2015-10-01

Mass Spectrometry Based Proteomics and Lipidomics Studies

Huan Kang
Brigham Young University

Mass Spectrometry Based Proteomics and Lipidomics Studies

Huan Kang

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Steven W. Graves, Chair
Beverly L. Roeder
John C. Price

Department of Chemistry and Biochemistry

Brigham Young University

October 2015

ABSTRACT

Mass Spectrometry Based Proteomics and Lipidomics Studies

Huan Kang
Department of Chemistry and Biochemistry, BYU
Master of Science

Mass spectrometry has emerged as having a vital role in various applications to biochemical fields. In this thesis, we have utilized a variety of mass spectrometry techniques for both bacteriophage proteomics and colostrum and milk lipidomics studies.

Our first study was the proteome characterization of Great Salt Lake bacteriophage NS01 with SDS-PAGE GEL to separate the viral proteins and high performance liquid chromatography (HPLC) coupled with an LTQ Orbitrap to identify the proteins after in-gel digestion. In this project, we have successfully identified 11 proteins with high confidence, p-values < 0.01, including coat protein gp88 with a coverage of 91% and tail protein gp86 with a coverage of 40.96%, which facilitated the classification of NS01 as a T7-like phage.

Our second study was the discovery of colostrum and milk biomarkers that can be used to predict the likelihood of development of production-related metabolic diseases (PRMDs) in dairy cows through a lipidomics approach. In this study, an electrospray ionization, time-of-flight mass spectrometer was applied to lipid profiling, quantification and significant biomolecule selection. A Q-Star quadrupole, orthogonal time-of-flight mass spectrometer and an Agilent 6530 accurate-mass quadrupole/time-of flight mass spectrometer were both used for lipid biomarker fragmentation and identification. According to linear discriminative statistical modeling, three panels of biomarkers were defined. A combination of 2 milk lipid predictors, including DG18:0/18:0 and TG 18:0/18:0/18:1, provided PRMD predictions with 75.0% sensitivity at 90.0% specificity. A combination of 3 colostrum lipid predictors, including TG16:0/18:1/18:3, DG16:0/16:0 and $C_{40}H_{60}NO$, provided PRMD prediction with 90.0% sensitivity at 86.4% specificity. Furthermore, a combination of 7 colostrum and milk biomarkers, including calculated differences between ‘shared’ markers found to be significantly different in both colostrum and milk, provided a predictive sensitivity of 87.5% at a specificity of 100%. Thus, three panels of lipid biomarkers have been discovered in 1-4 day postparturient dairy cow colostrum and milk that can be used to predict resistance or susceptibility prior to onset of clinically apparent PRMDs. These novel lipids could be used as important diagnostic predictors in the future. Therefore, mass spectrometry based proteomics and lipidomics approaches have been efficient tools in the biochemical research described in this thesis.

Keywords: proteomics, bacteriophage, lipidomics, mass spectrometry, biomarker, production-related metabolic diseases

ACKNOWLEDGEMENTS

I would never have come to accomplish my Master of Science degree without my three current committee members, Dr. Steven W. Graves, Dr. Beverly L. Roeder, and Dr. John C. Price, as well as Dr. John T. Prince and the substantial commitment of my many collaborators. My deepest thanks go to my advisor Dr. Steven Graves for supporting me and advising me the entire way and entrusting me with important tasks. Dr. Graves's support, knowledge, and expertise power this thesis in all aspects.

