

PIA allows us to perform an enrichment analysis for each KEGG term (i.e. pathways) based on a hypergeometric test (one-sided Fisher exact test) as described by Simoes and Emmert-Streib (Simoes and Emmert-Streib, 2012), with `pia.stats.enrichment()` command.

```
# Copy the example data file 'data.txt' in your current working directory
> pia.example()
# Perform the enrichment analysis
>  pia.stats.enrichment(in.file  =  "data.txt",  out.file  =  "enrichment_FA",
species = KEGG.species.bos)
Input file is imported!
Gene list specified... and correct!
Species code specified... and correct!
Enrichment analysis started ... and results exported!
Gene per pathway(s) table created and exported!
Pathway per gene(s) table created and exported!
```

The results are a series of '.txt' files with specific enrichment analysis results and with general descriptive information about single gene and pathway occurrences. For each pathway a p-value is calculated to estimate its probability of over-representation (Simoes and Emmert-Streib, 2012). This is useful to explore the list of relevant or interesting pathways when no previous restrictive biological assumptions are available.

```
# Summary of enrichment analysis results obtained with the example dataset
FA_enrich <- read.table("enrichment_FA_enrichment.txt", header = T)
head(FA_enrich)
pathway_ID   n_genes   all_genes   pvalue   padj   pathway_name
1   path:bta01100   8   1308   0.0008543217   0.2776545   Metabolic pathways
2   path:bta04920   2   72   0.0072621338   0.6386025   Adipocytokine signaling
pathway
3   path:bta05212   2   74   0.0076581146   0.6386025   Pancreatic cancer
4    path:bta04662    2    75    0.0078597229    0.6386025    B  cell  receptor
signaling pathway
```

```
5    path:bta00440    1    6    0.0108120071    0.7027805    Phosphonate and
phosphinate metabolism
6    path:bta04152    2    123    0.0202142955    0.8932402    AMPK signaling
pathway
FA_GxP <- read.table("enrichment_FA_GxP.txt", header = T)
head(FA_GxP)
n_genes   pathway_name   pathway_ID
1   8   Metabolic pathways   path:bta01100
2   2   MAPK signaling pathway   path:bta04010
3   2   Ras signaling pathway   path:bta04014
4   2   FoxO signaling pathway   path:bta04068
5   2   PI3K-Akt signaling pathway   path:bta04151
6   2   AMPK signaling pathway   path:bta04152

FA_PxG <- read.table("enrichment_FA_PxG.txt", header = T)
head(FA_PxG)
n_pathways   entrez_gene_id   ensembl_gene_id   gene_symbol
1   40   281073   ENSBTAG00000007591   CHUK
2   21   619066   ENSBTAG00000022927   RAC3
3   12   369023   ENSBTAG00000016253   G6PC3
4   8   616871   ENSBTAG00000001521   UQCRB
5   5   281152   ENSBTAG00000015980   FASN
6   3   510274   ENSBTAG00000001868   PCYT2
```

IV - 4. Validation

To evaluate the usefulness of the PIA approach, we used a publicly available dataset on human type 1 diabetes mellitus - T1DM (Qiu et al., 2014). In the reference study, the authors carried out a gene-based genome-wide association analysis and identified 452 significant genes. Among these genes, 171 were newly identified for type 1 diabetes mellitus, not previously described in literature. Fifty three out of 171 genes were further supported by replication or differential expression studies. Moreover, the authors reported four non-HLA genes (RASIP1, STRN4, BCAR1 and MYL2) and three HLA genes (FYN, HLA-J and PPP1R11) as validated by both replication and differential expression studies. We performed PIA considering the list of 171 newly identified genes, to verify the possible contribution of the PIA approach for candidate genes identification and more broadly for high-throughput data interpretation.

The validation datasets are downloadable along with the package and can be stored in the working directory using the command `pia.example(type="validation")` command.

IV - 4.1. post-GWAS dataset analysis

After data preparation, only 5 out of 171 genes had no entrez gene ID. These genes were excluded from the analysis and the list of 166 annotated genes was used to run the PIA function.

```
# Copy the example files used as validation set in the publication in the
current working directory
> pia.example(type="validation")
>  validation.genelist  <-  read.table("genelist_annotated_qiu2014.txt",
header=TRUE)
> length(validation.genelist$entrezgene)
[1] 166
> head(validation.genelist)
external_gene_name   entrezgene   ensembl_gene_id
1   ADAD1   132612   ENSG00000164113
2   ASCL2   430   ENSG00000183734
3   ATF7IP   55729   ENSG00000171681
4   BAK1   578   ENSG00000030110
5   BCAR1   9564   ENSG00000050820
6   BCL2A1   597   ENSG00000140379
```

Considering the complexity of the trait investigated, PIA was performed up to the third degree of interaction (Field and Tobias, 1997). The '*Type I diabetes mellitus*' (map04940), '*Insulin resistance*' (map04931) and '*AGE-RAGE signaling pathway in diabetic complications*' (map04933) pathways were chosen as FDI pathways (Greenbaum, 2002; Ramasamy et al., 2005). A summary of PIA results is reported in Table IV-2.

```
#Look for the homo sapiens KEGG specie code
list <- pia.speciesCode(string = "homo")
KEGG.species.homo = as.character(list[1,2]) #hsa
FDI = c("path:map04940", "path:map04931", "path:map04933")
levels = 3
# Run the PIA function
```

```
pia.script(in.file   =   “genelist_annotated_qiu2014.txt”,   out.file   =
"validation", species = KEGG.species.homo, FDI = FDI, levels = levels)
Input file imported!
Gene list specified... and correct!
Species code specified... and correct!
Pathway(s) is specified... and correct!
Prerequisite check passed!
PIA is running ...
Please wait... It could be a while depending on the number of pathways and
levels required!
PIA analysis completed and relative '.txt' files exported!
Preparing PIA diagram visualization!
Please wait... It could be a while depending on the number of pathways and
levels required!
Well done! Diagram visualization was created and exported.
```

Overall PIA results obtained from validations dataset are in line with reference study outcomes (Qiu et al., 2014), confirming the effectiveness of the PIA approach. In particular, 4 out of 7 genes validated both replication and differential expression studies (Qiu et al., 2014) were highlighted by PIA: *PTPN11*, *BCAR1*, *MYL2* and *FYN* (Figure 6). The other three genes (*RASIP1*, *STRN4* and *HLA-J*) were not detected by PIA since, although present in KEGG databases, they were not yet assigned to any pathway.

Along with these genes, PIA also highlighted other interesting genes (*ITPR3*, *BAK1*, *IL10*, *HMGB1* and *MICA*) not discussed by Qui et al. (Qiu et al., 2014), since validated only by the differential expression studies or only by the replication studies.

It is worth noting that PIA also highlighted other genes not discussed in the reference study (Qiu et al., 2014) but reported in literature as being associated with the susceptibility to T1DM disease, in some cases these genes are referred to in research conducted before the reference study. In particular, *CDK2* (Kim et al., 2017a), *RXRB* (Shi et al., 2016a), *MADCAM1* (Phillips et al., 2005a), *STAT4* (Bi et al., 2013), *BCL2A1* (Beyan et al., 2010a) and *SMAD7* (Chen et al., 2011). Simultaneously, it is worth noting that some genes were not highlighted by PIA, because (1) they fell out of the three pathway investigated levels (including *BRAP*, *FUT2*, *GNS*, *HIPK1*, *NUPR1*, *OR2B3*, *HIST1H4E*, *HIST1H2BF*, *OR2B3*, *OR2B6*, *OR2J2*, *OR5V1* and *SULT1A1* genes. (2) Although present in KEGG databases, were not yet assigned to any pathway. These

drawbacks clearly represent the main PIA limitations. A comparison among the PIA results and reference study (Qiu et al., 2014) is reported in Table IV-3.

IV - 4.2. Enrichment analysis

Accordingly to the reference study (Qiu et al., 2014), we performed the functional annotation clustering analysis of 452 identified T1DM genes, using the PIA enrichment analysis function.

```
# Copy the example files used as validation set in the publication in the
current working directory
> pia.example(type="validation")
enrich.genelist     <-     read.table("genelist_enrichment_qui2014.txt",
header=TRUE)
> head(enrich.genelist)
external_gene_name entrezgene ensembl_gene_id
1 OLFML3 56944 not_available
2 HIPK1 204851 not_available
3 IL10 3586 not_available
4 NSL1 25936 not_available
5 FAM46B 115572 not_available
6 LHX9 56956 not_available
#Look for the specie code matching the search string
> list <- pia.speciesCode(string = "homo")
> homo.species = as.character(list[1,2]) # hsa
# Run the PIA enrichment function
>   pia.stats.enrichment(in.file   =   "genelist_enrichment_qui2014.txt",
out.file = "validation", species=homo.species)
Input file is imported!
Gene list specified... and correct!
Species code specified... and correct!
Enrichment analysis started ...
and results exported!
Gene per pathway(s) table created and exported!
Pathway per gene(s) table created and exported!
```

The results showed as genes tend to be over-represented in immune diseases and immune system pathways (Supplementary Table IV-S1), according to Qiu et al. (Qiu et al., 2014).

```
#Summary of PIA enrichment result
> enrichment.result <- read.table("validation_enrichment.txt", header =
TRUE)
> head(enrichment.result)
pathway_ID n_genes all_genes pvalue padj pathway_name
1  path:hsa05322  26  133  1.338552e-22  4.403835e-20  Systemic  lupus
erythematosus
2 path:hsa05330 13 38 3.566961e-15 5.867652e-13 Allograft rejection
3 path:hsa04612 16 77 1.306151e-14 1.432412e-12 Antigen processing and
presentation
4 path:hsa05320 14 53 1.763583e-14 1.450547e-12 Autoimmune thyroid disease
5 path:hsa04940 13 43 2.264972e-14 1.490352e-12 Type I diabetes mellitus
6 path:hsa05332 12 41 3.705738e-13 2.031980e-11 Graft-versus-host disease
```

IV - 4.3. Expression dataset analysis

In order to create an example to illustrate visualization of gene expression values, we used FC valued obtained by Levy et al. (Levy et al., 2012) and considered by Qiu et al. (Qiu et al., 2014) as reference study for differential expression validation. Since the authors did not provide pathway estimation scores, we substituted those values with gene occurrences for each pathway of interest, obtained by PIA enrichment analysis.

```
#   Copy   the   example   data   files   "genelist_expr_Levy2012.txt"   and
'pathlist_expr_Levy2012.txt' in your current working directory
> pia.example(type="validation")
>   expr.validation.genelist   <-   read.table("genelist_expr_Levy2012.txt”,
header = TRUE)
> head(expr.validation.genelist)
ensembl_gene_id   external_gene_name   entrezgene   FC   pvalue
1   ENSG00000204252   HLA-DOA   3111   -0.535   0.000
2   ENSG00000239457   HLA-DOB   3112   -0.017   0.955
3   ENSG00000168384   HLA-DPA1   3113   -0.491   0.025
4   ENSG00000206239   HLA-DQA1   3117   -0.846   0.000
5   ENSG00000206237   HLA-DQB1   3119   -0.467   0.007
6   ENSG00000204592   HLA-E   3133   0.298   0.014
```

```
> expr.validation.path <- read.table("pathlist_expr_Levy2012.txt", header =
TRUE)
> head(expr.validation.path)
path_ID   value
1   path:map04514   16
2   path:map04940   14
3   path:map04151   6
4   path:map04210   5
5   path:map04630   5
6   path:map04010   4

#Look for the specie code matching the search string
> list <- pia.speciesCode(string = "homo")
> KEGG.species.homo = as.character(list[1,2]) #hsa
# Perform PIA on transcriptomic dataset
> pia.exprscript(in.file = "genelist_expr_Levy2012.txt", path.file =
"pathlist_expr_Levy2012.txt", out.file = "expression_data_validation",
species = KEGG.species.homo, pvalue = 0.05)
Input file imported!
Pathway input file imported!
Gene list specified... and correct!
Your path list colnames are correct!
Species code specified... and correct!
Pathway(s) is specified... and correct!
Prerequisite check passed!
PIA is running ...
Please wait... It could be a while depending on the number of pathways
required!
n. 18 of 30 genes passed the p-value filtering.
Well done! Diagram visualization was created and exported.
```

This PIA visualization (Figure IV-7) can help to interpret the results obtained from gene expression experiments by showing the nodes (i.e. genes) coloured according to gene FC values. The diagram showed the relationships among genes and pathways and allowed us to identify functionally related entities with possibly coordinated expression changes.

IV - 5. Conclusion

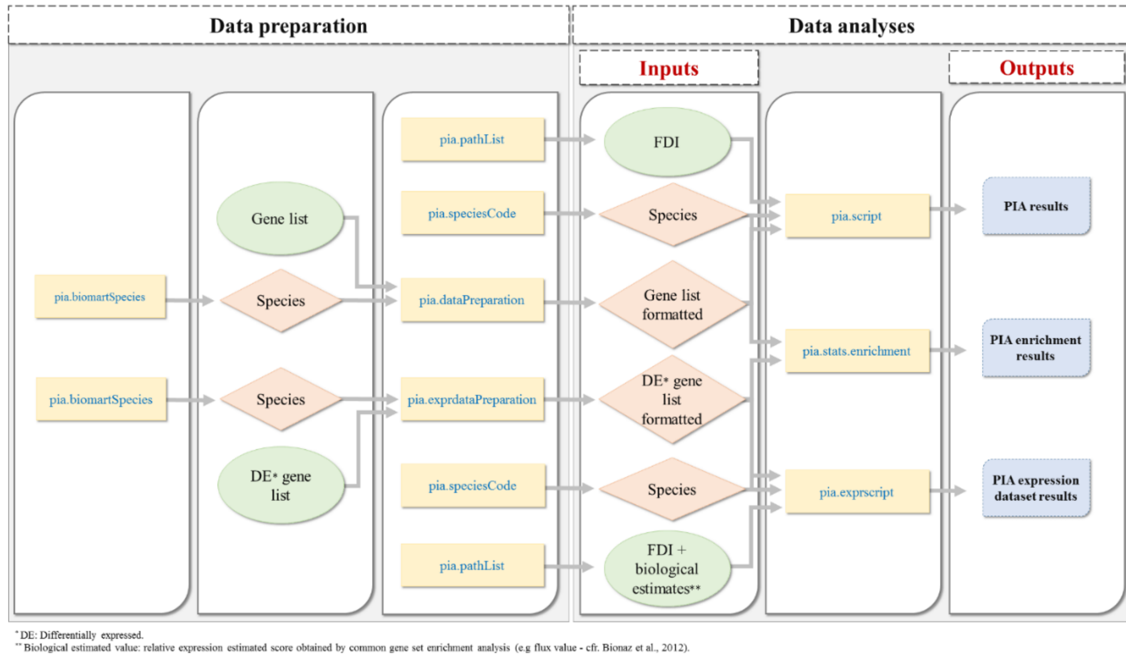
PIA represents a novel and useful approach to reduce the complexity of high-throughput data mining challenges and for candidate gene identification. PIA allows us to overcome the limitations of classical functional enrichment analysis providing network-based information among pathways and genes, and helping with the interpretation of genomic and transcriptomic analysis outcomes.

PIA is a package entirely built in R. The contribution of PIA in high-throughput data mining could be significant not only for well-documented species (i.e. *homo sapiens*), but also for less-annotated ones. PIA can work with all the species available in KEGG databases (more than 5,000 organisms). Although KEGG is a popular database for biological network information, the lack of gene or pathway information on the KEGG database could represent the main disadvantages of PIA.

The effectiveness of the PIA approach in terms result coherency was confirmed by the reference study validation. Ultimately, PIA produces time-saving advantages, creating a bibliographic list of genes that are biologically-involved with the trait investigated.

IV - 6. Figures and tables

Figure IV-1. The general architecture of the workflow of PIA package and schematic illustration of main functions.



* DE: Differentially expressed.
** Biological estimated value: relative expression estimated score obtained by common gene set enrichment analysis (e.g flux value - cfr. Bionaz et al., 2012).

Figure IV-2. An example of the '.html' file with the network-based visualization of PIA results. The green circles represent the candidate genes falling inside the pathways associated with the trait of interest or resulted in interaction. The violet rectangles represent the first-degree interaction (FDI or 1DI) pathways, directly connected to the trait of interest and showing candidate gene(s). The yellow rectangles represent the second degree (2DI) of pathways in interaction with FDI pathways and showing candidate gene(s). The orange rectangles represent the pathways investigated without any candidate gene.

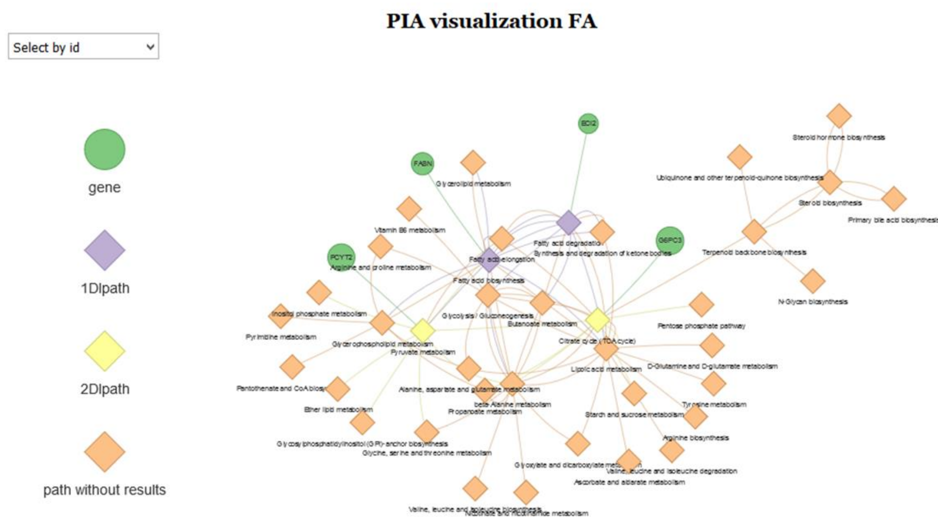


Figure IV-3. An example of node selection of PIA network-based visualization result.

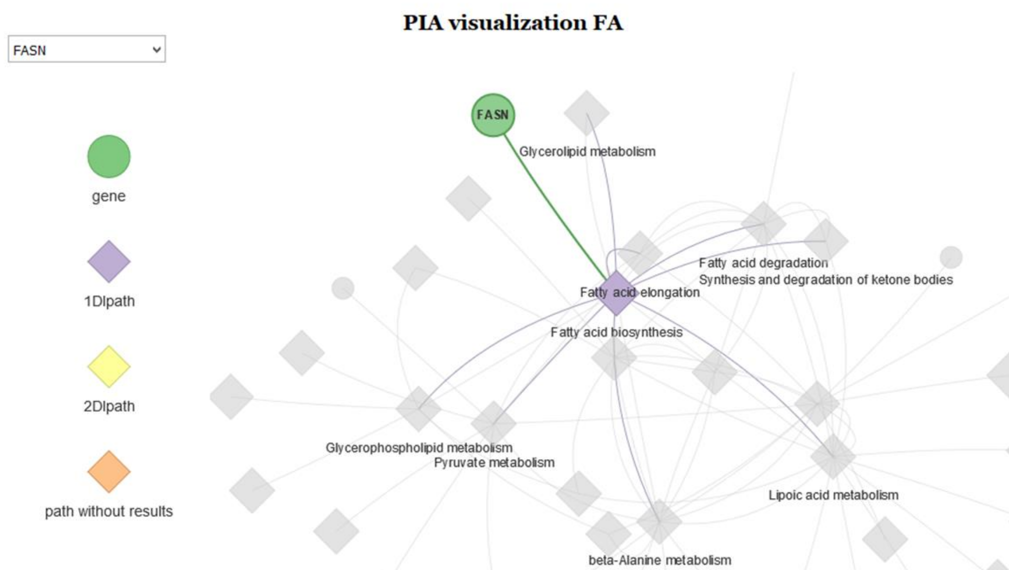


Figure IV-4. An example of the '.html' file with the network-based visualization of PIA result considering an expression dataset. The circles represent the genes coloured based on their fold change (FC) values. The rectangles represent the pathways of interest coloured based on their expression estimated scores (i.e. flux values obtained with Dynamic Impact Approach (Bionaz et al., 2012c)).