Both of the projects discussed in this thesis are a product of teamwork. First of all, for the bacteriophage NS01 study, thanks to Dr. David Belnap and David Eng for their preparation of the samples; Dr. Peter Shen, Dr. Ryan Taylor, Takuma Aoba for their support and guidance with the SDS-PAGE GEL protein separation and in-gel digestion which made it possible for me to fulfill my portion of that project. As for the second project, I want to express my eternal gratitude to Dr. Beverly L. Roeder for her idea to determine if first colostrum and milk lipidomic profiles can be used as biomarkers to predict resistance or risk for PRMDs in dairy cows. I am also indebted to Dr. Roeder for her supervision of several undergraduate students involved in the dairy cows study, the collection of colostrum and milk samples over a narrow window of time, the evaluation of animal health records, the management of animal data, and for providing student assistance in the lab to prep these samples for MS analysis. In particular, I wish to thank undergraduate students Holly Martin, Austin Cook, Kory Brown, Evan Buckmiller, and Alice Huang for their direct contributions to this work. Additional thanks go to Swati Anand, Chris Lau, Komal Kedia, Ying Ding and Bruce Jackson for their consistent assistance in the lab and to Dr. Dennis Eggett for his knowledge

and expertise in the statistical analyses. In addition, I would like to express my gratitude to Dr. John Price for his advisement and guidance in using mass spectrometers.

Last, but foremost, I would like to thank my parents for years of caring, support, and encouragement. My parents' personalities have always been an example for me and have given me a positive attitude and passion toward life.

Dedicated to:

My

Mom:

(Shuhua Li)

Dad:

(Guojun Kang)

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Mass spectrometry based lipidomics analysis

1.1.1 Lipid biomarkers overview

Lipids represent the largest and most diverse set of biomolecules. Their unifying characteristic is their hydrophobicity. Lipids have been known to have structural importance, but more recent research suggests many are biologically important active compounds mediating and regulating many cellular and systemic processes. Lipids are not directly genetically regulated. They are generated and metabolized by enzymes that are affected by the environment through such factors as nutrition and temperature. As is currently known, changes in lipids are seen in nearly all diseases states, including cardiovascular diseases, metabolic diseases like diabetes mellitus, obesity, Alzheimer's disease, arthritis, asthma, cancer, and importantly they may play significant roles in the pathophysiology of such diseases (Table 1). Useful biomarkers may result from downstream responses to the disease, but may also be upstream activators, regulators or general contributors to the disease. Therefore, specific lipids or groups of lipids may become clinically relevant biomarkers that can be used to specifically predict the risk for or can be used to diagnose particular diseases. Lipids have been researched as potential biomarkers in many diseases using various biological samples. For instance, the sphingolipidome has been shown to provide potential biomarkers for early-stage Alzheimer disease[1], type 1 diabetes[2], neurodegenerative disease[3], etc. Some lipids such as polyunsaturated fatty acids[4], triglycerides[5], glycerophospholipids[6], sphingolipids[6], and sterols[6] have participated in or been altered by metabolic disorders, and

Diseases	Lipids	Samples	Ref
Obesity	Ceramide, lysophosphatidylcholine, free fatty acids, TG	Plasma	[7], [8]
Diabetes	Lysophosphatidylcholine, free fatty acids, TG, cardiolipin	Plasma	[7], [9]
Mitochondrial dysfunction	Cardiolipin, TG and cholesterol esters	A549 cells	[10]
Alzheimer	Phospholipids, sphingolipids and related compounds, sulfatide	Serum, brain autopsy	[11], [12], [1], [13], [14]
Breast cancer	Phosphatidylcholine, total cholesterol, low-density lipoprotein cholesterol	Human breast tissue, plasma	[15], [16]
Ovarian cancer	Lysophosphatidic acid, lysophophatidylinositol, sphingosylphosphorylcholine, lysophosphorylcholine, lysophospholipids	Plasma	[17]
Asthma	Phosphatidylcholine, leukotriene B_4, prostaglandin D_2	Serum, mice	[18], [19]
Hepatic steatosis	Polyunsaturated fatty acids, glycerophospholipids, sphingolipids, sterols	Plasma	[4]

Table 1. Lipid abnormalities in associated diseases. TG: triglycerides

as such, are potential phenotypic or disease biomarkers.

Despite more research studies regarding protein biomarkers than lipid biomarkers, recent advances in lipid profiling, quantitative analysis and bioinformatics methods provide an opportunity to explore and define new roles for lipids and allow for them to be further characterized as biomarkers in order to predict or diagnose a range of diseases. In addition, in comparison with proteomic approaches, lipids are more stable, easier to handle, and may also provide a more economically efficient test for use in clinical practice than would protein mass spectrometry (MS) methods.

1.1.2 Mass spectrometry as a tool for lipidomics

By obtaining profile patterns of lipid molecules through MS analysis, metabolic changes can be observed under specific circumstances. Several such methods have been carried out including global[20], focused[21], and targeted[22] lipidomics using mass spectrometry.

Global lipidomic methods are applied to measure all or as many analytes as possible that are contained within extracted lipid samples without background information on any specific lipids or their fragments[23]. Put another way, this approach is unbiased and often considers hundreds of lipids, sometimes many more, in the discovery of biomarkers. Mass spectrometers with high resolution are always used for lipids profile studies to allow for the unambiguous resolution and measurement of all species observed. By combining liquid chromatographic separation with tandem mass spectrometry (MS/MS), fragmentation can be carried out, allowing for more information about the components that make up specific lipids, potentially identifying the specific lipid species. With the use of appropriate added

reference lipids, reasonably good quantification can also be determined.

A focused lipidomics approach aims to detect molecules of a particular lipid class by focusing on specific fragments or neutral losses between fragments that are highly associated with a single lipid subgroup[24]. Through this method, lipids that are present at abundances normally below the MS detection limit can also be detected, which helps to overcome MS ion suppression issues. Taking phospholipids as an example, product ion scanning at m/z 184 could be used to select for choline-containing phospholipids in the positive ion mode. Furthermore, neutral loss scanning at m/z 141, 185, 189 and 277 could be used for detection of phosphoethanolamine (PE), phosphatidylserine (PS), phosphoatidylglycerol (PG), and phosphatidylinositol (PI) species, respectively.

Targeted methods can be applied to choose one or a few specific molecules by precisely selecting m/z values for the lipid(s) of interest or by selection of its specific fragment ions. Through a targeted method, lipids can be either accurately quantified based their m/z values and their specific fragment ions or characterized by their fragmenting ions. The fragmentation information of the peak of interest can be obtained through specific product ion scanning as in other multiple reaction monitoring protein approaches.

In our experiments, we used a comprehensive or global lipidomics approach by means of electrospray ionization mass spectrometry (ESI-MS) to find peaks of interest (significant statistically, quantitatively different lipids) based on abundances between disease and control samples, and a targeted MS/MS method in order to fragment and further chemically characterize peaks of interest.

1.1.3 Electrospray ionization mass spectrometry (ESI-MS) as a tool for determining lipid identity

After the determination of lipids of interest, identification of the lipids is important in order to fully understand selected biomarkers and their role in biological functions. This was briefly discussed previously, but will be developed further in this section.

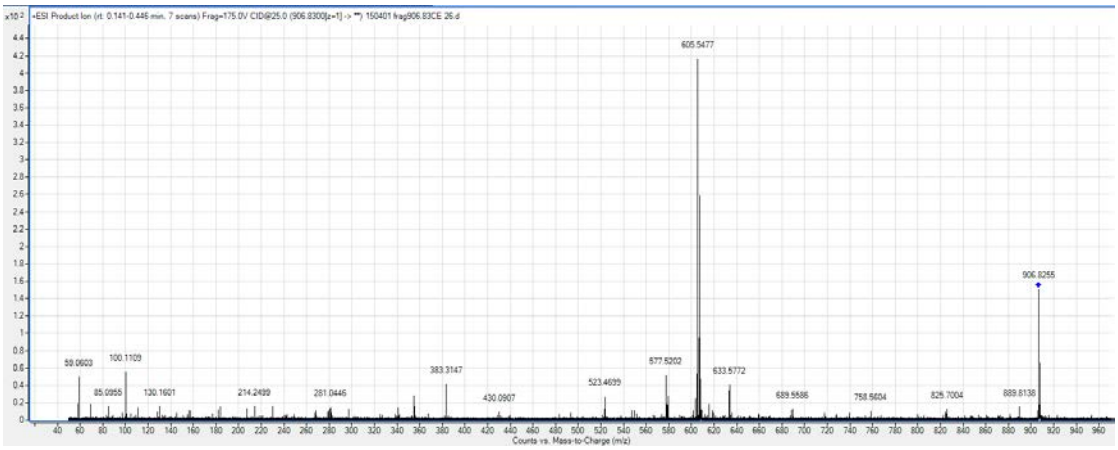
Tandem mass spectrometry involving ion-trap, triple quadrupole, or hybrid instruments such as quadrupole-time-of-flight, linear ion trap quadrupole are normally used for lipid identification with electrospray ionization as the ion source. Characterization of individual lipids using ESI-MS requires knowledge of the m/z value for the lipid of interest. Then fragmentation of the targeted peak or a target peak list based on m/z values is conducted through product-ion analysis after collision-induced dissociation (CID)[25]. By adjusting the collision energy[26], the abundance of the product ions of lower molecular mass will be different, which is useful for identification of the location of double bonds in fatty acyl chains. Optimal conditions for CID are determined by systematic investigations according to individual peak.