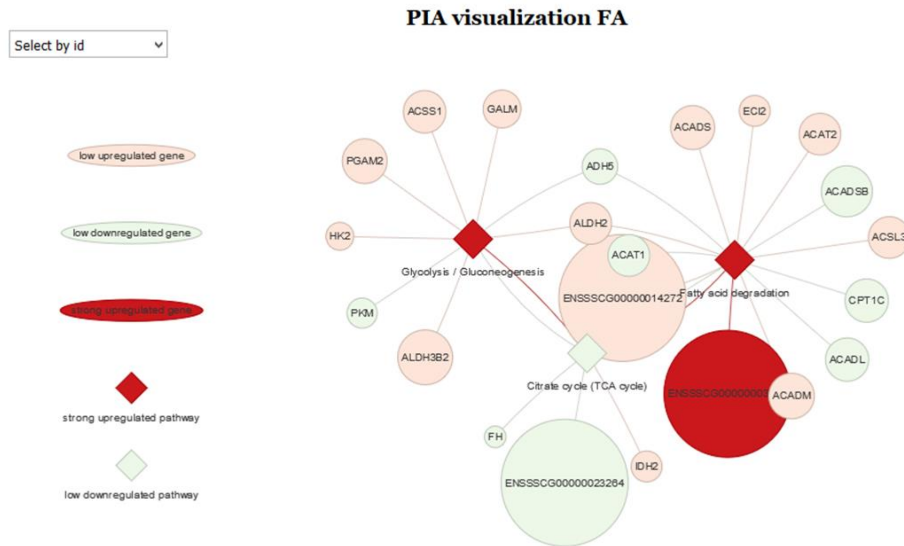


Figure IV-5. Node selection of PIA network-based visualization result obtained on expression dataset.

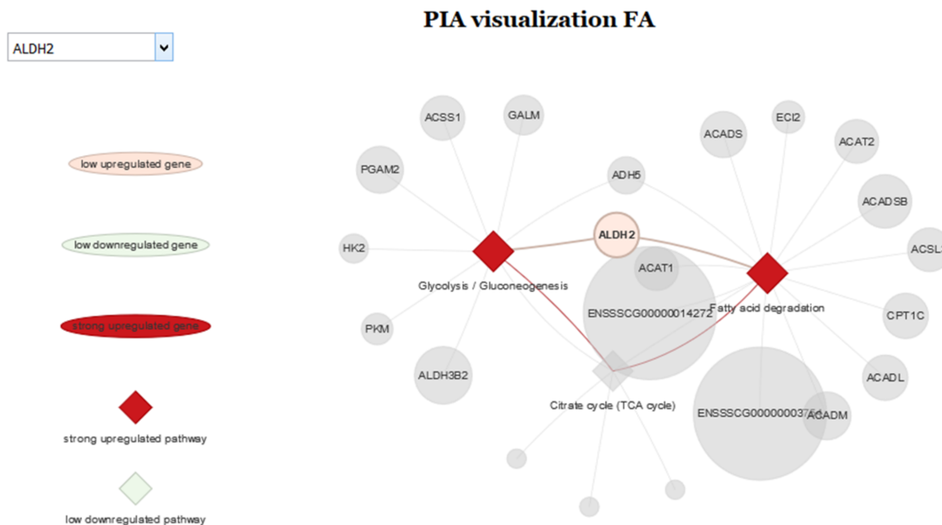


Figure IV-6. Network-based visualization of result obtained by PIA considering three interaction levels. The green circles represent the functional candidate genes falling inside the pathways associated with the trait of interest and/or resulted in interaction. The violet rectangles represent the first-degree (FDI or 1DI) interaction pathway, directly connected to the trait of interest (i.e. Type I diabetes mellitus', 'Insulin resistance', and 'AGE-RAGE signaling pathway in diabetic complications'). The yellow and blue rectangles represent the second (2DI) and third (3DI) pathways in interaction with FDI pathways, highlighted by the PIA and containing the gene of interest. The orange rectangles represent the pathways investigated showing no results.

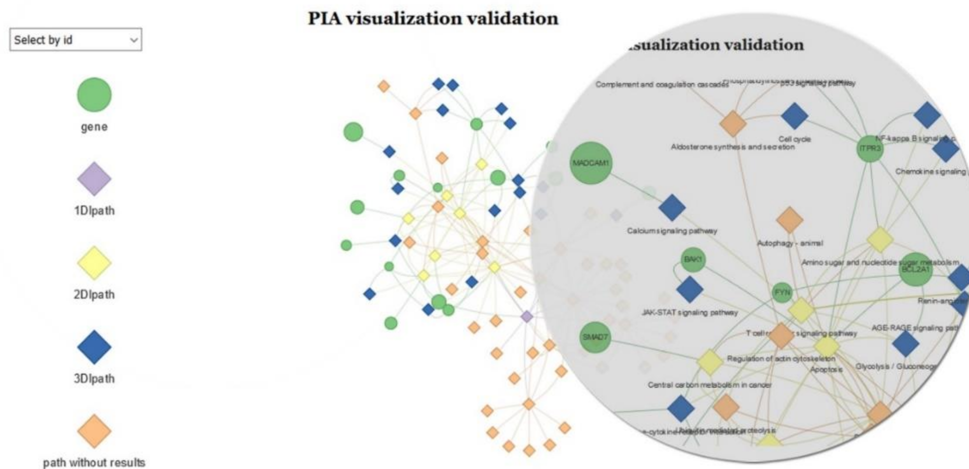


Figure IV-7. Network-based visualization of PIA result considering expression dataset (Levy et al., 2012; Qiu et al., 2014). The circles represent the genes coloured based on their fold change (FC) values. The rectangles represent the pathways of interest (i.e. Type I diabetes mellitus', 'Insulin resistance', and 'AGE-RAGE signaling pathway in diabetic complications') coloured based on their expression estimated scores (i.e. in our case genes occurrences).

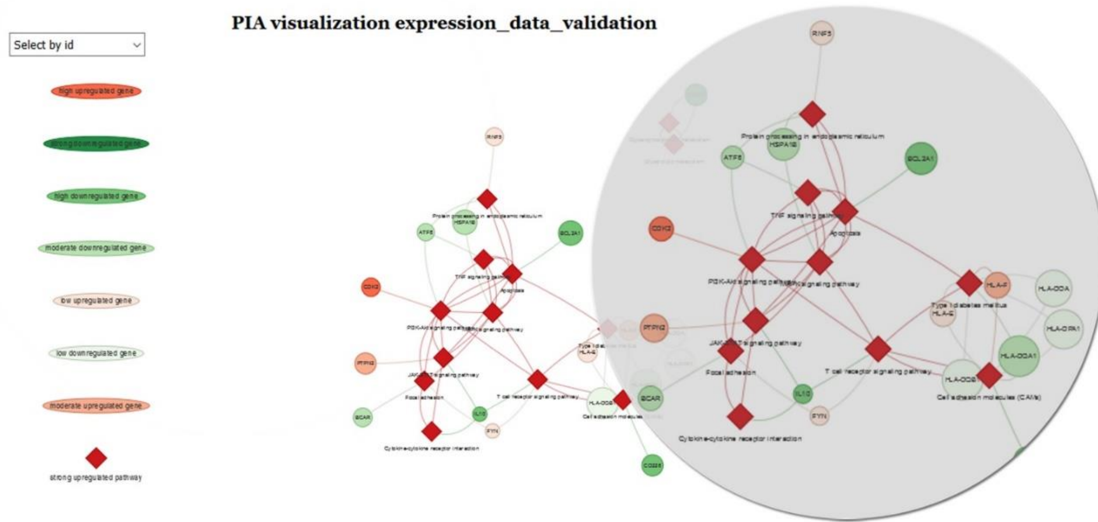


Table IV-1. Summary of node colour classification in network visualization obtained with `pia.exprscript()` function.

gene/pathway classification	with FC/estimated score value
low upregulated/downregulated	<25% of top up/downregulated gene/pathway value
moderate upregulated/downregulated	≥25% and < 50% of top up/downregulated gene/pathway value
high upregulated/downregulated	≥ 50% and < 75% of top up/downregulated gene/pathway value
strong upregulated/downregulated	≥ 75% of top up/downregulated gene/pathway value

Table IV-2. Summary of PIA results for three interaction levels, considering 'Type I diabetes mellitus', 'Insulin resistance', and 'AGE-RAGE signaling pathway in diabetic complications' as FDI or 1DI pathways.

PIA Interaction degree	*KEGG pathway*	*Gene*
1DI	Insulin resistance	*PTPN11*
2DI	PI3K-Akt signaling pathway	*CDK2*
	Apoptosis	*ITPR3, BAK1, BCL2A1*
	T cell receptor signaling pathway	*FYN, IL10*
	Calcium signaling pathway	*ITPR3*
	Jak-STAT signaling pathway	*STAT4, IL10, PTPN11*
	Cell cycle	*CDK2*
	TGF-beta signaling pathway	*SMAD7*
	Adipocytokine signaling pathway	*RXRB, PTPN11*
3DI	Chemokine signaling pathway	*BCAR1*
	NF-kappa B signaling pathway	*BCL2A1*
	FoxO signaling pathway	*CDK2, IL10*
	Phosphatidylinositol signaling system	*ITPR3*
	Cytokine-cytokine receptor interaction	*IL10*
	p53 signaling pathway	*CDK2*
	Autophagy - animal	*HMGB1*
	Protein processing in endoplasmic reticulum	*BAK1*
	Focal adhesion	*BCAR1, FYN, MYL2*
	Cell adhesion molecules (CAMs)	*MADCAM1*

Vascular smooth muscle contraction	*ITPR3*
Natural killer cell mediated cytotoxicity	*MICA, FYN, PTPN11*
Long-term potentiation	*ITPR3*
Long-term depression	*ITPR3*
Renin secretion	*ITPR3*
Aldosterone synthesis and secretion	*ITPR3*
Regulation of actin cytoskeleton	*MYL2, BCAR1*

Table IV-3. Comparison between PIA and reference study results (Qiu et al., 2014).

Genes* highlighted by PIA and consistent with main result in Qui et al. (Qiu et al., 2014) ***study***
PTPN11, BCAR1, MYL2 and *FYN*

* among the 7 genes validated both in replication and differential expression studies (Qiu et al., 2014)

Genes* highlighted by PIA, reported in literature as being associated to the susceptibility to T1DM disease, but not discussed in Qui et al. (Qiu et al., 2014)
ITPR3 (Qu et al., 2008), *BAK1* (Qiu et al., 2014), *IL10* (Hong et al., 2009), *HMGB1* (Zhang et al., 2010), *MICA* (Park et al., 2001), *CDK2* (Kim et al., 2017b), *RXRB* (Shi et al., 2016b), *MADCAM1* (Phillips et al., 2005b), *STAT4* (Bi et al., 2013b), *BCL2A1* (Beyan et al., 2010b), *SMAD7* (Chen et al., 2011)

* among the 23 genes validated in replication studies or 37 in differential expression studies (Qiu et al., 2014)

Genes* not highlighted by PIA, since falling inside no-investigated pathways
BRAP, *FUT2, GNS, HIPK1, NUPR1, OR2B3, HIST1H4E, HIST1H2BF, OR2B3, OR2B6, OR2J2, OR5V1, SULT1A1*

* among 166 out of 171 newly genes in reference study (Qiu et al., 2014)

Genes* detected by Qui et al. (Qiu et al., 2014)***, but not highlighted by PIA since not yet assigned to any KEGG pathways***
ADAD1, *ASCL2*, *ATF7IP*, BTN3A3, C6orf227, *CABP1, CCDC101, CEACAM7, CRYZL1, DEXI, ETF1P1, FAM46B, FAP, GCA, GGNBP1, GNL1, GP2, GUSBL1, HIST1H1A, HIST1H1T, HIST1H2BD, HIST1H3H, HIST1H4F, HIST1H4G, HIST1H4PS1, HLA-J, HORMAD2, IKZF1, IKZF3, KIAA0528, KIFC1, KRT222, LHX9, LOC144481, LOC284749, MAMSTR, MICG, MIR548H3, MIR600, MPZL3, NCAPD2, NSL1, OLFML3, OR12D1P, ORMDL3, PHF1, PLBD1, PLEKHA1, PPP1R10, PPP1R11, PRR3, PRSS16, RASIP1, RING1, SBK1, SCGN, SLC17A1, SLC17A2, SLC17A3, SLC17A4, SPRR2E, STRN4, TAPBPL, TMEM129, TMEM170A, VPS52, ZBTB9, ZNF192, ZNF274, ZNF322A, ZNF323, ZPBP2, ZZEF1*

* among 166 out of 171 newly genes in reference study (Qiu et al., 2014)

Genes* detected by Qui et al. (Qiu et al., 2014)***, but not highlighted by PIA since had no corresponding gene in KEGG databases***
GPR89P, HCG2P8, HCG4P3, HCG4P4, HCG4P9, HCGVIII-2, HCP5P2, LOC100127934, LOC100128077, LOC100128588, LOC100129387, LOC100130535, LOC100133214, LOC100270746, LOC100288130, LOC100506705, LOC100506979, LOC100507085, LOC340192, LOC402641, LYPLA2P1, OR2E1P, OR2U1P, OR2W6P, RPLP2P1, RPS10P1, RSPH1, SUMO2P, TRAJ57, TRAJ58, TRDD1, TRDD2, TRIM26, TRIM27, TRMEP1, TRNAA12, TRNAA19, TRNAA38, TRNAA40, TRNAA41, TRNAA5, TRNAF3, TRNAI1, TRNAI2, TRNAK43P, TRNAK8, TRNAL12, TRNAL47P, TRNAM15, TRNAM16, TRNAM4, TRNAM8, TRNAR10, TRNAS7, TRNAT11, TRNAT16, TRNAT7, TRNAV15, TRNAV27, TRNAV7, TRNAW2, TRNAY7, TRNAY8, UBD, VN1R14P

* among 166 out of 171 newly genes in reference study (Qiu et al., 2014)

IV - 7. References

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25–29. doi:10.1038/75556.

Beissbarth, T., and T.P. Speed. 2004. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinforma. Oxf. Engl. 20:1464–1465. doi:10.1093/bioinformatics/bth088.

Berg, B.H. van den, C. Thanthiriwatte, P. Manda, and S.M. Bridges. 2009. Comparing gene annotation enrichment tools for functional modeling of agricultural microarray data. BMC Bioinformatics 10:S9. doi:10.1186/1471-2105-10-S11-S9.

Beyan, H., R.C. Drexhage, L. van der Heul Nieuwenhuijsen, H. de Wit, R.C. Padmos, N.C. Schloot, H.A. Drexhage, and R.D. Leslie. 2010a. Monocyte gene-expression profiles associated with childhood-onset type 1 diabetes and disease risk: a study of identical twins. Diabetes 59:1751–1755. doi:10.2337/db09-1433.

Beyan, H., R.C. Drexhage, L. van der Heul Nieuwenhuijsen, H. de Wit, R.C. Padmos, N.C. Schloot, H.A. Drexhage, and R.D. Leslie. 2010b. Monocyte gene-expression profiles associated with childhood-onset type 1 diabetes and disease risk: a study of identical twins. Diabetes 59:1751–1755. doi:10.2337/db09-1433.

Bi, C., B. Li, Z. Cheng, Y. Hu, Z. Fang, and A. Zhai. 2013. Association study of STAT4 polymorphisms and type 1 diabetes in Northeastern Chinese Han population. Tissue Antigens 81:137–140. doi:10.1111/tan.12057.

Bionaz, M., K. Periasamy, S.L. Rodriguez-Zas, W.L. Hurley, and J.J. Loor. 2012. A novel dynamic impact approach (DIA) for functional analysis of time-course omics studies: validation

using the bovine mammary transcriptome. PloS One 7:e32455. doi:10.1371/journal.pone.0032455.

Chen, H.Y., X.R. Huang, W. Wang, J.H. Li, R.L. Heuchel, A.C.K. Chung, and H.Y. Lan. 2011. The protective role of Smad7 in diabetic kidney disease: mechanism and therapeutic potential. Diabetes 60:590–601. doi:10.2337/db10-0403.

Cirillo, E., L.D. Parnell, and C.T. Evelo. 2017. A review of pathway-based analysis tools that visualize genetic variants. Front. Genet. 8. doi:10.3389/fgene.2017.00174.

Curtis, R.K., M. Oresic, and A. Vidal-Puig. 2005. Pathways to the analysis of microarray data. Trends Biotechnol. 23:429–435. doi:10.1016/j.tibtech.2005.05.011.

Draghici, S., P. Khatri, R.P. Martins, G.C. Ostermeier, and S.A. Krawetz. 2003. Global functional profiling of gene expression. Genomics 81:98–104.

Durinck, S., P.T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4:1184–1191. doi:10.1038/nprot.2009.97.

Field, L.L., and R. Tobias. 1997. Unravelling a complex trait: the genetics of insulin-dependent diabetes mellitus. Clin. Investig. Med. Med. Clin. Exp. 20:41–49.

Greenbaum, C.J. 2002. Insulin resistance in type 1 diabetes. Diabetes Metab. Res. Rev. 18:192–200. doi:10.1002/dmrr.291.

Hong, E.-G., H.J. Ko, Y.-R. Cho, H.-J. Kim, Z. Ma, T.Y. Yu, R.H. Friedline, E. Kurt-Jones, R. Finberg, M.A. Fischer, E.L. Granger, C.C. Norbury, S.D. Hauschka, W.M. Philbrick, C.-G. Lee, J.A. Elias, and J.K. Kim. 2009. Interleukin-10 prevents diet-induced insulin resistance by attenuating macrophage and cytokine response in skeletal muscle. Diabetes 58:2525–2535. doi:10.2337/db08-1261.

Huang, D.W., B.T. Sherman, and R.A. Lempicki. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:1–13. doi:10.1093/nar/gkn923.

Ihaka, R., and R. Gentleman. 1996. R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. 5:299–314. doi:10.2307/1390807.

Joyce, A.R., and B.Ø. Palsson. 2006. The model organism as a system: integrating "omics" data sets. Nat. Rev. Mol. Cell Biol. 7:198–210. doi:10.1038/nrm1857.

Kanehisa, M. 2016. KEGG Bioinformatics Resource for Plant Genomics and Metabolomics.. Methods Mol. Biol. Clifton NJ 1374:55–70. doi:10.1007/978-1-4939-3167-5_3.

Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45:D353–D361. doi:10.1093/nar/gkw1092.

Khatri, P., and S. Drăghici. 2005. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinforma. Oxf. Engl. 21:3587–3595. doi:10.1093/bioinformatics/bti565.

Khatri, P., M. Sirota, and A.J. Butte. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput. Biol. 8:e1002375. doi:10.1371/journal.pcbi.1002375.

Kim, S.Y., J.-H. Lee, M.J. Merrins, O. Gavrilova, X. Bisteau, P. Kaldis, L.S. Satin, and S.G. Rane. 2017. Loss of cyclin dependent kinase 2 in the pancreas links primary β-cell dysfunction to progressive depletion of β-cell mass and diabetes. J. Biol. Chem. jbc.M116.754077. doi:10.1074/jbc.M116.754077.

Levy, H., X. Wang, M. Kaldunski, S. Jia, J. Kramer, S.J. Pavletich, M. Reske, T. Gessel, M. Yassai, M.W. Quasney, M.K. Dahmer, J. Gorski, and M.J. Hessner. 2012. Transcriptional

signatures as a disease-specific and predictive inflammatory biomarker for type 1 diabetes. Genes Immun. 13:593–604. doi:10.1038/gene.2012.41.

Palombo, V., M. Milanesi, S. Sgorlon, S. Capomaccio, M. Mele, E. Nicolazzi, P. Ajmone-Marsan, F. Pilla, B. Stefanon, and M. D'Andrea. 2018. Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. J. Dairy Sci. doi:10.3168/jds.2018-14413.

Park, Y., H. Lee, C.B. Sanjeevi, and G.S. Eisenbarth. 2001. MICA polymorphism is associated with type 1 diabetes in the Korean population. Diabetes Care 24:33–38.

Phillips, J.M., K. Haskins, and A. Cooke. 2005. MAdCAM-1 is needed for diabetes development mediated by the T cell clone, BDC-2·5. Immunology 116:525–531. doi:10.1111/j.1365-2567.2005.02254.x.

Qiu, Y.-H., F.-Y. Deng, M.-J. Li, and S.-F. Lei. 2014. Identification of novel risk genes associated with type 1 diabetes mellitus using a genome-wide gene-based association analysis. J. Diabetes Investig. 5:649–656. doi:10.1111/jdi.12228.

Qu, H.-Q., L. Marchand, A. Szymborski, R. Grabs, and C. Polychronakos. 2008. The association between type 1 diabetes and the ITPR3 gene polymorphism due to linkage disequilibrium with HLA class II. Genes Immun. 9:264–266. doi:10.1038/gene.2008.12.

Ramasamy, R., S.J. Vannucci, S.S.D. Yan, K. Herold, S.F. Yan, and A.M. Schmidt. 2005. Advanced glycation end products and RAGE: a common thread in aging, diabetes, neurodegeneration, and inflammation. Glycobiology 15:16R-28R. doi:10.1093/glycob/cwi053.

Rk, C., O. M, and V.-P. A. 2005. Pathways to the analysis of microarray data. Trends Biotechnol. 23:429–435. doi:10.1016/j.tibtech.2005.05.011.

Shi, A., J. Wen, G. Liu, H. Liu, Z. Fu, J. Zhou, Y. Zhu, Y. Liu, X. Guo, and J. Xu. 2016. Genetic variants in vitamin D signaling pathways and risk of gestational diabetes mellitus. Oncotarget 7:67788–67795. doi:10.18632/oncotarget.11984.

Simoes, R. de M., and F. Emmert-Streib. 2012. Bagging statistical network inference from large-scale gene expression data. PLOS ONE 7:e33624. doi:10.1371/journal.pone.0033624.

Tavazoie, S., J.D. Hughes, M.J. Campbell, R.J. Cho, and G.M. Church. 1999. Systematic determination of genetic network architecture. Nat. Genet. 22:281–285. doi:10.1038/10343.

Zhang, S., J. Zhong, P. Yang, F. Gong, and C.-Y. Wang. 2010. HMGB1, an innate alarmin, in the pathogenesis of type 1 diabetes. Int. J. Clin. Exp. Pathol. 3:24–38.

Chapter V - GENERAL CONCLUSION

Since the beginning of domestication, livestock species were selected to fulfill human needs. Breed creation intensified the differentiation between animal populations and promoted the development of intensive selection schemes. Most of the animal breeding theory we are still using today, was developed in the first half of the 20th century, when innovative statistical techniques (i.e. best linear unbiased prediction - BLUP) were used to select various traits that optimize animals performance and select those with most optimal combinations, i.e. estimation of breeding value (EBV) (Henderson, 1984). Subsequently, the increasing availability of DNA information provided promising opportunities to enhance animal breeding theory. In particular, new advances in animal genotyping fostered the development of marker-assisted selection (MAS) (Dekkers, 2004) and more recently of so-called genomic selection (GS) (Meuwissen, 2007). All of this is producing positive genetic trends in many productive traits, particularly increasing reliability of genomic EBV (GEBV) compared with parent average estimates (Hayes et al., 2009). Despite the fact that advances have resulted in more accurate selection results and a faster genetic improvement across generations, much more is expected (Hayes et al., 2009). In fact, we still have a poor knowledge about gene biology of phenotypes under selection. A deeper understanding of animal genome organization and information would further increase the accuracy of genomic evaluation by incorporating prior knowledge. In this regard, it is expected that the new and revolutionary advent of high-throughput 'omics' (HTO) technologies has the capability to spearhead the progress of systems biology, including applications on animal production and health traits (Suravajhala et al., 2016).

In thesis the biology of livestock complex traits, such as lipid metabolism and colostrogenesis/lactogenesis transition respectively in bovine and pig species, has been investigated using state-of-art genomic and transcriptomic analyses. In particular, these goals have been achieved by complementary approaches and different methods. In chapter 2, a genome-wide association study (GWAS) on bovine milk was performed with the aim to identify genomic regions or genes associated with fatty acids (FA) profile and investigate genetic differences between Italian Simmental (IS) and Italian Holstein (IH) breeds. Along with single-SNP GWAS, an innovative post-GWAS pipeline was applied. It was mainly based on a gene-centric association (Capomaccio et al., 2015) and pathways interaction investigation (see chapter 4) approaches. This helped us to dissect and prioritize the GWAS association signals with the aim of finding candidate genes affecting breed-specific FA composition. In particular,

according to previous results reported in literature, the effects of well-established genes associated with milk fat yield and content were confirmed by our study. Moreover, other possible candidate genes, several of them directly or indirectly involved in 'Lipid Metabolism', were also identified. Overall, the findings improve our understanding of genetic architecture in IS and IH cows and reflect breed-specific genomic features. The differences are explainable by the different productive characteristics and divergent selective breeding history of two breeds under study and represent further molecular information useful in breeding programs.

In chapter 3, an RNA sequencing analysis was performed on sow mammary gland from 14 days prior to parturition to day 1 in lactation to provide a comprehensive transcriptome profiling to better elucidate the biological mechanism of swine colostrogenesis/lactogenesis. This transition step plays a key role in piglet survival and growth, which represents a major problem especially in modern pig production where piglet mortality is high during the first days of life (Theil et al., 2014). In fact, colostrum and transient milk are pivotal sources of antibodies, energy and nutrients for any neonate and especially for piglets that are deficient in fat depots and brown adipose tissue (Salmon, 2000; Pastorelli et al., 2009). With the goal of highlighting the differentially expressed genes (DEG) among the different time points under study, we applied a well-established post-sequencing analysis pipeline based on the edgeR (Robinson et al., 2010) and limma-voom (Law et al., 2014) methods. Furthermore, functional bioinformatics tools such as the Dynamic Impact Approach (DIA) (Bionaz et al., 2012) and Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA) were used for pathway analysis and to identify transcription regulators and their networks. This study produced a huge amount of genomic information that provided us a better understanding of metabolic and signalling pathways involved in the sow peripartum period. In fact, although the precise timing for the transition from colostrogenesis to lactogenesis in swine remains unclear, our data supported the hypothesis that the transition occurs before parturition. This is likely attributable to upregulation of a wide array of genes along with the activation of transcription regulators controlling lipid synthesis and endoplasmic reticulum biogenesis and stress response. In summary, the transcriptome changes greatly between 6 and 2 days prepartum and these changes are highly likely to be involved in coordinating the synthesis of colostrum and main milk components (i.e. protein, fat, lactose and antimicrobial factors) as revealed by influenced pathways.

In addition to the objectives and related studies summarized above, this thesis also introduced an in-house bioinformatics tool performing a new pathway analysis useful for post-genomic

and -transcriptomic data mining. In chapter 4, the PIA (Pathway Interaction Analysis) R package was described. The basic idea of PIA is to implement a pathway analysis taking into account a network of both upstream and downstream pathways in interaction. Pathway analysis (PA) is commonly applied for the secondary analysis of high-throughput experiments (Khatri and Drăghici, 2005). Nevertheless, the majority of PA tools freely available performs analysis within a single pathway (Curtis et al., 2005). This represent a clear limitation in ‘omic’ research that has its strength in a holistic approach. Specifically, focus on a single pathways of interest may reduce the information obtainable by the modern HT platforms. We believe that PIA can help to interpret HT data and identify candidate genes for function that can influence multiple and complex biological processes. In particular, we think that PIA data visualization, in the form of interactive pathway diagrams and gene-pathway biological interactions, can enhance interpretation of scientific data, increase understanding of the conclusions drawn, and promote discussion and follow-up research questions. In this regard, PIA was validated using a publicly available dataset on human type 1 diabetes mellitus (Qiu et al., 2014) and showed remarkable advantages in terms of effectiveness and time-saving.

In conclusion, with the general aim of providing new genetic information for animal breeding, this thesis has explored the possibilities offered by HTO technologies in the genomic and transcriptomic field, such as High Density genotyping and Next Generation Sequencing, with established and innovative bioinformatics procedures. The single-research results were significant and more broadly they dimostrated that the omic approach represents the gold standard method to give insight into the most complex biological mechanisms. In this regard, omic data analysis represents a revolutionary gain that increasingly depend on researchers capable of creating and implementing effective and integrative pipelines that comprise integrated (multi)omics approaches instead of distinct and monothematic ones (Suravajhala et al., 2016; Manzoni et al., 2018). This clearly requires the cooperation of multidisciplinary teams. It is early days yet, but what is certain is that we finally have the great opportunity to pinpoint key elements of biological questions that would have been impossible decades ago.

V - 1. References

Bionaz, M., K. Periasamy, S.L. Rodriguez-Zas, W.L. Hurley, and J.J. Loor. 2012. A novel dynamic impact approach (DIA) for functional analysis of time-course omics studies: validation using the bovine mammary transcriptome. PloS One 7:e32455. doi:10.1371/journal.pone.0032455.

Capomaccio, S., M. Milanesi, L. Bomba, E. Vajana, and P. Ajmone-Marsan. 2015. MUGBAS: a species free gene-based programme suite for post-GWAS analysis. Bioinforma. Oxf. Engl. 31:2380–2381. doi:10.1093/bioinformatics/btv144.

Curtis, R.K., M. Oresic, and A. Vidal-Puig. 2005. Pathways to the analysis of microarray data. Trends Biotechnol. 23:429–435. doi:10.1016/j.tibtech.2005.05.011.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82 E-Suppl:E313-328. doi:10.2527/2004.8213_supplE313x.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646.

Henderson, C.R. 1984. Applications of linear models in animal breeding. Appl. Linear Models Anim. Breed.

Khatri, P., and S. Drăghici. 2005. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinforma. Oxf. Engl. 21:3587–3595. doi:10.1093/bioinformatics/bti565.

Law, C.W., Y. Chen, W. Shi, and G.K. Smyth. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15:R29. doi:10.1186/gb-2014-15-2-r29.

Manzoni, C., D.A. Kia, J. Vandrovcova, J. Hardy, N.W. Wood, P.A. Lewis, and R. Ferrari. 2018. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief. Bioinform. 19:286–302. doi:10.1093/bib/bbw114.

Meuwissen, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie 124:321–322. doi:10.1111/j.1439-0388.2007.00708.x.

Palombo, V., M. Milanesi, S. Sgorlon, S. Capomaccio, M. Mele, E. Nicolazzi, P. Ajmone-Marsan, F. Pilla, B. Stefanon, and M. D'Andrea. 2018. Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. J. Dairy Sci. doi:10.3168/jds.2018-14413.

Pastorelli, G., M. Neil, and I. Wigren. 2009. Body composition and muscle glycogen contents of piglets of sows fed diets differing in fatty acids profile and contents. Livest. Sci. 123:329–334. doi:10.1016/j.livsci.2008.11.023.

Qiu, Y.-H., F.-Y. Deng, M.-J. Li, and S.-F. Lei. 2014. Identification of novel risk genes associated with type 1 diabetes mellitus using a genome-wide gene-based association analysis. J. Diabetes Investig. 5:649–656. doi:10.1111/jdi.12228.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. 26:139–140. doi:10.1093/bioinformatics/btp616.

Salmon, H. 2000. Mammary gland immunology and neonate protection in pigs. Homing of lymphocytes into the MG. Adv. Exp. Med. Biol. 480:279–286. doi:10.1007/0-306-46832-8_32.

Suravajhala, P., L.J.A. Kogelman, and H.N. Kadarmideen. 2016. Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. Genet. Sel. Evol. GSE 48. doi:10.1186/s12711-016-0217-x.

Theil, P.K., C. Lauridsen, and H. Quesnel. 2014. Neonatal piglet survival: impact of sow nutrition around parturition on fetal glycogen deposition and production and composition of colostrum and transient milk. Animal 8:1021–1030. doi:10.1017/S1751731114000950.

Vailati-Riboni, M., Palombo V., and Loor J.J. 2017. What are omics sciences? Periparturient Diseases of Dairy Cows: A Systems Biology Approach pp. 1-7.

SUPPLEMENTARY MATERIAL

Supplementary Table II-S1. Significant genes (i.e. FDR q-values less than or equal to 0.05) obtained with MUGBAS analysis and associated with milk fatty acid (FA) traits in Italian Simmental (IS) and Italian Holstein (IH). Ensembl gene id, false discovery rate (FDR) q-value statistics, gene symbol, chromosome and genome location are reported for each gene. SNP name, GWA p-value and genome location are reported for each best SNP.

breed	trait	Gene						Best SNP			
		Ensembl gene id	FDR q-value	Gene symbol	Chr	Start	End	BestSNP	BestSNP p-value	BestSNP Location	Candidate
IH	FA14	ENSBTAG00000007123	0,02384	ENSA	3	20141694	20146997	BovineHD0300006393	0,00001	20151200	BestCandidate
IH	FA14	ENSBTAG00000015154	0,02384	NA	3	20172325	20176960	BovineHD0300006393	0,00001	20151200	PossibleFalsePositive
IH	FA14	ENSBTAG00000020338	0,04291	CCT6B	19	15448008	15483845	UA-IFASA-7088	0,00008	15506868	BestCandidate
IH	FA14	ENSBTAG00000020357	0,01192	GOLPH3L	3	20082069	20126434	BovineHD0300006393	0,00001	20151200	PossibleFalsePositive
IH	FA14	ENSBTAG00000044530	0,00000	SNORA70	19	15745579	15745700	BovineHD1900004261	0,00002	15725123	BestCandidate
IH	FA15	ENSBTAG00000012225	0,00000	KPNA2	19	49569375	49579204	BovineHD1900013834	0,00007	49530233	BestCandidate
IH	FA16	ENSBTAG00000014358	0,03179	EVA1B	3	110118545	110120286	ARS-BFGL-NGS-102149	0,00003	110078547	BestCandidate
IH	FA16	ENSBTAG00000024097	0,03179	NA	3	110014938	110020632	ARS-BFGL-NGS-102149	0,00003	110078547	PossibleFalsePositive
IH	FA16	ENSBTAG00000038617	0,03179	SH3D21	3	110121046	110134882	ARS-BFGL-NGS-102149	0,00003	110078547	PossibleFalsePositive
IH	FA18	ENSBTAG00000000090	0,00000	SEC31B	26	21248525	21280208	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA18	ENSBTAG00000000091	0,00000	NDUFB8	26	21283406	21288806	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA18	ENSBTAG00000000092	0,00000	HIF1AN	26	21291076	21300157	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA18	ENSBTAG00000001017	0,01445	SLK	26	24786360	24844579	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA18	ENSBTAG00000002010	0,04865	PI4K2A	26	18684202	18710907	BovineHD2600004833	0,00001	18761989	PossibleFalsePositive
IH	FA18	ENSBTAG00000002261	0,00000	LBX1	26	21894402	21896298	BovineHD2600005648	0,00000	21926490	BestCandidate
IH	FA18	ENSBTAG00000002430	0,01335	COL17A1	26	24848931	24896345	BovineHD2600006436	0,00000	24918578	BestCandidate
IH	FA18	ENSBTAG00000002880	0,01445	SORBS1	26	16721039	16919269	Hapmap54846-rs29022328	0,00035	16753151	BestCandidate
IH	FA18	ENSBTAG00000003294	0,04865	MRPL43	26	21685254	21692758	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA18	ENSBTAG00000003296	0,04865	C10orf2	26	21692833	21698205	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA18	ENSBTAG00000003298	0,04865	LZTS2	26	21703370	21710503	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA18	ENSBTAG00000004612	0,01093	SORCS3	26	26294032	26397191	BTA-61038-no-rs	0,00001	26267018	BestCandidate
IH	FA18	ENSBTAG00000007243	0,02577	CFAP58	26	25185701	25286776	UA-IFASA-4715	0,00014	25314352	BestCandidate
IH	FA18	ENSBTAG00000007476	0,00000	BTRC	26	22001775	22172725	BTB-00932332	0,00000	22118554	BestCandidate
IH	FA18	ENSBTAG00000007591	0,04865	CHUK	26	20966010	21008277	Hapmap31825-BTA-158647	0,00002	21056547	PossibleFalsePositive
IH	FA18	ENSBTAG00000007594	0,01445	CWF19L1	26	21010622	21035035	Hapmap31825-BTA-158647	0,00002	21056547	PossibleFalsePositive
IH	FA18	ENSBTAG00000007948	0,01445	SORCS1	26	27810198	28389391	BovineHD2600007498	0,00000	28146982	BestCandidate
IH	FA18	ENSBTAG00000008100	0,02577	GOLGA7B	26	18858709	18871716	BovineHD2600004851	0,00009	18820468	BestCandidate
IH	FA18	ENSBTAG00000008102	0,00681	CRTAC1	26	18869719	19013761	BovineHD2600004938	0,00004	19015156	BestCandidate
IH	FA18	ENSBTAG00000010739	0,01445	BLOC1S2	26	21043927	21053587	Hapmap31825-BTA-158647	0,00002	21056547	PossibleFalsePositive
IH	FA18	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive

IH	FA18	ENSBTAG00000012077	0,00000	SLF2	26	21630816	21670831	BovineHD2600005581	0,00000	21564772	BestCandidate
IH	FA18	ENSBTAG00000012107	0,00000	SLC25A28	26	20466104	20476402	BovineHD2600005302	0,00000	20463679	BestCandidate
IH	FA18	ENSBTAG00000012857	0,01036	CNNM2	26	23795436	23972895	BovineHD2600006134	0,00000	23847594	BestCandidate
IH	FA18	ENSBTAG00000012914	0,00000	WNT8B	26	21240570	21244552	BTB-00931481	0,00000	21226405	BestCandidate
IH	FA18	ENSBTAG00000014574	0,01036	CFAP43	26	24953967	25055692	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA18	ENSBTAG00000018566	0,04865	SFRP5	26	18782564	18787376	BovineHD2600004833	0,00001	18761989	BestCandidate
IH	FA18	ENSBTAG00000018604	0,04865	SEMA4G	26	21677156	21690795	BovineHD2600005595	0,00000	21629048	BestCandidate
IH	FA18	ENSBTAG00000021071	0,00132	TRIM8	26	23536131	23548589	BovineHD2600006067	0,00000	23497760	BestCandidate
IH	FA18	ENSBTAG00000021397	0,00000	NKX2-3	26	20400227	20403772	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA18	ENSBTAG00000021566	0,00000	PAX2	26	21470763	21546730	BovineHD2600005557	0,00000	21479224	BestCandidate
IH	FA18	ENSBTAG00000023629	0,00000	NA	26	20527171	20527751	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA18	ENSBTAG00000023947	0,04865	AVPI1	26	18711832	18719926	BovineHD2600004833	0,00001	18761989	PossibleFalsePositive
IH	FA18	ENSBTAG00000027715	0,01445	NA	26	21057820	21058839	Hapmap31825-BTA-158647	0,00002	21056547	BestCandidate
IH	FA18	ENSBTAG00000032761	0,02577	CALHM3	26	24303869	24309613	BovineHD2600006239	0,00008	24238250	BestCandidate
IH	FA18	ENSBTAG00000037006	0,00477	5S_rRNA	26	25366087	25366206	BovineHD2600006631	0,00003	25441091	BestCandidate
IH	FA18	ENSBTAG00000040263	0,04865	NA	26	25038735	25038826	Hapmap53060-rs29020888	0,00008	25032529	BestCandidate
IH	FA18	ENSBTAG00000043222	0,01445	SNORA12	26	21014545	21014692	Hapmap31825-BTA-158647	0,00002	21056547	PossibleFalsePositive
IH	FA18	ENSBTAG00000043254	0,00251	U6	26	20371949	20372055	BovineHD2600005288	0,00000	20427852	BestCandidate
IH	FA18	ENSBTAG00000044153	0,01445	SH3PXD2A	26	24413186	24469653	ARS-BFGL-NGS-1092	0,00000	24531763	BestCandidate
IH	FA18	ENSBTAG00000044522	0,04865	bta-mir-339b	7	18379980	18380037	ARS-BFGL-NGS-111581	0,00006	18419552	BestCandidate
IH	FA18	ENSBTAG00000045060	0,04865	SNORA70	7	18373015	18373100	ARS-BFGL-NGS-111581	0,00006	18419552	PossibleFalsePositive
IH	FA18	ENSBTAG00000045130	0,00000	SNORA70	26	21566316	21566435	BovineHD2600005579	0,00000	21555707	BestCandidate
IH	FA18	ENSBTAG00000045148	0,02577	bta-mir-2393	26	25360584	25360637	BovineHD2600006606	0,00009	25375304	BestCandidate
IH	FA18	ENSBTAG00000045703	0,00000	COX15	26	20533690	20550733	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA18	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate
IH	FA18	ENSBTAG00000047077	0,00000	ENTPD7	26	20494069	20524328	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA18	ENSBTAG00000047450	0,04865	SFR1	26	24945604	24949627	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA18	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IH	FA19	ENSBTAG00000000856	0,02980	FBXL6	14	1766767	1769754	ARS-BFGL-NGS-4939	0,00000	1801116	PossibleFalsePositive
IH	FA19	ENSBTAG00000000857	0,02980	NA	14	1763994	1766621	ARS-BFGL-NGS-4939	0,00000	1801116	PossibleFalsePositive
IH	FA19	ENSBTAG00000004969	0,03973	LRRC14	14	1610263	1613725	BovineHD1400000206	0,00001	1679844	PossibleFalsePositive
IH	FA19	ENSBTAG00000004970	0,03973	LRRC24	14	1604105	1609477	BovineHD1400000206	0,00001	1679844	PossibleFalsePositive
IH	FA19	ENSBTAG00000005311	0,02980	POLR3H	5	113138060	113150601	Hapmap25014-BTA-123017	0,00006	113100000	PossibleFalsePositive
IH	FA19	ENSBTAG00000006429	0,02980	ACO2	5	113089139	113138185	Hapmap25014-BTA-123017	0,00006	113100000	BestCandidate
IH	FA19	ENSBTAG00000007749	0,03406	TONSL	14	1681494	1692498	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA19	ENSBTAG00000008355	0,03406	CPSF1	14	1728207	1742670	ARS-BFGL-NGS-4939	0,00000	1801116	PossibleFalsePositive
IH	FA19	ENSBTAG00000010276	0,03973	RECQL4	14	1614027	1620509	BovineHD1400000206	0,00001	1679844	BestCandidate
IH	FA19	ENSBTAG00000011064	0,03406	ADCK5	14	1742714	1756301	ARS-BFGL-NGS-4939	0,00000	1801116	PossibleFalsePositive
IH	FA19	ENSBTAG00000026320	0,03406	VPS28	14	1693641	1698490	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA19	ENSBTAG00000026356	0,02980	DGAT1	14	1795351	1804562	ARS-BFGL-NGS-4939	0,00000	1801116	BestCandidate
IH	FA19	ENSBTAG00000035158	0,02980	TMEM249	14	1770660	1772329	ARS-BFGL-NGS-4939	0,00000	1801116	PossibleFalsePositive
IH	FA19	ENSBTAG00000035254	0,03406	CYHR1	14	1663923	1677519	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA19	ENSBTAG00000042638	0,02980	SNORA11	5	113110355	113110471	Hapmap25014-BTA-123017	0,00006	113100000	PossibleFalsePositive
IH	FA19	ENSBTAG00000046026	0,03406	SLC39A4	14	1719732	1724220	UFL-rs134432442	0,00000	1736599	BestCandidate
IH	FA19	ENSBTAG00000046031	0,03973	C8orf82	14	1602474	1605012	BovineHD1400000206	0,00001	1679844	PossibleFalsePositive
IH	FA19	ENSBTAG00000046208	0,02980	SCRT1	14	1782901	1788087	ARS-BFGL-NGS-4939	0,00000	1801116	PossibleFalsePositive
IH	FA21	ENSBTAG00000000312	0,00993	GRINA	14	2018559	2021709	BovineHD1400000262	0,00001	1967325	PossibleFalsePositive
IH	FA21	ENSBTAG00000000856	0,00000	FBXL6	14	1766767	1769754	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA21	ENSBTAG00000000857	0,00000	NA	14	1763994	1766621	UFL-rs134432442	0,00000	1736599	BestCandidate
IH	FA21	ENSBTAG00000004761	0,00000	FOXH1	14	1654701	1656256	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000004969	0,00119	LRRC14	14	1610263	1613725	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000004970	0,00119	LRRC24	14	1604105	1609477	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000007186	0,00993	ARHGAP39	14	1563866	1600378	BovineHD1400000188	0,00000	1588879	BestCandidate
IH	FA21	ENSBTAG00000007749	0,00000	TONSL	14	1681494	1692498	BovineHD1400000206	0,00000	1679844	BestCandidate
IH	FA21	ENSBTAG00000007753	0,00000	KIFC2	14	1656950	1663804	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000007834	0,00000	PPP1R16A	14	1628814	1633988	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000007835	0,00000	GPT	14	1623903	1626907	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000007838	0,00000	MFSD3	14	1620570	1622643	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000008355	0,00000	CPSF1	14	1728207	1742670	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000009677	0,00993	PARP10	14	2024713	2031386	BovineHD1400000262	0,00001	1967325	PossibleFalsePositive
IH	FA21	ENSBTAG00000010276	0,00119	RECQL4	14	1614027	1620509	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000011064	0,00000	ADCK5	14	1742714	1756301	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000018455	0,03288	COMMD5	14	1531491	1533526	BovineHD1400000188	0,00000	1588879	PossibleFalsePositive
IH	FA21	ENSBTAG00000018456	0,03288	ZNF7	14	1517535	1523303	BovineHD1400000188	0,00000	1588879	PossibleFalsePositive
IH	FA21	ENSBTAG00000020751	0,03288	HSF1	14	1806081	1825793	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA21	ENSBTAG00000024647	0,00993	NA	16	5600155	5959906	BovineHD1600001740	0,00037	6027593	BestCandidate

IH	FA21	ENSBTAG00000026320	0,00000	VPS28	14	1693641	1698490	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000026350	0,03288	SPATC1	14	1970919	1998401	BovineHD1400000262	0,00001	1967325	BestCandidate
IH	FA21	ENSBTAG00000026356	0,00000	DGAT1	14	1795351	1804562	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA21	ENSBTAG00000035158	0,00000	TMEM249	14	1770660	1772329	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA21	ENSBTAG00000035254	0,00000	CYHR1	14	1663923	1677519	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000038487	0,04768	ZNF613	18	58130465	58141877	BovineHD4100013794	0,00028	58178448	BestCandidate
IH	FA21	ENSBTAG00000045478	0,03288	bta-mir-2308	14	1566933	1567001	BovineHD1400000188	0,00000	1588879	PossibleFalsePositive
IH	FA21	ENSBTAG00000046026	0,00000	SLC39A4	14	1719732	1724220	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000046031	0,00119	C8orf82	14	1602474	1605012	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA21	ENSBTAG00000046208	0,00000	SCRT1	14	1782901	1788087	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000000856	0,00000	FBXL6	14	1766767	1769754	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000000857	0,00000	NA	14	1763994	1766621	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000003561	0,01084	STK35	13	53482007	53494820	BovineHD1300015127	0,00013	53472741	BestCandidate
IH	FA25	ENSBTAG00000004761	0,00000	FOXH1	14	1654701	1656256	BovineHD1400000206	0,00000	1679844	BestCandidate
IH	FA25	ENSBTAG00000004969	0,00000	LRRC14	14	1610263	1613725	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA25	ENSBTAG00000004970	0,00000	LRRC24	14	1604105	1609477	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA25	ENSBTAG00000007186	0,01084	ARHGAP39	14	1563866	1600378	BovineHD1400000188	0,00000	1588879	BestCandidate
IH	FA25	ENSBTAG00000007749	0,00000	TONSL	14	1681494	1692498	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000007753	0,00000	KIFC2	14	1656950	1663804	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000007834	0,00000	PPP1R16A	14	1628814	1633988	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA25	ENSBTAG00000007835	0,00000	GPT	14	1623903	1626907	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA25	ENSBTAG00000007838	0,00000	MFSD3	14	1620570	1622643	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA25	ENSBTAG00000008355	0,00000	CPSF1	14	1728207	1742670	UFL-rs134432442	0,00000	1736599	BestCandidate
IH	FA25	ENSBTAG00000010276	0,00000	RECQL4	14	1614027	1620509	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive
IH	FA25	ENSBTAG00000011064	0,00000	ADCK5	14	1742714	1756301	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000020751	0,03973	HSF1	14	1806081	1825793	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000026320	0,00000	VPS28	14	1693641	1698490	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000026356	0,00000	DGAT1	14	1795351	1804562	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000035158	0,00000	TMEM249	14	1770660	1772329	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000035254	0,00000	CYHR1	14	1663923	1677519	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000045478	0,02073	bta-mir-2308	14	1566933	1567001	BovineHD1400000188	0,00000	1588879	PossibleFalsePositive
IH	FA25	ENSBTAG00000046026	0,00000	SLC39A4	14	1719732	1724220	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA25	ENSBTAG00000046031	0,00000	C8orf82	14	1602474	1605012	BovineHD1400000206	0,00000	1679844	PossibleFalsePositive

IH	FA25	ENSBTAG00000046208	0,00000	SCRT1	14	1782901	1788087	UFL-rs134432442	0,00000	1736599	PossibleFalsePositive
IH	FA28	ENSBTAG00000007887	0,02980	GRIN2A	25	8554540	8655588	BovineHD4100017129	0,00005	8553268	BestCandidate
IH	FA28	ENSBTAG00000008942	0,02980	NGEF	3	113219680	113266318	BovineHD0300032706	0,00001	113237636	BestCandidate
IH	FA28	ENSBTAG00000012357	0,02980	NA	14	67016468	67233590	BovineHD1400018723	0,00004	66956213	PossibleFalsePositive
IH	FA28	ENSBTAG00000018058	0,02980	C2orf82	3	113208933	113216709	BovineHD0300032706	0,00001	113237636	PossibleFalsePositive
IH	FA28	ENSBTAG00000024085	0,02980	NA	3	113204847	113205856	BovineHD0300032706	0,00001	113237636	PossibleFalsePositive
IH	FA28	ENSBTAG00000030020	0,02980	bta-mir-599	14	66995149	66995243	BovineHD1400018723	0,00004	66956213	PossibleFalsePositive
IH	FA28	ENSBTAG00000037291	0,02980	bta-mir-875	14	66995017	66995093	BovineHD1400018723	0,00004	66956213	BestCandidate
IH	FA28	ENSBTAG00000040347	0,00000	GPC6	12	68232526	68740385	BTB-00501758	0,00000	68227222	BestCandidate
IH	FA44	ENSBTAG00000004262	0,02384	ZNF454	7	2347258	2366394	BovineHD0700000648	0,00000	2375559	PossibleFalsePositive
IH	FA44	ENSBTAG00000004817	0,04768	NA	18	15498689	15499310	BovineHD1800004897	0,00047	15568505	BestCandidate
IH	FA44	ENSBTAG00000007122	0,02384	ZFP2	7	2371313	2394402	BovineHD0700000648	0,00000	2375559	BestCandidate
IH	FA44	ENSBTAG00000016187	0,02384	GRM6	7	2313237	2324783	BovineHD0700000648	0,00000	2375559	PossibleFalsePositive
IH	FA44	ENSBTAG00000034973	0,04768	NA	7	2302464	2303051	BovineHD0700000648	0,00000	2375559	PossibleFalsePositive
IH	FA46	ENSBTAG00000000354	0,03443	PDE6G	19	51768184	51769908	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000000355	0,03443	OXLD1	19	51763604	51765917	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000000356	0,03443	CCDC137	19	51757841	51763491	BovineHD1900014452	0,00003	51762152	BestCandidate
IH	FA46	ENSBTAG00000000411	0,03901	HGS	19	51733073	51746201	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000005923	0,04218	ABTB2	15	65548291	65735623	ARS-BFGL-NGS-105277	0,00018	65530435	BestCandidate
IH	FA46	ENSBTAG00000006189	0,02384	ACTG1	19	51868429	51871276	UA-IFASA-8764	0,00003	51796076	PossibleFalsePositive
IH	FA46	ENSBTAG00000016776	0,03443	BAHCC1	19	51911293	51945652	BovineHD1900014513	0,00015	51891411	BestCandidate
IH	FA46	ENSBTAG00000019104	0,03443	FAAP100	19	51833821	51842965	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000019105	0,03973	NPLOC4	19	51781191	51829647	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000024932	0,03443	FSCN2	19	51845406	51851192	UA-IFASA-8764	0,00003	51796076	BestCandidate
IH	FA46	ENSBTAG00000030200	0,03901	ARL16	19	51746397	51748740	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000040573	0,03443	TSPAN10	19	51771254	51775213	BovineHD1900014452	0,00003	51762152	PossibleFalsePositive
IH	FA46	ENSBTAG00000047677	0,03443	bta-mir-3533	19	51870131	51870214	UA-IFASA-8764	0,00003	51796076	PossibleFalsePositive
IH	FA51	ENSBTAG00000008884	0,04768	MAPRE3	11	72598140	72653468	ARS-BFGL-NGS-115114	0,00016	72598008	BestCandidate
IH	FA51	ENSBTAG00000011324	0,04768	EMILIN1	11	72539296	72547260	ARS-BFGL-NGS-115114	0,00016	72598008	PossibleFalsePositive
IH	FA51	ENSBTAG00000011325	0,04768	KHK	11	72525752	72538457	ARS-BFGL-NGS-115114	0,00016	72598008	PossibleFalsePositive
IH	FA51	ENSBTAG00000011328	0,04768	CGREF1	11	72521482	72524704	ARS-BFGL-NGS-115114	0,00016	72598008	PossibleFalsePositive
IH	FA51	ENSBTAG00000013403	0,04768	AGBL5	11	72555805	72573334	ARS-BFGL-NGS-115114	0,00016	72598008	PossibleFalsePositive
IH	FA51	ENSBTAG00000042923	0,04768	SNORA62	11	72613721	72613870	ARS-BFGL-NGS-115114	0,00016	72598008	PossibleFalsePositive

IH	FA56	ENSBTAG00000015369	0,02384	MLLT11	3	19762896	19767967	BovineHD0300006279	0,00002	19749099	BestCandidate
IH	FA58	ENSBTAG00000015369	0,02384	MLLT11	3	19762896	19767967	BovineHD0300006279	0,00002	19749099	BestCandidate
IH	FA68	ENSBTAG00000002782	0,03973	ZC3H7B	5	112977785	113010230	BovineHD0500032611	0,00004	112963765	BestCandidate
IH	FA68	ENSBTAG00000005311	0,01192	POLR3H	5	113138060	113150601	Hapmap25014-BTA-123017	0,00000	113100000	PossibleFalsePositive
IH	FA68	ENSBTAG00000006429	0,03973	ACO2	5	113089139	113138185	Hapmap25014-BTA-123017	0,00000	113100000	BestCandidate
IH	FA68	ENSBTAG00000013051	0,04768	PRM3	25	9994132	9994413	ARS-BFGL-NGS-32794	0,00028	9965984	PossibleFalsePositive
IH	FA68	ENSBTAG00000015109	0,04768	TOB2	5	113060940	113070989	Hapmap25014-BTA-123017	0,00000	113100000	PossibleFalsePositive
IH	FA68	ENSBTAG00000016580	0,00000	TEF	5	113028012	113037944	Hapmap25014-BTA-123017	0,00000	113100000	PossibleFalsePositive
IH	FA68	ENSBTAG00000032880	0,04768	NA	25	9995790	9996533	ARS-BFGL-NGS-32794	0,00028	9965984	PossibleFalsePositive
IH	FA68	ENSBTAG00000032884	0,04768	Tnp2	25	9988822	9990357	ARS-BFGL-NGS-32794	0,00028	9965984	BestCandidate
IH	FA68	ENSBTAG00000042633	0,03973	U6	5	112947483	112947589	BovineHD0500032611	0,00004	112963765	PossibleFalsePositive
IH	FA68	ENSBTAG00000042638	0,03973	SNORA11	5	113110355	113110471	Hapmap25014-BTA-123017	0,00000	113100000	PossibleFalsePositive
IH	FA78	ENSBTAG00000004761	0,04470	FOXH1	14	1654701	1656256	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000004863	0,04470	RIC3	15	44962852	45024904	ARS-BFGL-NGS-118098	0,00011	44942917	BestCandidate
IH	FA78	ENSBTAG00000004969	0,04470	LRRC14	14	1610263	1613725	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000004970	0,04470	LRRC24	14	1604105	1609477	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000007749	0,04470	TONSL	14	1681494	1692498	BovineHD1400000206	0,00005	1679844	BestCandidate
IH	FA78	ENSBTAG00000007834	0,04470	PPP1R16A	14	1628814	1633988	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000007835	0,04470	GPT	14	1623903	1626907	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000007838	0,04470	MFSD3	14	1620570	1622643	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000008355	0,04470	CPSF1	14	1728207	1742670	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000010276	0,04470	RECQL4	14	1614027	1620509	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000011064	0,04470	ADCK5	14	1742714	1756301	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000026320	0,04470	VPS28	14	1693641	1698490	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000035254	0,04470	CYHR1	14	1663923	1677519	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000040347	0,04470	GPC6	12	68232526	68740385	BovineHD1200018600	0,00013	68157879	BestCandidate
IH	FA78	ENSBTAG00000046026	0,04470	SLC39A4	14	1719732	1724220	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA78	ENSBTAG00000046031	0,04470	C8orf82	14	1602474	1605012	BovineHD1400000206	0,00005	1679844	PossibleFalsePositive
IH	FA81	ENSBTAG00000000090	0,00000	SEC31B	26	21248525	21280208	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA81	ENSBTAG00000000091	0,00000	NDUFB8	26	21283406	21288806	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA81	ENSBTAG00000000092	0,00000	HIF1AN	26	21291076	21300157	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA81	ENSBTAG00000001017	0,00722	SLK	26	24786360	24844579	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA81	ENSBTAG00000002010	0,01289	PI4K2A	26	18684202	18710907	BovineHD2600004833	0,00000	18761989	PossibleFalsePositive
IH	FA81	ENSBTAG00000002261	0,00000	LBX1	26	21894402	21896298	BovineHD2600005648	0,00000	21926490	BestCandidate
IH	FA81	ENSBTAG00000002430	0,00114	COL17A1	26	24848931	24896345	BovineHD2600006436	0,00000	24918578	BestCandidate
IH	FA81	ENSBTAG00000002880	0,00722	SORBS1	26	16721039	16919269	BovineHD2600004205	0,00103	16786406	BestCandidate
IH	FA81	ENSBTAG00000003294	0,02838	MRPL43	26	21685254	21692758	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA81	ENSBTAG00000003296	0,02838	C10orf2	26	21692833	21698205	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA81	ENSBTAG00000003298	0,02838	LZTS2	26	21703370	21710503	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA81	ENSBTAG00000004612	0,00722	SORCS3	26	26294032	26397191	BTA-61038-no-rs	0,00001	26267018	BestCandidate
IH	FA81	ENSBTAG00000007243	0,00298	CFAP58	26	25185701	25286776	UA-IFASA-4715	0,00001	25314352	BestCandidate
IH	FA81	ENSBTAG00000007476	0,00000	BTRC	26	22001775	22172725	BTB-00932332	0,00000	22118554	BestCandidate
IH	FA81	ENSBTAG00000007594	0,04584	CWF19L1	26	21010622	21035035	Hapmap31825-BTA-158647	0,00003	21056547	PossibleFalsePositive
IH	FA81	ENSBTAG00000007948	0,00722	SORCS1	26	27810198	28389391	BovineHD2600007489	0,00000	28125807	BestCandidate
IH	FA81	ENSBTAG00000008100	0,00477	GOLGA7B	26	18858709	18871716	BovineHD2600004851	0,00001	18820468	BestCandidate
IH	FA81	ENSBTAG00000008102	0,00000	CRTAC1	26	18869719	19013761	BovineHD2600004938	0,00000	19015156	BestCandidate
IH	FA81	ENSBTAG00000010739	0,04584	BLOC1S2	26	21043927	21053587	Hapmap31825-BTA-158647	0,00003	21056547	PossibleFalsePositive
IH	FA81	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IH	FA81	ENSBTAG00000012077	0,00000	SLF2	26	21630816	21670831	BovineHD2600005581	0,00000	21564772	BestCandidate
IH	FA81	ENSBTAG00000012107	0,00000	SLC25A28	26	20466104	20476402	BovineHD2600005302	0,00000	20463679	BestCandidate
IH	FA81	ENSBTAG00000012857	0,00207	CNNM2	26	23795436	23972895	BovineHD2600006134	0,00000	23847594	BestCandidate
IH	FA81	ENSBTAG00000012914	0,00000	WNT8B	26	21240570	21244552	BTB-00931481	0,00000	21226405	BestCandidate
IH	FA81	ENSBTAG00000014574	0,00114	CFAP43	26	24953967	25055692	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA81	ENSBTAG00000018564	0,04334	ZFYVE27	26	18753615	18776150	BovineHD2600004833	0,00000	18761989	BestCandidate
IH	FA81	ENSBTAG00000018566	0,00722	SFRP5	26	18782564	18787376	BovineHD2600004833	0,00000	18761989	PossibleFalsePositive
IH	FA81	ENSBTAG00000018604	0,02838	SEMA4G	26	21677156	21690795	BovineHD2600005595	0,00000	21629048	BestCandidate
IH	FA81	ENSBTAG00000021068	0,04584	SUFU	26	23452980	23517250	BovineHD2600006067	0,00000	23497760	BestCandidate
IH	FA81	ENSBTAG00000021071	0,00841	TRIM8	26	23536131	23548589	BovineHD2600006067	0,00000	23497760	PossibleFalsePositive
IH	FA81	ENSBTAG00000021397	0,00000	NKX2-3	26	20400227	20403772	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA81	ENSBTAG00000021566	0,00000	PAX2	26	21470763	21546730	BovineHD2600005557	0,00000	21479224	BestCandidate
IH	FA81	ENSBTAG00000023629	0,00000	NA	26	20527171	20527751	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA81	ENSBTAG00000023947	0,01289	AVPI1	26	18711832	18719926	BovineHD2600004833	0,00000	18761989	PossibleFalsePositive
IH	FA81	ENSBTAG00000027715	0,04584	NA	26	21057820	21058839	Hapmap31825-BTA-158647	0,00003	21056547	BestCandidate
IH	FA81	ENSBTAG00000032761	0,01289	CALHM3	26	24303869	24309613	BovineHD2600006239	0,00006	24238250	BestCandidate
IH	FA81	ENSBTAG00000033344	0,01882	NA	10	18889118	18907145	BovineHD1000006246	0,00009	18944621	BestCandidate
IH	FA81	ENSBTAG00000037006	0,00114	5S_rRNA	26	25366087	25366206	UA-IFASA-4715	0,00001	25314352	PossibleFalsePositive

IH	FA81	ENSBTAG00000038540	0,04584	NA	26	25060219	25073503	Hapmap53060-rs29020888	0,00009	25032529	PossibleFalsePositive
IH	FA81	ENSBTAG00000040263	0,03881	NA	26	25038735	25038826	Hapmap53060-rs29020888	0,00009	25032529	BestCandidate
IH	FA81	ENSBTAG00000040290	0,04584	GSTO2	26	25103430	25119670	Hapmap53060-rs29020888	0,00009	25032529	PossibleFalsePositive
IH	FA81	ENSBTAG00000043222	0,04584	SNORA12	26	21014545	21014692	Hapmap31825-BTA-158647	0,00003	21056547	PossibleFalsePositive
IH	FA81	ENSBTAG00000043254	0,00207	U6	26	20371949	20372055	BovineHD2600005288	0,00000	20427852	BestCandidate
IH	FA81	ENSBTAG00000044153	0,00530	SH3PXD2A	26	24413186	24469653	ARS-BFGL-NGS-1092	0,00000	24531763	BestCandidate
IH	FA81	ENSBTAG00000044306	0,04584	snoU83D	26	29357369	29357450	BTB-01883871	0,00004	29388641	BestCandidate
IH	FA81	ENSBTAG00000045130	0,00000	SNORA70	26	21566316	21566435	BovineHD2600005579	0,00000	21555707	BestCandidate
IH	FA81	ENSBTAG00000045148	0,00530	bta-mir-2393	26	25360584	25360637	UA-IFASA-4715	0,00001	25314352	PossibleFalsePositive
IH	FA81	ENSBTAG00000045703	0,00000	COX15	26	20533690	20550733	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA81	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate
IH	FA81	ENSBTAG00000047077	0,00000	ENTPD7	26	20494069	20524328	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA81	ENSBTAG00000047450	0,00722	SFR1	26	24945604	24949627	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA81	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IH	FA82	ENSBTAG00000000090	0,00000	SEC31B	26	21248525	21280208	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA82	ENSBTAG00000000091	0,00000	NDUFB8	26	21283406	21288806	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA82	ENSBTAG00000000092	0,00000	HIF1AN	26	21291076	21300157	BTB-00931481	0,00000	21226405	PossibleFalsePositive
IH	FA82	ENSBTAG00000001017	0,00000	SLK	26	24786360	24844579	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA82	ENSBTAG00000002010	0,00116	PI4K2A	26	18684202	18710907	BovineHD2600004833	0,00000	18761989	PossibleFalsePositive
IH	FA82	ENSBTAG00000002261	0,00000	LBX1	26	21894402	21896298	BovineHD2600005648	0,00000	21926490	BestCandidate
IH	FA82	ENSBTAG00000002430	0,00000	COL17A1	26	24848931	24896345	BovineHD2600006436	0,00000	24918578	BestCandidate
IH	FA82	ENSBTAG00000002880	0,00935	SORBS1	26	16721039	16919269	Hapmap58930-rs29010490	0,00087	16822073	BestCandidate
IH	FA82	ENSBTAG00000003294	0,00000	MRPL43	26	21685254	21692758	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA82	ENSBTAG00000003296	0,00000	C10orf2	26	21692833	21698205	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA82	ENSBTAG00000003298	0,00000	LZTS2	26	21703370	21710503	BovineHD2600005595	0,00000	21629048	PossibleFalsePositive
IH	FA82	ENSBTAG00000003576	0,02235	POLL	26	22197461	22205845	BovineHD2600005698	0,00000	22122641	BestCandidate
IH	FA82	ENSBTAG00000003741	0,00497	NEURL1	26	24330861	24403159	ARS-BFGL-NGS-2180	0,00001	24477962	BestCandidate
IH	FA82	ENSBTAG00000003989	0,04520	NA	26	25088448	25097722	Hapmap53060-rs29020888	0,00013	25032529	PossibleFalsePositive
IH	FA82	ENSBTAG00000004318	0,00063	ARL3	26	23563111	23761992	BovineHD2600006067	0,00000	23497760	PossibleFalsePositive
IH	FA82	ENSBTAG00000004612	0,00063	SORCS3	26	26294032	26397191	BovineHD2600006943	0,00000	26242200	BestCandidate
IH	FA82	ENSBTAG00000007044	0,00000	CALHM1	26	24284483	24287960	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000007243	0,00876	CFAP58	26	25185701	25286776	UA-IFASA-4715	0,00005	25314352	BestCandidate
IH	FA82	ENSBTAG00000007476	0,00000	BTRC	26	22001775	22172725	BTB-00932332	0,00000	22118554	BestCandidate
IH	FA82	ENSBTAG00000007588	0,01255	ERLIN1	26	20921806	20963567	BovineHD2600005436	0,00000	21008688	BestCandidate
IH	FA82	ENSBTAG00000007591	0,00116	CHUK	26	20966010	21008277	Hapmap31825-BTA-158647	0,00000	21056547	PossibleFalsePositive
IH	FA82	ENSBTAG00000007594	0,00159	CWF19L1	26	21010622	21035035	Hapmap31825-BTA-158647	0,00000	21056547	PossibleFalsePositive
IH	FA82	ENSBTAG00000007948	0,00311	SORCS1	26	27810198	28389391	BovineHD2600007489	0,00000	28125807	BestCandidate
IH	FA82	ENSBTAG00000008100	0,00159	GOLGA7B	26	18858709	18871716	BovineHD2600004851	0,00001	18820468	BestCandidate
IH	FA82	ENSBTAG00000008102	0,00000	CRTAC1	26	18869719	19013761	BovineHD2600004938	0,00000	19015156	BestCandidate
IH	FA82	ENSBTAG00000008936	0,03221	ABCC2	26	20613538	20684065	BovineHD2600005345	0,00000	20630551	BestCandidate
IH	FA82	ENSBTAG00000008939	0,00935	DNMBP	26	20694707	20777487	BovineHD2600005345	0,00000	20630551	PossibleFalsePositive
IH	FA82	ENSBTAG00000009709	0,01538	TAF5	26	24211948	24225837	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000009713	0,01538	NA	26	24226663	24231943	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000009715	0,00159	PDCD11	26	24232143	24274045	BovineHD2600006239	0,00000	24238250	BestCandidate
IH	FA82	ENSBTAG00000009719	0,00000	CALHM2	26	24275663	24281013	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000009879	0,01255	NA	26	24153121	24183826	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000010739	0,00063	BLOC1S2	26	21043927	21053587	Hapmap31825-BTA-158647	0,00000	21056547	PossibleFalsePositive
IH	FA82	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IH	FA82	ENSBTAG00000011960	0,03221	GOT1	26	20285687	20310044	ARS-BFGL-NGS-23064	0,00000	20365711	BestCandidate
IH	FA82	ENSBTAG00000012077	0,00000	SLF2	26	21630816	21670831	BovineHD2600005581	0,00000	21564772	BestCandidate
IH	FA82	ENSBTAG00000012107	0,00000	SLC25A28	26	20466104	20476402	BovineHD2600005302	0,00000	20463679	BestCandidate
IH	FA82	ENSBTAG00000012857	0,00000	CNNM2	26	23795436	23972895	BovineHD2600006134	0,00000	23847594	BestCandidate
IH	FA82	ENSBTAG00000012858	0,03221	NT5C2	26	23983206	24080557	ARS-BFGL-NGS-111090	0,00000	23920913	BestCandidate
IH	FA82	ENSBTAG00000012914	0,00000	WNT8B	26	21240570	21244552	BTB-00931481	0,00000	21226405	BestCandidate
IH	FA82	ENSBTAG00000014335	0,02567	CYP17A1	26	23694362	23700404	Hapmap49372-BTA-91009	0,00003	23689229	PossibleFalsePositive
IH	FA82	ENSBTAG00000014336	0,01538	WBP1L	26	23674348	23684808	Hapmap49372-BTA-91009	0,00003	23689229	BestCandidate
IH	FA82	ENSBTAG00000014574	0,00063	CFAP43	26	24953967	25055692	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA82	ENSBTAG00000015019	0,03221	OBFC1	26	24700354	24737868	ARS-BFGL-NGS-118189	0,00000	24786731	BestCandidate
IH	FA82	ENSBTAG00000018564	0,01255	ZFYVE27	26	18753615	18776150	BovineHD2600004833	0,00000	18761989	BestCandidate
IH	FA82	ENSBTAG00000018566	0,00355	SFRP5	26	18782564	18787376	BovineHD2600004833	0,00000	18761989	PossibleFalsePositive
IH	FA82	ENSBTAG00000018604	0,00000	SEMA4G	26	21677156	21690795	BovineHD2600005595	0,00000	21629048	BestCandidate
IH	FA82	ENSBTAG00000021068	0,03221	SUFU	26	23452980	23517250	BovineHD2600006067	0,00000	23497760	BestCandidate
IH	FA82	ENSBTAG00000021071	0,00000	TRIM8	26	23536131	23548589	BovineHD2600006067	0,00000	23497760	PossibleFalsePositive
IH	FA82	ENSBTAG00000021246	0,01255	NA	26	23728581	23741299	Hapmap49372-BTA-91009	0,00003	23689229	PossibleFalsePositive
IH	FA82	ENSBTAG00000021397	0,00000	NKX2-3	26	20400227	20403772	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA82	ENSBTAG00000021566	0,00000	PAX2	26	21470763	21546730	BovineHD2600005557	0,00000	21479224	BestCandidate

IH	FA82	ENSBTAG00000023629	0,00000	NA	26	20527171	20527751	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA82	ENSBTAG00000023947	0,00116	AVPI1	26	18711832	18719926	BovineHD2600004833	0,00000	18761989	PossibleFalsePositive
IH	FA82	ENSBTAG00000027715	0,00063	NA	26	21057820	21058839	Hapmap31825-BTA-158647	0,00000	21056547	BestCandidate
IH	FA82	ENSBTAG00000032761	0,00000	CALHM3	26	24303869	24309613	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000036127	0,01255	NA	26	23747358	23764458	Hapmap49372-BTA-91009	0,00003	23689229	PossibleFalsePositive
IH	FA82	ENSBTAG00000036423	0,04078	bta-mir-146b	26	22930890	22930995	ARS-BFGL-NGS-107403	0,00000	22889812	BestCandidate
IH	FA82	ENSBTAG00000037006	0,00063	5S_rRNA	26	25366087	25366206	BovineHD2600006631	0,00001	25441091	BestCandidate
IH	FA82	ENSBTAG00000038540	0,03221	NA	26	25060219	25073503	Hapmap53060-rs29020888	0,00013	25032529	PossibleFalsePositive
IH	FA82	ENSBTAG00000038879	0,01538	bta-mir-1307	26	24230071	24230219	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000038993	0,01538	NA	26	24215411	24215535	BovineHD2600006239	0,00000	24238250	PossibleFalsePositive
IH	FA82	ENSBTAG00000040263	0,02890	NA	26	25038735	25038826	Hapmap53060-rs29020888	0,00013	25032529	BestCandidate
IH	FA82	ENSBTAG00000040290	0,04078	GSTO2	26	25103430	25119670	Hapmap53060-rs29020888	0,00013	25032529	PossibleFalsePositive
IH	FA82	ENSBTAG00000042922	0,02235	U6	26	20894775	20894881	BovineHD2600005427	0,00010	20959354	BestCandidate
IH	FA82	ENSBTAG00000043222	0,00159	SNORA12	26	21014545	21014692	Hapmap31825-BTA-158647	0,00000	21056547	PossibleFalsePositive
IH	FA82	ENSBTAG00000043254	0,00000	U6	26	20371949	20372055	BovineHD2600005288	0,00000	20427852	BestCandidate
IH	FA82	ENSBTAG00000044153	0,00063	SH3PXD2A	26	24413186	24469653	ARS-BFGL-NGS-1092	0,00000	24531763	BestCandidate
IH	FA82	ENSBTAG00000044887	0,03221	SCARNA18	26	20756299	20756381	BovineHD2600005359	0,00009	20688560	PossibleFalsePositive
IH	FA82	ENSBTAG00000045082	0,03221	SCARNA17	26	20756089	20756231	BovineHD2600005359	0,00009	20688560	BestCandidate
IH	FA82	ENSBTAG00000045130	0,00000	SNORA70	26	21566316	21566435	BovineHD2600005579	0,00000	21555707	BestCandidate
IH	FA82	ENSBTAG00000045148	0,01255	bta-mir-2393	26	25360584	25360637	UA-IFASA-4715	0,00005	25314352	PossibleFalsePositive
IH	FA82	ENSBTAG00000045703	0,00000	COX15	26	20533690	20550733	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA82	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate
IH	FA82	ENSBTAG00000047077	0,00000	ENTPD7	26	20494069	20524328	BovineHD2600005302	0,00000	20463679	PossibleFalsePositive
IH	FA82	ENSBTAG00000047450	0,00000	SFR1	26	24945604	24949627	BovineHD2600006436	0,00000	24918578	PossibleFalsePositive
IH	FA82	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA04	ENSBTAG00000002089	0,00794	FAM217A	23	49935473	49947452	BovineHD2300014545	0,00012	49876414	PossibleFalsePositive
IS	FA04	ENSBTAG00000015176	0,00000	C6orf201	23	49921820	49930811	BovineHD2300014545	0,00012	49876414	PossibleFalsePositive
IS	FA04	ENSBTAG00000015178	0,00794	ECI2	23	49907476	49923551	BovineHD2300014545	0,00012	49876414	BestCandidate
IS	FA06	ENSBTAG00000015980	0,00000	FASN	19	51384922	51403614	BovineHD1900014372	0,00000	51386735	BestCandidate
IS	FA06	ENSBTAG00000019317	0,00000	GPS1	19	51422926	51427197	BovineHD1900014372	0,00000	51386735	PossibleFalsePositive
IS	FA06	ENSBTAG00000047760	0,00000	DUS1L	19	51416679	51422719	BovineHD1900014372	0,00000	51386735	PossibleFalsePositive
IS	FA07	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA07	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate

IS	FA07	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA09	ENSBTAG00000011575	0,00000	RFNG	19	51429344	51431873	BovineHD1900014372	0,00000	51386735	PossibleFalsePositive
IS	FA09	ENSBTAG00000015980	0,00000	FASN	19	51384922	51403614	BovineHD1900014372	0,00000	51386735	BestCandidate
IS	FA09	ENSBTAG00000019317	0,00000	GPS1	19	51422926	51427197	BovineHD1900014372	0,00000	51386735	PossibleFalsePositive
IS	FA09	ENSBTAG00000019321	0,04765	CCDC57	19	51271301	51351984	BovineHD1900014372	0,00000	51386735	PossibleFalsePositive
IS	FA09	ENSBTAG00000047760	0,00000	DUS1L	19	51416679	51422719	BovineHD1900014372	0,00000	51386735	PossibleFalsePositive
IS	FA10	ENSBTAG00000019343	0,02383	MDFI	23	15467450	15482249	BovineHD2300015562	0,00000	15443294	BestCandidate
IS	FA15	ENSBTAG00000000039	0,02803	SIRT7	19	51585609	51591804	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000000040	0,02803	MAFG	19	51576540	51582086	ARS-BFGL-NGS-90673	0,00004	51581082	BestCandidate
IS	FA15	ENSBTAG00000000042	0,00681	PYCR1	19	51567328	51571991	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000000044	0,00681	MYADML2	19	51562314	51565450	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000001868	0,02803	PCYT2	19	51592128	51599282	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000003436	0,03971	RHBDF2	19	55893571	55902692	BovineHD1900015816	0,00003	55957032	BestCandidate
IS	FA15	ENSBTAG00000004632	0,01649	NA	19	51499066	51528822	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000008747	0,01649	DCXR	19	51466698	51469340	ARS-BFGL-NGS-39983	0,00001	51395684	PossibleFalsePositive
IS	FA15	ENSBTAG00000011575	0,00000	RFNG	19	51429344	51431873	BovineHD1900014372	0,00000	51386735	BestCandidate
IS	FA15	ENSBTAG00000013898	0,02803	NPB	19	51601521	51601969	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000015980	0,00000	FASN	19	51384922	51403614	BovineHD1900014364	0,00000	51349695	PossibleFalsePositive
IS	FA15	ENSBTAG00000019317	0,00000	GPS1	19	51422926	51427197	BovineHD1900014364	0,00000	51349695	PossibleFalsePositive
IS	FA15	ENSBTAG00000019321	0,00000	CCDC57	19	51271301	51351984	BovineHD1900014364	0,00000	51349695	BestCandidate
IS	FA15	ENSBTAG00000022927	0,01649	RAC3	19	51470292	51471813	ARS-BFGL-NGS-39983	0,00001	51395684	PossibleFalsePositive
IS	FA15	ENSBTAG00000044643	0,01649	bta-mir-2346	19	51516011	51516089	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA15	ENSBTAG00000047043	0,01649	NA	19	51461750	51463698	ARS-BFGL-NGS-39983	0,00001	51395684	BestCandidate
IS	FA15	ENSBTAG00000047760	0,00000	DUS1L	19	51416679	51422719	BovineHD1900014364	0,00000	51349695	PossibleFalsePositive
IS	FA15	ENSBTAG00000047973	0,01649	NA	19	51499260	51515723	ARS-BFGL-NGS-90673	0,00004	51581082	PossibleFalsePositive
IS	FA18	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA18	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate
IS	FA18	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA21	ENSBTAG00000001521	0,02383	UQCRB	14	70329414	70334124	BovineHD1400019742	0,00002	70351638	BestCandidate
IS	FA21	ENSBTAG00000001522	0,02383	MTERF3	14	70306723	70324461	BovineHD1400019742	0,00002	70351638	PossibleFalsePositive
IS	FA25	ENSBTAG00000007594	0,03971	CWF19L1	26	21010622	21035035	BovineHD2600005454	0,00000	21102710	PossibleFalsePositive
IS	FA25	ENSBTAG00000010739	0,01430	BLOC1S2	26	21043927	21053587	BovineHD2600005454	0,00000	21102710	PossibleFalsePositive
IS	FA25	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005454	0,00000	21102710	BestCandidate

IS	FA25	ENSBTAG00000027715	0,01430	NA	26	21057820	21058839	BovineHD2600005454	0,00000	21102710	PossibleFalsePositive
IS	FA25	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005454	0,00000	21102710	PossibleFalsePositive
IS	FA25	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005454	0,00000	21102710	PossibleFalsePositive
IS	FA30	ENSBTAG00000048175	0,00000	NA	5	102894599	102923181	ARS-BFGL-NGS-40945	0,00001	102855072	BestCandidate
IS	FA32	ENSBTAG00000013334	0,03404	CSF3R	3	110001845	110012507	BovineHD0300031616	0,00022	109936441	BestCandidate
IS	FA32	ENSBTAG00000016252	0,02978	LSM12	19	44509434	44529146	BTA-45551-no-rs	0,00029	44597888	PossibleFalsePositive
IS	FA32	ENSBTAG00000016253	0,02978	G6PC3	19	44531725	44536439	BTA-45551-no-rs	0,00029	44597888	BestCandidate
IS	FA32	ENSBTAG00000033502	0,03574	TMEM225	29	27570026	27573346	BovineHD2900008070	0,00000	27500510	PossibleFalsePositive
IS	FA32	ENSBTAG00000039964	0,03404	NA	29	27444678	27445628	BovineHD2900008070	0,00000	27500510	PossibleFalsePositive
IS	FA32	ENSBTAG00000040239	0,03404	NA	29	27523131	27524063	BovineHD2900008070	0,00000	27500510	BestCandidate
IS	FA32	ENSBTAG00000046173	0,02978	ALG12	5	120924137	120932984	BovineHD0500035302	0,00005	120870313	BestCandidate
IS	FA32	ENSBTAG00000047801	0,02978	CRELD2	5	120933369	120940333	BovineHD0500035302	0,00005	120870313	PossibleFalsePositive
IS	FA38	ENSBTAG00000009637	0,00000	SLC12A2	7	26973462	27064924	BTB-00264267	0,00000	27097343	BestCandidate
IS	FA65	ENSBTAG00000039738	0,04765	TMIGD3	3	31900284	31986507	BovineHD0300009990	0,00010	31918422	BestCandidate
IS	FA72	ENSBTAG00000032829	0,02383	GHITM	28	39435546	39448963	ARS-BFGL-NGS-118113	0,00007	39482527	BestCandidate
IS	FA81	ENSBTAG00000007914	0,03574	NA	4	92315807	92316363	BovineHD0400025748	0,00003	92356201	BestCandidate
IS	FA81	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA81	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate
IS	FA81	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA82	ENSBTAG00000007594	0,04765	CWF19L1	26	21010622	21035035	BovineHD2600005454	0,00020	21102710	BestCandidate
IS	FA82	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA82	ENSBTAG00000021566	0,04765	PAX2	26	21470763	21546730	BTB-00931586	0,00000	21409429	BestCandidate
IS	FA82	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005467	0,00000	21149234	BestCandidate
IS	FA82	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005467	0,00000	21149234	PossibleFalsePositive
IS	FA83	ENSBTAG00000007591	0,00681	CHUK	26	20966010	21008277	BovineHD2600005441	0,00000	21040105	PossibleFalsePositive
IS	FA83	ENSBTAG00000007594	0,00000	CWF19L1	26	21010622	21035035	BovineHD2600005441	0,00000	21040105	PossibleFalsePositive
IS	FA83	ENSBTAG00000010739	0,00000	BLOC1S2	26	21043927	21053587	BovineHD2600005441	0,00000	21040105	BestCandidate
IS	FA83	ENSBTAG00000010742	0,00000	PKD2L1	26	21056133	21108833	BovineHD2600005441	0,00000	21040105	PossibleFalsePositive
IS	FA83	ENSBTAG00000027715	0,00000	NA	26	21057820	21058839	BovineHD2600005441	0,00000	21040105	PossibleFalsePositive
IS	FA83	ENSBTAG00000043222	0,00893	SNORA12	26	21014545	21014692	BovineHD2600005441	0,00000	21040105	PossibleFalsePositive
IS	FA83	ENSBTAG00000045728	0,00000	NA	26	21141592	21148318	BovineHD2600005472	0,00000	21169203	BestCandidate
IS	FA83	ENSBTAG00000047957	0,00000	SCD	26	21132751	21133969	BovineHD2600005472	0,00000	21169203	PossibleFalsePositive
IS	FA84	ENSBTAG00000003532	0,02383	TLE4	8	55837381	55992436	BTA-119741-no-rs	0,00002	55927704	BestCandidate

Supplementary Table II-S2. Significant genes (i.e. FDR q-values less than or equal to 0.05) obtained with MUGBAS analysis divided for each chromosome (Chr), breed and fatty acid (FA).

Chr	***Breed***	***FA***	***FA code***	***Gene[1]***
3	Italian Holstein	C14-iso	FA14	*ENSA, MCL1, GOLPH3L*
		C15-iso	FA16	*EVA1B, MRPS15, SH3D21*
		C17:1 c9	FA28	*NGEF, C2orf82, ENSBTAG00000024085*
		SFA	FA56	*MLLT11*
		UFA	FA57	*MLLT11*
		PUFA	FA58	*MLLT11, GABPB2*
	Italian Simmental	C18:1 t6-8	FA32	*CSF3R*
		BCFA	FA65	*TMIGD3*
4	Italian Simmental	ID 10-1/(10+10-1)	FA81	*ENSBTAG00000007914*
5	Italian Holstein	C15:0	FA19	*POLR3H, ACO2, SNORA11*
		OCFA	FA68	*ZC3H7B, POLR3H, ACO2, TOB2, TEF, U6, SNORA11*
	Italian Simmental	C18:1 t4	FA30	*ENSBTAG00000048175*
		C18:1 t6-8	FA32	*ALG12, CRELD2*
7	Italian Holstein	C14:1c9	FA18	*bta-mir-339b, SNORA70*
		C20:0	FA44	*ZNF454, ZFP2, GRM6, ENSBTAG00000034973*
	Italian Simmental	C18:1 c11	FA38	*SLC12A2*
8	Italian Simmental	ID 18-1/(18+18-1)	FA84	*TLE4*
9	Italian Holstein	PUFA	FA58	*SESN1*
10	Italian Holstein	ID 10-1/(10+10-1)	FA81	*SENP8*
11	Italian Holstein	C20:4 c5,c8,c11,c14	FA51	*MAPRE3, EMILIN1, KHK, CGREF1, AGBL5, SNORA62*
12	Italian Holstein	C17:1 c9	FA28	*GPC6*
12	Italian Holstein	BCFA + OCFA	FA78	*GPC6*
13	Italian Holstein	C16:1 c9	FA25	*STK35*
14	Italian Holstein	C15:0	FA19	*FBXL6, SLC52A2, LRRC14, LRRC24, TONSL, CPSF1, RECQL4, ADCK5, VPS28, DGAT1, TMEM249, CYHR1, SLC39A4, C8orf82, SCRT1*
		C16:0	FA21	*GRINA, FBXL6, SLC52A2, FOXH1, LRRC14, LRRC24, ARHGAP39, TONSL, KIFC2, PPP1R16A, GPT, MFSD3, CPSF1, PARP10, RECQL4, ADCK5, COMMD5, ZNF7, HSF1, VPS28, SPATC1, DGAT1, TMEM249, CYHR1, bta-mir-2308, SLC39A4, C8orf82, SCRT1*

		C16:1 c9	FA25	*FBXL6, SLC52A2, FOXH1, LRRC14, LRRC24, ARHGAP39, TONSL, KIFC2, PPP1R16A, GPT, MFSD3, CPSF1, RECQL4, ADCK5, HSF1, VPS28, DGAT1, TMEM249, CYHR1, bta-mir-2308, SLC39A4, C8orf82, SCRT1*
		C17:1 c9	FA28	*ENSBTAG00000012357, MIR599, MIR875*
		BCFA + OCFA	FA78	*FOXH1, LRRC14, LRRC24, TONSL, PPP1R16A, GPT, MFSD3, CPSF1, RECQL4, ADCK5, VPS28, CYHR1, SLC39A4, C8orf82*
	Italian Simmental	C16:0	FA21	*UQCRB, MTERF3*
15	Italian Holstein	BCFA + OCFA	FA78	*RIC3*
		C18:2 9,11 c/t	FA46	*ABTB2*
16	Italian Holstein	C16:0	FA21	*LOC781004*
		PUFA	FA58	*ZNF281*
18	Italian Holstein	C16:0	FA21	*ZNF613*
		C20:0	FA44	*ENSBTAG00000004817*
		PUFA	FA58	*PPFIA3, HRC, TRPM4, ZNF350, ZNF677, ENSBTAG00000039969, ZNF112, bta-mir-2900*
19	Italian Holstein	C14-iso	FA14	*CCT6B, SNORA70*
		C14:0	FA15	*KPNA2*
		C18:2 9,11 c/t	FA46	*PDE6G, OXLD1, CCDC137, HGS, ACTG1, BAHCC1, FAAP100, NPLOC4, FSCN2, ARL16, TSPAN10, bta-mir-3533*
	Italian Simmental	C10:0	FA06	*FASN, GPS1, DUS1L*
		C12:0	FA09	*RFNG, FASN, GPS1, CCDC57, DUS1L*
		C14:0	FA15	*SIRT7, MAFG, PYCR1, MYADML2, PCYT2, RHBDF2, ENSBTAG00000004632, DCXR, RFNG, NPB, FASN, GPS1, CCDC57, RAC3, bta-mir-2346, DUS1L, ENSBTAG00000047973*
		C18:1 t6-8	FA32	*LSM12, G6PC3*
23	Italian Simmental	C7:0	FA04	*FAM217A, C6orf201, ECI2*
		C13-iso	FA10	*MDFI*
25	Italian Holstein	C17:1 c9	FA28	*GRIN2A*
		OCFA	FA68	*PRM3, PRM2, Tnp2*

26	Italian Holstein	C14:1c9	FA18	*SEC31B, NDUFB8, HIF1AN, SLK, PI4K2A, LBX1, COL17A1, SORBS1, MRPL43, C10orf2, LZTS2, SORCS3, CFAP58, BTRC,* ***CHUK*****,** ***CWF19L1*****,** *SORCS1, GOLGA7B, CRTAC1,* ***BLOC1S2*****,** ***PKD2L1*****,** *SLF2, SLC25A28, CNNM2, WNT8B, CFAP43, SFRP5, SEMA4G, TRIM8, NKX2-3,* ***PAX2*****,** *ENSBTAG00000023629, AVPI1, ENSBTAG00000027715, CALHM3, 5S_rRNA, ENSBTAG00000040263,* ***SNORA12*****,** *U6, SH3PXD2A, SNORA70, bta-mir-2393, COX15,* ***SCD1*****,** *ENTPD7, SFR1,* ***SCD***
		ID 10-1/(10+10-1)	FA81	*SEC31B, NDUFB8, HIF1AN, SLK, PI4K2A, LBX1, COL17A1, SORBS1, MRPL43, C10orf2, LZTS2, SORCS3, CFAP58, BTRC,* ***CWF19L1*****,** *SORCS1, GOLGA7B, CRTAC1,* ***BLOC1S2*****,** ***PKD2L1*****,** *SLF2, SLC25A28, CNNM2, WNT8B, CFAP43, ZFYVE27, SFRP5, SEMA4G, SUFU, TRIM8, NKX2-3,* ***PAX2*****,** *ENSBTAG00000023629, AVPI1,* ***LOC101906134*****,** *CALHM3, 5S_rRNA, GSTO1, ENSBTAG00000040263, GSTO2,* ***SNORA12*****,** *U6, SH3PXD2A, snoU83D, SNORA70, bta-mir-2393, COX15,* ***SCD1*****,** *ENTPD7, SFR1,* ***SCD***
		ID 14-1/(14+14-1)	FA82	*SEC31B, NDUFB8, HIF1AN, SLK, PI4K2A, LBX1, COL17A1, SORBS1, MRPL43, C10orf2, LZTS2, POLL, NEURL1, GSTO1, ARL3, SORCS3, CALHM1, CFAP58, BTRC, ERLIN1, CHUK,* ***CWF19L1*****,** *SORCS1, GOLGA7B, CRTAC1, ABCC2, DNMBP, TAF5, USMG5, PDCD11, CALHM2, PCGF6,* ***BLOC1S2*****,** ***PKD2L1*****,** *GOT1, SLF2, SLC25A28, CNNM2, NT5C2, WNT8B, CYP17A1, WBP1L, CFAP43, OBFC1, ZFYVE27, SFRP5, SEMA4G, SUFU, TRIM8, BORCS7, NKX2-3,* ***PAX2*****,** *ENSBTAG00000023629, AVPI1,* ***LOC101906134*****,** *CALHM3, AS3MT, bta-mir-146b, 5S_rRNA, GSTO1, bta-mir-1307, ENSBTAG00000038993, ENSBTAG00000040263, GSTO2, U6,* ***SNORA12*****,** *U6, SH3PXD2A, SCARNA18, SCARNA17, SNORA70, bta-mir-2393, COX15,* ***SCD1*****,** *ENTPD7, SFR1,* ***SCD***
	Italian Simmental	C10:1c9	FA07	***PKD2L1, SCD1, SCD***
		C14:1c9	FA18	***PKD2L1, SCD1, SCD***
		C16:1 c9	FA25	***CWF19L1, BLOC1S2, PKD2L1, LOC101906134, SCD1, SCD***
		ID 10-1/(10+10-1)	FA81	***PKD2L1, SCD1, SCD***
		ID 14-1/(14+14-1)	FA82	***CWF19L1, PKD2L1, PAX2, SCD1, SCD***
		ID 16-1/(16+16-1)	FA83	***CHUK, CWF19L1, BLOC1S2, PKD2L1, LOC101906134, SNORA12, SCD1, SCD***
28	Italian Simmental	PUFA/SFA	FA72	*GHITM*
29	Italian Simmental	C18:1 t6-8	FA32	*TMEM225, LOC781828, LOC781509*

[1] Significant genes overlapping between the two breed are shown in bold

Supplementary Table III-S1. Summary of upstream transcription regulators (TR) in -2vs-14 and +1vs-14 time comparisons obtained by IPA.

-2vs-14 comparison						
Upstream Regulator	**Expr Fold Change**	**Molecule Type**	**Predicted Activation State**	**Activation z-score**	**p-value of overlap**	**Target molecules in dataset**
XBP1		transcription regulator	Activated	5,436	6,23E-14	APBB2,APP,ARFGAP3,CALR,COPB2,COPE,COPZ1,CREB3L1,DAD1,EDEM2,GGCX,GOLGA3,GOLGA4,GOLPH3L,KDELR3,LMAN2,NUCB2,ORMDL3,PDIA3,PDIA4,PDIA6,RPN1,RPN2,SDF2L1,SEC61A1,SEC61B,SERP1,SRPRA,SSR2,SSR3,YIF1A
SREBP1		transcription regulator	Activated	2,227	0,00594	ACSL1,BHLHE40,CYP7A1,FADS1,GPAM,GRSF1,GSR,IL10,LGALS3,LSS,PCK2,RDH11,STXBP1
ATF6		transcription regulator	Activated	2	0,0076	APP,CALR,DERL3,NUCB2,PDIA4,UNC13B
FOS	-1,556	transcription regulator	Inhibited	-2,102	0,00274	ANXA4,APLN,AQP3,ARF1,B4GALT1,Ccnjl,CRABP2,CTSV,EDN1,FOLR2,FOXA1,GSR,IL10,KRT8,LGALS3,MET,MXD1,PGAM2,PLAUR,RARG,RBBP4,RBBP7,RIPK4,SDC1,SNX3,SUMO2,VIM,XDH
KLF4	-1,194	transcription regulator	Inhibited	-2,156	0,00104	ACVR1,CITED1,CRABP2,CTNNB1,CYP1A1,DSP,DUSP1,HES1,HEY2,IL10,NOS3,NOV,NRP1,NTF4,PLAUR,TWIST2,VIM
FBXW7	-1,023	transcription regulator	Inhibited	-2,209	0,000137	APP,CD36,CREB3L1,DGAT1,FGFBP1,GPAM

+1vs-14 comparison						
Upstream Regulator	**Expr Fold Change**	**Molecule Type**	**Predicted Activation State**	**Activation z-score**	**p-value of overlap**	**Target molecules in dataset**
XBP1		transcription regulator	Activated	6,084	1,09E-28	ALG2,APBB2,APP,ARCN1,ARFGAP3,BET1,BLZF1,CALR,CAT,COPB2,COPE,COPZ1,CREB3,CREB3L1,CXCL2,DAD1,DDIT3,DDOST,DNAJB11,DNAJB9,DNAJC3,EDEM2,EIF2AK3,ERP29,FKBP2,FKBP7,GGCX,GOLGA3,GOLGA4,GOLPH3L,GORASP2,HLA-DRA,HSP90B1,HSPA13,HYOU1,ICAM1,KDELR3,LMAN1,LMAN2,MAP1LC3B,MGAT2,MOGS,MYC,NOS2,NR1H3,NUCB2,ORMDL3,PDIA3,PDIA4,PDIA6,PIGA,PRNP,RABAC1,RPN1,RPN2,SDF2L1,SEC11C,SEC23B,SEC31A,SEC61A1,SEC61B,SEC61G,SEC63,SERP1,SMC3,SOD1,SPARC,SRP19,SRP54,SRP68,SRPRA,SSR1,SSR2,SSR3,STX5,SYVN1,TNFSF11,TRAM1,TXNDC11,USO1,VAMP4,XRCC6,YIF1A
IRF7	1,988	transcription regulator	Activated	4,693	0,0000507	ADAR,CASP4,CCL8,CD40,CD69,CMPK2,CXCL10,DDX58,GBP1,HERC5,IFI44,IFIT1,IFIT3,IL15,IRF7,IRF9,ISG15,JAK2,MAP3K8,MCL1,MX1,NAMPT,PARP14,PSME1,PSME2,RSAD2,STAT1,STAT2,TLR4,TMBIM6,UBA7,UBE2L6,USP18,ZBP1,ZC3HAV1
TP53	-1,703	transcription regulator	Activated	4,254	7,57E-39	ACAT1,ACP2,ACSL3,ACTA2,ACTL6A,ACTN4,ADAMTSL4,ADH5,AHCY,AIFM2,AKR1B1,ALDH18A1,ALDH4A1,ANKH,ANTXR1,ANXA4,APBB2,APOE,APP,ARAP2,ASF1B,ASXL1,ATF3,ATG2B,ATG4A,ATXN1,AURKA,AURKB,BCAP31,BCL3,BHLHE40,BLZF1,BNIP3,BRCA1,BTG1,BTG2,C1QC,CA9,CALU,CAMK2N1,CARHSP1,CARS,CASP2,CASP4,CASP9,CAT,CCNA2,CCNB1,CCNB2,CCNE2,CCNG1,CCNL1,CD47,CD59,CDC20,CDC6,CDK1,CDK2,CDKN2D,CEP55,CGREF1,CITED2,CKAP2,CKM,CNN1,COL1A2,COL4A1,COL4A2,COMT,CP,CREB3,CRYAB,CSK,CSRP1,CTGF,CTNNB1,CTSB,CTSK,CYB5A,CYP51A1,CYR61,DBI,DCK,DDIT3,DDR1,DGKA,DHCR24,DLGAP5,DRAM1,DSN1,DSTN,DUSP1,DUSP5,E2F5,E2F8,EIF2AK3,ERCC3,ESPL1,EXO1,F11R,FAM83D,FANCI,FDPS,FERMT2,FGF13,FGFBP1,FIGNL1,FOS,FOXM1,FOXO3,FUBP1,FUCA1,GADD45G,GATM,GBP1,GLRX,GLUL,GNA13,GNL3,GSR,GSTP1,H2AFZ,HERC5,HIC1,HJURP,HK2,HLA-DQA1,HMGCS1,HMMR,HS3ST1,HSPD1,ICAM1,ID2,IDH1,IDH2,IDH3G,IFI30,IGDCC4,IGFBP7,IL10,IL10RA,INPP4A,IRF5,IRF7,IRF9,ISG15,ITGA2,JUN,KAT2B,KIFC1,KIT,KRT14,KSR1,LAPTM4A,LGALS3,LIMK2,LMAN2,LOX,LPIN1,LSS,LYZ,MAD2L1,MAFB,MAN2A1,MAP3K8,MAP4K2,MAPRE3,MBNL2,MCL1,MCM2,MCM3,MELK,MMP23B,MPI,MPZL2,MRPL46,MSH2,MT-CO2,MT-CYB,MT-ND5,MVK,MX1,MYC,MYL9,MYO1C,MYO6,NAMPT,NCAPG,NEK2,NFKBIA,NOLC1,NOS2,NOS3,NPM1,NR2F1,NRARP,NRP1,OMA1,P4HA1,PARD6B,PBK,PCBP4,PCLAF,PCNA,PDCD6IP,PDIA6,PDK1,PEX2,PHKG1,PIM1,PLA2G16,PLK2,PMM1,POLA1,POLD1,POLD2,POLE2,POLK,PPARD,PPM1A,PPM1F,PPP1CC,PRDX3,PRDX6,PRKAB1,PRKAB2,PRKAG2,PRNP,PROM1,PSEN2,PSMD3,PSRC1,PTGS1,PTTG1,RAB8A,RACGAP1,RAD51,RAD51AP1,RALY,RB1CC1,RBBP4,RBBP7,RBM3,RFC4,RGS12,RNASE4,RPN1,RPN2,RPS25,RPS6KA2,RPSA,RRM2B,S100B,SEC23B,SEC61A1,SEC61B,SEMA6A,SERPINE1,SESN1,SLC16A1,SLC19A1,SMARCB1,SMC3,SMC4,SNX5,SOD1,SOD2,SPC25,SPDL1,SPHK2,SPP1,SQLE,SRC,SRSF3,ST14,STARD4,STAT1,STEAP3,STMN1,SUCLG1,TBL1X,TDP2,TIGAR,TIMP2,TINAGL1,TMED7,TMEM127,TOP2A,TP53,TP53INP1,TPD52L1,TRAP1,TRIM6,TSC2,TTC28,TUBB,UBE2C,UBE2T,UNC5B,USO1,USP9X,VAMP4,VASN,VCAN,VDR,VIM,VRK1,WDHD1,WSB2,XPNPEP1,XPO1,YPEL3,ZFP36L1,ZYX
NUPR1		transcription regulator	Activated	4,1	8,02E-13	ACSS1,ANP32A,ARMC7,ATF3,AURKA,B3GAT3,BNIP3,BRCA1,BTG1,C8orf58,CAMK2N1,CASP2,CCNA2,CCNB2,CCNF,CDCA3,CDK2,CENPC,CENPI,CITED2,COL1A2,CROT,CXADR,CXCR4,CYR61,DDIT3,DGCR8,DHCR24,DNMT3B,DSN1,DUSP5,E2F8,ELL2,ERCC6L,ESPL1,EXO1,FANCD2,FCF1,FCHSD2,FLVCR1,FOXO3,FUCA1,GK,GPCPD1,GRAMD3,GSTA4,HJURP,HK2,IL13RA1,ITPR3,KIF11,KIF20A,KIF2C,KIFC1,KXD1,LMNB1,LRP8,MGLL,MGME1,MMD,MXD1,MYC,MYD88,NAAA,NCKIPSD,NEIL3,NFIL3,NR1D2,NRBF2,OSBPL6,OSER1,P4HA2,PARP1,PARP9,PCYOX1,PDK1,PIM1,POLA2,POLE2,PRNP,PTPRJ,RAD51,RBMS1,RELB,RILPL2,RNF19B,RPA1,SAMD4A,SERPINE1,SKA2,SKP2,SLC39A8,SPAG5,SPC25,SPDL1,STX3,SUOX,TDRKH,TFAP2A,TMEM167B,TMEM19,TP53,TRERF1,TUBGCP5,UAP1,UNC5B,ZC3HAV1,ZFAND2A,ZFP36L1
NFATC2	2	transcription regulator	Activated	4,092	0,0000323	ABCA1,CCNA2,CCNB1,CCNF,CD40,CITED2,CMPK2,CRYAB,CX3CR1,CXCL10,DAB2,DGKA,E2F5,FGL2,HDAC1,IFIT3,IL10,IL15,IL18,IRF7,ISG15,MERTK,MYC,NFKBIZ,PDZD2,PLK2,PTPRK,RSAD2,SRC,STAT1,STAT2,TLR3,USP25
CDKN2A		transcription regulator	Activated	3,973	1,52E-11	ASF1B,AURKB,BTG2,CAPG,CCNA2,CCNB1,CCNG1,CDCA5,CDCA7L,CDK1,CDK2,CDKN2C,CDKN2D,CENPK,CHAF1A,CITED2,CNOT6L,CTGF,CXCL10,DCK,DCTN4,DDR1,DONSON,DUSP1,Esrra,FANCA,FBL,FOS,FOXA1,GADD45G,GNA13,GNL3,HOXB9,IL15,ITGAV,ITGB3,JAK2,JUN,KCNK1,KIFC1,LGALS3,MAD2L1,MCL1,MCMBP,MELK,MYC,MYCN,NPM1,ODC1,P4HA2,PCNA,POLD2,POLK,RAB27A,RAD51AP1,RBBP7,RFC4,RRM2B,SERPINE1,SH3BP2,SKP2,TCF19,TLR4,TNFRSF1A,TP53,TP53INP1,UBR7,VIM,VRK1,ZNF385A
IRF3	1,356	transcription regulator	Activated	3,721	0,00108	ADAM9,ANXA4,B2M,B4GALT5,CD69,CMPK2,CXCL10,DDX58,GBP1,IFI44,IFIT1,IFIT3,IL10,IL15,IRF5,IRF7,ISG15,MARCH6,NOS2,PARP14,PRNP,RSAD2,STAT1,STAT2,TFAP2C,TLR3,TLR4,TNFAIP3,UBE2L6,USP18,VIM,ZBP1
SREBP1		transcription regulator	Activated	3,517	0,0000298	ABCA1,ACACB,ACADS,BHLHE40,CFD,CXCL10,CYB5A,CYP51A1,CYP7A1,DBI,DPY19L3,FADS1,FADS2,FDPS,GRSF1,GSR,HK2,HMGCS1,HSPA13,IDH1,IL10,INSIG1,LGALS3,LPIN1,LSS,MSMO1,NOS2,NPC1,NR1H3,NSDHL,PCK2,PPAT,RDH11,SCD,SERPINE1,SQLE,STARD4,STXBP1,SUCLG1,TM7SF2,TP53

RB1	-1,014	transcription regulator	Activated	3,509	0,00000197	ACTC1,ANGPT2,ASF1B,ATP6V1D,AURKB,BNIP3,BRCA1,CASP4,CASP9,CCNA2,CCNB1,CCNE2,CDC6,CDCA5,CDCA7L,CDK1,CDK2,CENPK,CHAF1A,CHRNB1,CITED2,CKM,CNOT6L,COL4A3BP,CSN2,CTGF,DCK,DCTN4,DDIT3,DONSON,Esrra,FANCA,FOS,KCNK1,KIT,LACTB,LCK,LIG1,MCM2,MCM3,MCMBP,MELK,MTOR,MYB,MYC,ORC1,PARP1,PCLAF,PCNA,PGAM2,PGRMC1,PIM1,PSEN2,PTGS1,RAB27A,RAD51,RAD51AP1,RFC4,RSL1D1,SDHD,SERPINE1,SH3BP2,SKP2,SOD2,SRPRA,TCF19,TOPBP1,TP53,UBR7,VDAC1,VRK1
KDM5B		transcription regulator	Activated	3,434	0,000000579	ARL6IP5,AURKA,BRCA1,BUB3,CCNB1,CDCA3,CDK1,CLDN1,CTGF,CYB5A,DDIT3,DHCR24,DLGAP5,FOXA1,HMMR,HSD17B8,INSIG1,ISG15,KIF2C,LRCH4,MAPK8IP3,MCM2,MCM3,NEDD9,OSER1,PBK,PGR,PSIP1,RNF40,SNRPD1,SOX9,SWAP70,TNFSF13,TOP2A,UBR7
IRF5	-2,296	transcription regulator	Activated	3,206	0,000836	CMPK2,CXCL10,CXCL2,CXCR4,DDX58,IFI44,IFIT1,IFIT3,IRF5,IRF7,ISG15,NAMPT,PRKRA,RSAD2,STAT1,STAT2,UBE2L6
ECSIT	-1,026	transcription regulator	Activated	2,942	0,00148	BCL3,CD83,IRF7,NFKBIA,NFKBIE,PIM1,RELB,SOD2,TNFAIP3
SMARCB1	-1,466	transcription regulator	Activated	2,863	0,00000201	ACAT1,ACTR2,AURKA,BAG3,BNIP3,BTG1,CCNA2,CDC6,CDK1,CDKN2C,CIDEB,COL1A2,CXCR4,DSC2,F10,F11R,FASTK,FOS,GJB1,HES1,HP,KIF11,LBP,MCM2,MCM3,MX1,MYC,PGLYRP1,PLK4,POLA1,PRKAB2,RAB3B,RAD51AP1,RFC5,SKP2,SMARCB1,SMC4,SPARC,TP53
RELA	1,089	transcription regulator	Activated	2,856	0,00000822	ACTA2,ACTN4,AHR,ALCAM,APOE,APP,ARFGAP3,ARHGDIB,B2M,BCL3,BEX2,BLVRA,BTG2,CD14,CD40,CD59,CD69,CFB,CHI3L1,CITED2,COL1A2,COL4A3BP,CSN2,CTGF,CTSB,CXCL10,CXCL2,CXCR4,CYBB,DDIT3,DUSP1,EPAS1,ERAP1,FOS,FSCN1,GBP1,GRK5,HES1,ICAM1,IL10,IRF7,ISG15,JUN,KIT,LUM,MST1R,MT-CYB,MYB,MYC,NAMPT,NFKBIA,NFKBIB,NFKBIE,NOS2,ORAI1,PRDX6,PRKCD,RELB,SAA1,SOD2,SOX9,STIM1,SYTL1,THOC1,TLR2,TNFAIP3,TP53,TWIST2,VIM,XIAP
CREB1		transcription regulator	Activated	2,819	0,0000398	ABCA1,ADORA2A,APOE,ATF3,ATP2B4,ATP6V0D1,AURKA,BAG3,BCAP29,BHLHE40,BNIP3,BTG2,CADPS2,CARS,CCNA2,CCNB1,CCNB2,Ccnjl,CGREF1,CLMP,CNN1,COCH,CRYM,CSRP2,CXCL2,CXCR4,CYP51A1,CYR61,DCX,DGKA,DUSP1,ENTPD1,ESD,FGF13,FGL2,FOS,GADD45B,GADD45G,GDF11,GLA,GLS,HLA-DQB1,HLA-DRA,HMGCS1,HSD11B1,IRF7,JUN,KDELR3,KIAA1549L,LAPTM4B,LCN2,LGALS8,LITAF,LMO1,LSM11,LSS,MCL1,MEST,MRPS18B,MSMO1,MVK,MYC,NFIL3,NNAT,NOS2,NPC2,NREP,NRP1,PCNA,PCSK1,RAD54L,RBMX,SCD,SEC63,SIAH2,SLC16A1,SLC19A1,SLC35G2,SOD2,SQSTM1,STAC2,STAT3,STMN1,SULT2B1,TFAP2A,TINF2,TNFSF11,TP53INP2,VIM
STAT1	1,642	transcription regulator	Activated	2,797	0,00000982	ABCA1,ALDH1A3,ANGPT2,APOE,B2M,BATF2,BCL6,BTG1,CASP2,CASP4,CD14,CD40,CDK2,CFB,CLIC5,CMPK2,CSN2,CXCL10,CXCL2,FCER1G,FCGR2B,FGL2,FOS,FURIN,GBP1,HES1,HTRA1,ICAM1,IFIT1,IFIT3,IL10,IL15,IRF5,IRF7,IRF9,ISG15,JAK2,JUN,LCN2,LY96,MX1,MYC,NOS2,PARP9,PIM1,PSME1,PSME2,RSAD2,SMARCB1,STAT1,STAT2,TAPBPL,TLR3,TLR4,TP53,UBD,USP18
CEBPA		transcription regulator	Activated	2,786	0,0000205	ABCA3,AKR1B1,ANPEP,ARL6IP5,BTG1,BTG2,CA2,CCNA2,CCNB2,CD14,CFD,CHI3L1,COL1A2,CRABP2,CREG1,CSN2,CTSK,CXCR4,CYP7A1,F8,FOS,FOXM1,FOXO3,G0S2,GBP1,GGH,GLRX,GRHL3,GSTP1,H1FX,HGF,HP,HSD11B1,ICAM1,ID2,IL10,ISG15,JUN,KCNMB1,KRT14,LCK,LCN2,LITAF,LTF,MT-CO1,MYC,MYCN,NFATC2,NFIL3,NRP1,OXTR,PCNA,PLXND1,PPARA,PPARD,PTGS1,PTPRC,S100A9,SAA1,SCAP,SCD,SERPINE1,SMPDL3A,SOD1,SOD2,SPP1,TFAP2A,TNFRSF1A,UBE2I,VCAN,VDR,ZBTB48,ZMIZ1
ID3		transcription regulator	Activated	2,777	0,00132	BCL3,BCL6,CCNB1,CCNE2,CCNG1,CD72,CD83,CDC6,CDK1,CDKN2C,CXCR4,DUSP1,FCER1G,FOXO3,GADD45B,GADD45G,GNL3,HOMER2,ICAM1,IL10,IL10RA,IL17RB,IRF5,ITGB3,JAK2,MAP3K14,MYB,MYC,RPS6KA2,TNFRSF1A,TNFSF11,TP53,ZBTB16

EPAS1	2,056	transcription regulator	Activated	2,586	0,000000224	ANGPT2,ANGPTL4,ATG5,BHLHE40,BNIP3,C1QA,CA9,CAT,CCR5,CDCP1,CHMP2B,CITED2,CKM,CKMT2,CLDN1,CTGF,CXCL2,CXCR4,CYP51A1,DDIT3,FOS,GADD45B,GCHFR,GLS,HMGCS1,ITGAV,ITGB3,LOX,MAFF,MANF,NFIL3,NOS3,NRARP,OMA1,PAN2,RB1CC1,SCAP,SERPINE1,SLC29A1,SOD1,SOD2,SOX9,STC2,TEK,TMEM45A,TNFAIP3
FOXO3	4,974	transcription regulator	Activated	2,573	0,000000116	BNIP3,CAT,CCNB1,CCNE2,CLDN1,CTGF,CTSV,CXCL10,CYR61,DDIT3,FOS,FOXM1,FOXO3,FOXO4,GABARAPL1,GADD45B,GLUL,GTF2I,IL10,IMPDH2,LCN2,MAX,MXD1,MXD3,MYC,NAMPT,NFKBIA,NOS2,NOS3,PAK1,PRDX3,PRDX5,PRNP,Rcan1,SESN1,SH2B3,SKP2,SOD1,SOD2,TNFRSF1A,UBE2C,VIM,YBX1
HIF1A	-1,027	transcription regulator	Activated	2,523	2,08E-08	ACTA2,ANGPTL4,ANKRD37,APOE,AQP9,ATG9A,ATP7A,AURKA,BHLHE40,BNIP3,BRCA1,CA9,CARS,CCR5,CDCP1,CITED2,CLDN1,CTGF,CTPS1,CXCL2,CXCR4,CYB5A,CYR61,EMC9,EPAS1,FOS,FSCN1,FURIN,GADD45B,GCHFR,GHR,GLYR1,HES6,HK2,HP,ID2,IL10,ITGAV,ITGB3,JUN,KRT14,LOX,MAFF,MANF,MCL1,METTL23,MITF,MST1R,MT-CO3,MT-ND1,MYC,NOS2,NOS3,NPM1,NRARP,P4HA1,P4HA2,PDK1,PKM,PPARA,PROM1,R3HCC1L,SERPINE1,SLC29A1,SLC39A7,SOD2,SOX9,SP1,ST3GAL1,STAT3,STC2,TLR2,TMEM19,TMEM45A,TP53,TPI1,VIM
PDX1		transcription regulator	Activated	2,523	0,00195	ACTG1,AKR1B1,AKR7A2,ATF3,ATG9A,CAMK2N1,CAT,CRELD2,CROT,CXCL2,CXCR4,DUSP5,FH,GALNT18,GRN,ID2,IDH1,INSIG1,JUN,KAT7,LTA4H,MAN1A1,MAOB,MAP2,MITF,MT-ND1,MYC,NQO1,PCNA,PCSK1,RSAD2,SPP1
IRF1	-1,049	transcription regulator	Activated	2,51	0,000000224	B2M,CASP2,CCNB1,CD40,CDK2,CFB,CXCL10,CXCL16,CXCL2,CYBB,EIF4A3,ERAP1,FGL2,IFIT1,IFIT3,IL10,IL15,IL18,IRF5,IRF7,IRF9,ISG15,JAK2,MX1,MYB,MYC,NOS2,ODC1,PCNA,PIGR,PLA2G16,PSME1,PSME2,RSAD2,SP1,STAT1,STAT2,STAT3,TLR3,TP53
CREM	1,344	transcription regulator	Activated	2,457	0,00144	ABCA1,ACTC1,ANXA4,APOE,ATF3,BHLHE40,BTG2,CCNB1,CRABP2,CYP51A1,DUSP1,FOS,GADD45B,HLA-DRA,HMGCS1,LSS,MCL1,MEST,MSMO1,MVK,NFIL3,NOS2,NPC2,PCSK1,Pln,RYR2,SIAH2,SLC16A1,SMC4
TCF7L2		transcription regulator	Activated	2,438	0,0000136	ACAA1,ACADL,ACSL3,ADIPOR2,APOD,AQP9,ARAP2,BMP4,CAMK2N1,CARHSP1,CCNYL1,CREB3L2,CSRP1,CTGF,CTNNAL1,CYP51A1,CYP7A1,DHCR24,DHRS7,DRAM2,ENTPD5,EPAS1,EPCAM,EPS15,ERBIN,EVI2B,FH,G2E3,GJB1,GLTP,GLUL,GRAMD3,ID2,IDH1,IL10,KAT2B,LRRN1,MAN1A1,MTMR2,MYC,MYO6,NKAIN1,NPC1,NPC2,PIGA,PPP1R16B,PRRG1,RALGDS,RCBTB1,RHOA,RNASE4,SDC2,SOX10,SPP1,STK17B,STRN,TBC1D14,TMEM125,TSPAN2,TWF1
TCF3	-1,449	transcription regulator	Activated	2,437	0,000185	ACACB,ARSA,ATF3,AURKA,AZGP1,BCL6,CA2,CCNA2,CCNB1,CCNB2,CCNE2,CDC45,CDKN2C,CGREF1,CKM,CTSV,GADD45B,HS3ST1,ID2,JUN,KIF11,KIF2C,KIF4A,KIT,KLHDC2,MAD2L1,MSMO1,MYC,MYCN,NFIL3,PLK4,PTGS1,RACGAP1,RASSF4,RPS3A,SCIN,SEMA3G,SULT2B1,TCF3,TMEFF1,TOP2A,XRCC6
MXI1		transcription regulator	Activated	2,397	0,000104	CCNB1,FOXM1,ID2,IFT20,IMPDH2,MYC,MYCN,ODC1
GATA1		transcription regulator	Activated	2,395	0,000592	ANGPT2,BTG2,CA2,CALR,CD36,CDK2,CDKN2C,CDKN2D,CYBB,CYFIP1,F10,FYB,HHEX,ITGB3,KIT,LYZ,MCM3,MITF,MSH2,MYB,MYC,MYCN,NCL,NFIL3,PCNA,PIM1,POLA2,POLD1,RAD54L,RFC4,RPL22,SCIN,SLC19A1,SLC50A1,SPTB,SRC,SRM,TEK,TOP2A,UAP1,USP25

NFKBIA	1,874	transcription regulator	Activated	2,35	1,59E-08	AMPD3,AURKB,AZGP1,BCL3,BNIP3,BRF2,BTG2,CASP4,CCNA2,CCNB1,CCNE2,CD40,CD69,CDC45,CDC6,CDK2,CDKN2D,CHI3L1,CHI3L2,COL1A2,CP,CRYAB,CSK,CTNNB1,CTSB,CTSZ,CXCL10,CXCL2,CXCR4,DAG1,DDIT3,ERAP1,FGFR4,FOS,FOXM1,FSCN1,GADD45B,GADD45G,GRK5,GRN,HES1,HK2,ICAM1,IL10,IL15,IRX3,ISG15,ITGA2,ITGAV,JUN,LCN2,LITAF,MR1,MT-CO3,MT-CYB,MYC,NBR1,NFKBIA,NFKBIB,NFKBIE,NOS2,OGN,PCNA,PIM1,RABEP2,RELB,RGS4,RPS18,RPSA,S100A9,SCFD1,SDC1,SOD1,SOD2,SOD3,SOX9,TFCP2L1,TIMP2,TLR2,TLR4,TNFAIP3,TP53,TRADD,USP9X,XIAP,XRCC6
ATF4	1,621	transcription regulator	Activated	2,289	0,00534	ATF3,CA9,CALR,CANX,CSN2,CTNNB1,DDIT3,DDR2,HSP90B1,IGFBP7,JUN,KLF9,LALBA,LGALS3,MAP1LC3B,MCL1,MID1IP1,NR1H3,OSMR,PCK2,SLC1A5,STAT3,STC2,TNFRSF12A,TNFSF11
SREBP2		transcription regulator	Activated	2,262	0,0000324	ABCA1,CYB5A,CYP51A1,DBI,FADS2,FDPS,GTF2I,HES6,HMGCS1,IDH1,INSIG1,LSS,MSMO1,MVK,NSDHL,RDH11,SCD,SQLE,STARD4,TM7SF2
RBL1		transcription regulator	Activated	2,208	0,000199	AURKB,CASP4,CCNA2,CDC6,CDK1,CDK2,FOS,HES1,MCM2,MCM3,MTOR,MYC,MYCN,NEK2,ORC1,PCNA,PPP1R8,SKP2,TP53
NFKB1		transcription regulator	Activated	2,204	0,00000292	ADORA1,AKR1B1,APOE,APP,B2M,BAMBI,BCL3,BTG2,CD40,CD59,CFB,CHI3L1,COL1A2,CREB3,CSN2,CTSB,CXCL10,CXCL2,CYBB,DUSP1,FANCD2,FOS,FSCN1,GBP1,GRK5,ICAM1,IKBKG,IL10,IL18,ISG15,MAP3K8,MCL1,MYB,MYC,NFKBIA,NFKBIB,NOS2,ORAI1,PRKCD,RACK1,RELB,SAR1A,SOD2,SOX9,STAT1,STIM1,SYTL1,TLR2,TNFAIP3,TP53,XIAP
SMARCA4		transcription regulator	Activated	2,161	0,00000259	ABCA1,ACTA1,ACTA2,ACTN4,AHR,AIM1,ALDH1A3,ALDH2,ARHGDIB,ASCC1,ATG9A,ATP2B4,AZGP1,BAMBI,BMP4,C15orf52,C1orf54,CCNA2,CCNE2,CDC6,CDK2,CDKN2D,CHODL,CKM,CLDND1,CNTN1,CP,CTGF,CTSB,CXCR4,DES,DYSF,ENTPD3,EPHB2,FADS3,FBLIM1,FCHSD2,FMO2,FOS,GADD45G,GBP1,GCHFR,GCLC,GSTO1,GSTP1,HLA-DRA,HS3ST1,ICAM1,IFI30,IFIT1,IGFBP7,ITGAV,JUN,KCNJ2,KIT,LAMA3,LDB1,LGALS3,LOX,LRAT,LUM,MAFF,MAOB,MT-ND2,MXD1,MYB,MYC,NCALD,NFKBIZ,NPC2,NR2F2,NRIP3,NRP1,PAEP,PDK1,Pdlim3,PLPP3,RAC2,RBM4B,SDC2,SERPINE1,SLC35G2,SOD3,SPP1,SRPX,ST3GAL1,STARD10,TAGLN,TAPBPL,TIMP2,TLR2,TMEM171,TNFSF13,TUBB,TWF1,TXNRD1,UBD,UNC13D,VIM
STAT2	1,684	transcription regulator	Activated	2,147	0,000299	CXCL10,GBP1,GHDC,IFIT1,IFIT3,IL10,IRF5,IRF7,IRF9,ISG15,MX1,RSAD2,SF3A1,USP18
E2F6		transcription regulator	Activated	2,111	0,0000131	BRCA1,CDC45,CDC6,DCTN4,DHPS,GINS2,LIG1,MCM2,MCM3,MYC,PCLAF,POLA2,RAD51,RAD51AP1,RBBP4,RFC4,RPA2,RYR2,SERPINE1
TOB1	1,099	transcription regulator	Activated	2,111	0,0099	CCNA2,CDK2,HJURP,MBNL2,NBR1,PCBP4,SPDL1,TMED7,TP53,UBE2T,WDHD1
PPARGC1B		transcription regulator	Activated	2,062	0,00568	ABCA1,ACACB,ACADL,ACADM,DHCR24,FDPS,HK2,LSS,MITF,MVK,PDK4,SCD,SQLE

MEF2D		transcription regulator	Activated	2,03	0,0012	ACTA1,CCNE2,CDC6,CDK2,CKM,COL1A2,CTGF,FOS,JUN,MCM3,PCNA,PPARA
CCNE1	1,162	transcription regulator	Inhibited	-2,138	0,00143	BRCA1,CCNA2,CCNB1,CDC45,CDC6,MCM2,PCNA,TP53
CCND1		transcription regulator	Inhibited	-2,274	8,48E-12	ALDH1A3,AURKA,AZGP1,BRCA1,C7,CCNA2,CCNE2,CDC45,CDC6,CDCA7L,CDK2,CDKN2C,CENPH,CENPK,CEP55,CLSPN,CPED1,DHCR24,DONSON,DTL,E2F7,E2F8,ESCO2,FAM83D,FOXM1,GAS2L3,HECTD2,HJURP,HOMER2,HSPB8,ITGAV,ITGB3,KIF11,KIF20A,KIF20B,KIF2C,KIF4A,KLHL24,KRT14,MAFF,MELK,MESDC2,MFSD6,MTMR1,MYC,PBLD,PCLAF,PCNA,PGR,PSMC3IP,PSRC1,PTPRC,RAB3B,RACGAP1,RAD51,RFC5,RMI2,SPC25,SPP1,STARD4,STXBP1,TM7SF2,TP53,TP53INP1,TP53INP2,TSC2,UAP1,ZNF367
E2F3	-1,447	transcription regulator	Inhibited	-2,463	0,000479	ARPC1A,C1orf198,CCNA2,CCNB1,CCNB2,CDC45,CDC6,CDCA3,CDK1,CDK2,DAG1,HOXB9,MAD2L1,MAL2,MCM2,MCM3,MYB,MYC,MYCN,ORC1,PCLAF,PCNA,POLA1,POLA2,PPP1R8,PTTG1,RAD51,SERPINE1,TOPBP1,UBE2C
TRIM24		transcription regulator	Inhibited	-2,623	0,000074	CA2,CMPK2,CNOT6,CSRP1,CXCL10,DDX58,GLUL,IFI44,IFIT3,IRF7,IRF9,ISG15,JAK2,LGALS3,PCLAF,PRPS2,SERPINE1,SPP1,STAT1,STAT2,TLR2,TRIM6-TRIM34,UBA7,USP18
WT1		transcription regulator	Inhibited	-2,624	0,0000606	AHCY,AMHR2,ANPEP,BTG2,CDC45,CDKN2C,CHAF1B,CIRBP,CKM,COL4A1,CTGF,CTNNB1,CTSV,CXCL10,FDPS,GSR,HSP90B1,IL10,LGALS3,LMAN1,LSS,MCL1,MSLN,MYC,MYCN,NCSTN,NUCB1,ODC1,PCK2,PDIA4,RPL19,SDC1,SERPINE1,SLC35G2,SOX9,SQLE,SQSTM1,TFAP2A,TRAP1,TSPAN5,TYROBP,VDR,YBX1,ZMIZ1,ZNF7
E2F1		transcription regulator	Inhibited	-2,705	1,84E-20	ACADL,ADIPOR2,ANGPT2,AURKA,AURKB,BMP4,BNIP3,BRCA1,BUB3,CA2,CALD1,CALR,CASP9,CCNA2,CCNB1,CCNB2,CCNE2,CCNF,CCT2,CCT4,CDC20,CDC45,CDC6,CDK1,CDK2,CDKN2C,CDKN2D,CITED2,COPS8,CRABP2,CRYAB,CTNNB1,CTSB,CYB5A,DDIT3,DUSP1,E2F8,EIF3I,ERH,Esrra,EXOSC9,FANCD2,FOS,FOXM1,FOXO3,HELLS,HES1,HNRNPK,HNRNPR,HSP90B1,HSPD1,HSPE1,ICAM1,KIT,KRT14,LACTB,LCK,LTA4H,MAD2L1,MAP3K14,MAPK14,MCL1,MCM2,MCM3,MMP16,MSH2,MTBP,MYB,MYC,MYCN,NCL,NDUFC1,NFKBIA,NFKBIB,NRP1,NUDC,ORC1,PCLAF,PCNA,PDCD5,PDK1,PDK4,PHB,PHC1,PLK2,POLA1,POLA2,POLD1,PPP1R8,PRPS2,RACGAP1,RAD51,RAD54L,RAN,RBBP4,RFC4,RFC5,RHOQ,RPA2,RSL1D1,SERPINE1,SMARCB1,SMC4,SOD2,SP1,SRPRA,STK17B,STMN1,TCF3,TOP2A,TOPBP1,TP53,TP53INP1,TRAP1,TXNRD1,UCHL5,VIM,VRK1,ZNF672
RCAN1		transcription regulator	Inhibited	-2,752	0,00109	ACTA1,CCNA2,CCNF,CD36,FOS,ICAM1,MYH1,NOS3,SOD1
MYB	-12,288	transcription regulator	Inhibited	-2,85	0,00165	ANPEP,BRCA1,CCNB1,CDK1,COL1A2,COL4A1,COPA,CXCR4,ERBIN,JUN,KIT,MYB,MYC,PCNA,POLA1,SLC1A5,SLC25A3,SPP1,VAV1,VIM
MITF	2,231	transcription regulator	Inhibited	-3,059	2,07E-11	AIM1,ALCAM,APOE,ATP6V1C1,AURKB,BRCA1,CCNB1,CCNF,CD151,CDCA3,CDK2,CENPH,CENPO,CEP55,CHAF1A,CTSK,DSN1,DSTYK,ESPL1,FANCA,FMOD,FOS,GPRIN1,HAUS8,HES1,ITPKB,KIF20A,KIF4A,KIFC1,KIT,LGALS3,LIG1,MCM2,MICAL1,MITF,NCAPD2,NUF2,PIF1,POLD1,POLE2,PSEN2,PSMC3IP,RAB27A,RFC5,RHOQ,RRAGD,SDC1,SEMA6A,SERPINE1,SLC7A8,SNW1,SOX10,SOX9,SPAG5,SPC25,STXBP1,TACC3,TFAP2A,TMEM251,TP53,UBE2C,VAT1

E2F2		transcription regulator	Inhibited	-3,13	0,00000963	CCNA2,CCNB1,CCNB2,CDC45,CDC6,CDK1,CDKN2C,CDKN2D,MCM2,MCM3,MYB,MYC,MYCN,ORC1,PCNA,POLA1,RAD51,SERPINE1,TOPBP1,TP53
NKX2-3		transcription regulator	Inhibited	-3,201	0,000183	ANGPT2,ANGPTL4,ANKRD37,ARHGDIB,BATF2,BMP4,BTG1,C19orf66,CCNB2,CD36,CEP55,CMPK2,CRYAB,CXADR,CXCL16,DDX58,F2RL1,GALNT15,GBP1,GHR,HMMR,HNRNPA0,MAP2,MYD88,MYO5A,NOS3,NR2F1,PARP10,PARP14,PARP9,RPL23,SAMD9,SRPX,STAT1,STAT2,TXNDC12,UACA,UBA7,UBE2L6,USP18,ZC3HAV1
TAL1	1,003	transcription regulator	Inhibited	-3,282	0,00304	ARSA,AZGP1,BCL6,CCNB1,CD69,CENPU,DSCC1,GALNT7,GUCY1A3,HELLS,ID2,IL10RA,JUN,KIF20A,KIT,LMO1,MAP2,MAP3K1,MCM2,MELK,MEST,MSLN,MYB,MYC,NCAPG,NFKBIZ,RASD1,RIPK4,RPS3A,TMEFF1,TNFAIP3,UBXN1,XRCC6
FOXM1	-3,416	transcription regulator	Inhibited	-3,675	0,0000112	AURKB,CCNA2,CCNB1,CCNB2,CCNE2,CCNF,CDC20,CDK1,CDK2,CTNNB1,CYP7A1,FOS,FOXM1,KIF20A,LOX,MYC,NEK2,Nes,PCNA,PLK4,PROM1,SKP2,STAT3,STMN1,TOP2A,TP53,TWIST2,VCAN,VIM
TBX2	-1,445	transcription regulator	Inhibited	-4,537	1,36E-09	ASF1B,ATF3,AURKA,AURKB,BHLHE40,CCNA2,CCNB1,CCNL1,CDC6,CDCA3,CDCA5,CDK1,CDKN2C,CHAF1B,CKAP2,DDIT3,E2F7,E2F8,FOXM1,HELLS,LIG1,MAD2L1,MCM2,MXD3,NCAPD2,NFIL3,PKMYT1,SEPT10
MYC	-3,001	transcription regulator	Inhibited	-5,543	8,68E-23	ABCA1,ABCC3,ABCD1,ACAT1,ACSS1,ACTA1,ACTN4,ACVR1,AHCY,ALCAM,ALDH18A1,ANGPT2,ANXA4,ANXA5,APP,ARHGAP25,ATPIF1,AURKB,BCL6,BRCA1,CAMK2N1,CANX,CASP9,CCNA2,CCNB1,CCNB2,CCNE2,CCT3,CD151,CD47,CD69,CDC20,CDK1,CDK2,CHRNB1,CHST15,CITED1,COL1A2,COL4A1,COL4A2,CRABP2,CRYAB,CSRP2,CTNNB1,CTSB,CTSV,CXCL10,DBI,DCTPP1,DDIT3,DKC1,DNPH1,DSP,DUSP1,DUSP5,EEF2,EIF3D,EPCAM,ERAP1,EXOSC7,EZH1,FADS2,FAM129A,FAP,FBL,FMOD,FOS,FOXM1,FRZB,GADD45B,GADD45G,GAMT,GCLC,GFPT1,GGH,GGT1,GLG1,GLS,GLUL,GLYR1,GOLGA2,GPC1,GRHL3,GSR,H2AFZ,HERC5,HES1,HK2,HNRNPA1,HSPD1,HSPE1,ICAM1,ID2,IDH1,IDH2,IFIT1,IL10,IL17RB,IMPA2,IQGAP2,IREB2,IRF7,IRF9,IRX3,JUN,KAT2A,KRT14,LOX,LRRN1,LUM,LYZ,MAD2L1,MAN2A1,MAX,MCL1,METAP2,MGAT1,MITF,MOGS,MRE11A,MSH2,MTBP,MYC,MYCN,MYL9,MYO1C,NCL,NFKBIA,NOLC1,NOP56,NPM1,NQO1,NRP1,NUCB1,NUDC,ODC1,PAICS,PAK1,PARP1,PCNA,PDK1,Pdlim3,PFAS,PHB,PHB2,PKM,POLD1,POLR1B,POLR2G,PPARA,PPARD,PPAT,PRDX3,PRMT1,PROM1,PTBP1,PTPRC,RAB40B,RAD51,RARG,RBBP4,RBBP7,RHOA,RHPN1,RPL13,RPL19,RPL22,RPL23,RPL26,RPL27,RPL3,RPL30,RPL35,RPL5,RPL6,RPS18,RPS19,RPS20,RPS6,RRM2B,RSAD2,SDCBP,SERPINE1,SKP2,SLC16A1,SLC1A5,SNRPD1,SNRPN,SOD2,SOX9,SPARC,SPP1,SRM,ST3GAL1,STMN1,SUMO2,SUMO3,TAF1D,TAT,TCF3,TIMP2,TLN1,TNFRSF12A,TNFSF11,TP53,TPI1,TRAP1,TSC2,TWIST2,TXNRD1,UBE2C,UBE2I,USP18,VDAC2,VIM,VPS72,XPO1,YBX1,ZFP36L1
MYCN	-3,43	transcription regulator	Inhibited	-5,901	1,46E-17	ABCA1,ABCA3,ABCC3,ABCD1,ACTG1,ACTN4,B2M,CITED2,COL4A1,CTGF,E2F5,EEF1G,EEF2,EZH1,FAU,HK2,HSPD1,ID2,IGFBP7,ITGA2,MAGT1,MRE11A,MYCN,NCL,NPM1,NUCB1,ODC1,PDIA4,PDK1,PHB,PSMB7,RBBP4,RBBP7,RNF11,RPL11,RPL12,RPL13,RPL18,RPL19,RPL22,RPL23,RPL26,RPL27,RPL29,RPL3,RPL30,RPL35,RPL37,RPL37A,RPL38,RPL4,RPL5,RPL6,RPS13,RPS17,RPS19,RPS20,RPS25,RPS26,RPS28,RPS3A,RPS4X,RPS5,RPS6,RPS8,SDC2,SERPINE1,SORD,SPARC,TAGLN,TIMP2,TMED9,TNFRSF1A,TP53,TPI1,TUBB,TUFM,VIM,ZFAND5,ZYX

Supplementary Table IV-S1. Summary of PIA enrichment analysis (adjusted p-value ≤ 0.05).

pathway_name	*pathway_ID*	*n_genes*	*all_genes*	*pvalue*	*padj*
Systemic lupus erythematosus	path:hsa05322	26	133	1.94E-08	6.40E-06
Allograft rejection	path:hsa05330	13	38	4.25E-01	7.01E+01
Antigen processing and presentation	path:hsa04612	16	77	1.62E+00	1.75E+02
Autoimmune thyroid disease	path:hsa05320	14	53	2.13E+00	1.75E+02
Type I diabetes mellitus	path:hsa04940	13	43	2.69E+00	1.78E+02
Graft-versus-host disease	path:hsa05332	12	41	4.34E+01	2.39E+03
Viral myocarditis	path:hsa05416	13	59	2.44E+02	1.15E+04
Asthma	path:hsa05310	10	31	1.43E+03	5.89E+03
Staphylococcus aureus infection	path:hsa05150	12	56	2.54E+03	8.87E+04
Herpes simplex infection	path:hsa05168	19	185	2.69E+03	8.87E+04
Intestinal immune network for IgA production	path:hsa04672	11	49	1.04E+04	3.13E+05
Phagosome	path:hsa04145	16	152	7.98E+04	2.19E+05
Alcoholism	path:hsa05034	17	180	1.24E+05	3.14E+06
Inflammatory bowel disease (IBD)	path:hsa05321	11	65	2.62E+05	6.18E+06
Cell adhesion molecules (CAMs)	path:hsa04514	15	144	3.27E+05	7.19E+06
Epstein-Barr virus infection	path:hsa05169	17	201	6.76E+05	1.39E+07
Th1 and Th2 cell differentiation	path:hsa04658	11	92	1.12E+07	2.17E+08
Leishmaniasis	path:hsa05140	10	74	1.34E+07	2.46E+08

Human T-cell leukemia virus 1 infection	path:hsa05166	16	219	1.54E+07	2.67E+07
Viral carcinogenesis	path:hsa05203	15	201	2.96E+07	4.89E+08
Th17 cell differentiation	path:hsa04659	11	107	5.34E+07	8.39E+08
Rheumatoid arthritis	path:hsa05323	10	90	8.73E+07	1.31E+09
Toxoplasmosis	path:hsa05145	11	113	9.28E+07	1.33E+09
Hematopoietic cell lineage	path:hsa04640	10	97	1.76E+08	2.42E+07
Influenza A	path:hsa05164	12	171	9.36E+07	0.000123590160476991
Tuberculosis	path:hsa05152	11	179	7.67E+09	0.000973268639997763
Human cytomegalovirus infection	path:hsa05163	10	225	0.00207695741895569	0.0253850351205695

ACKNOWLEDGMENTS

Because this may be my only opportunity to thank these individuals in writing, I apologize if I am a bit verbose in my thanks.

My supervisor, Mariasilvia D'Andrea deserves thanks for many things. Thank you so much for taking a chance on me, and for creating a supportive research environment which allowed me to perform my studies and activities. You have provided support at key moments, while allowing me to work independently much of the time. Without your guidance, this PhD would not have been achievable.

Within the research environment created by my supervisor, I have crossed paths with many professors and postdocs who have influenced and enhanced my research. The direction of my work has been strongly influenced by the members of the Mammalian NutriPhysioGenomics Laboratory group at the University of Illinois, notably J. Juan Loor and Mario Vailati-Riboni. My research would not have been possible without the aid and support of distinguished Italian professors that, through their writings and direct communication, provided me the benefits of their own valuable experience and helped my skills to flourish, notably Bruno Stefanon, Fabio Pilla, Giustino Gaspa, Marcello Mele, Marco Milanesi, Nicolo' Pietro Paolo Macciotta, Paolo Ajmone Marsan and Stefano Capomaccio.

To all of those whom I worked with during these past marvellous three years, I hope we have the chance to continue working together and know you will always be an inspiration to me.