Multiple sector mass spectrometry instruments and MS-MS techniques can be applied and combined for lipid biomarker(s) characterization. For example, in our studies, both QSTAR and Qq-TOF[27] instruments were used for biomarker characterization. The QSTAR can select the targeted peak with low resolution, unit resolution and high resolution. However, it has a lower sensitivity and resolution power compared with Qq-TOF. Thus, the QSTAR may have a hard time detecting the peak of interest if it has a low abundance. As for the Qq-TOF, it can detect low abundance peaks.

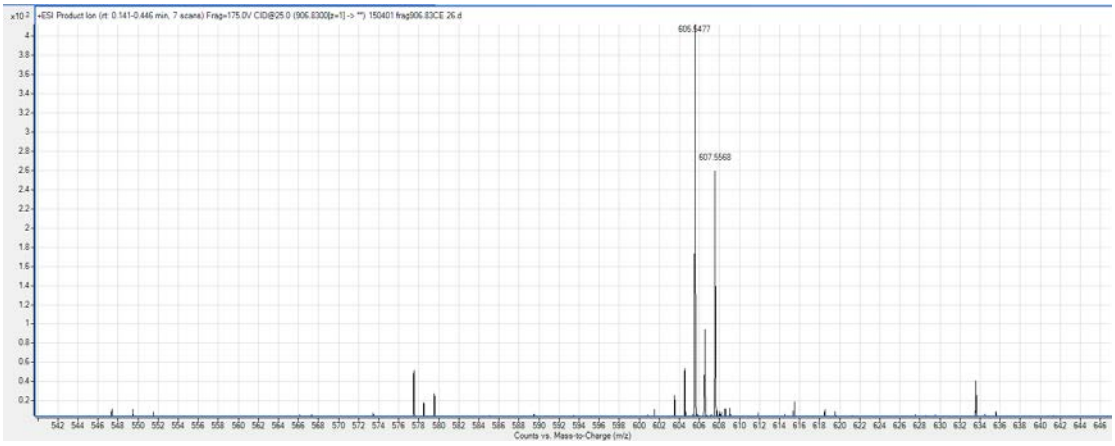
Extensive MS-MS characterizations of the most common lipid classes, including mass spectra, have been published in the literature. Taking triacylglycerols as an example, the sn-1, sn-2 or sn-3 component of triacylglycerols can be determined based on neutral loss values with fragmentation as recorded in the product-ion scan. Triacylglycerols are neutral hydrophobic lipids defined as fatty acid triesters of glycerol. Therefore, positive ions such as ammonium acetate, sodium, and lithium that can form positive adducts with neutral lipids are always added to the solvents in order to produce lipid ions that can be detected in positive ion mode of mass spectrometry.

Product-ion mass spectra of ammoniated triacylglycerol species ($[M+NH_4]^+$) at an m/z value of 906.83 after CID at 25 eV yielded product ions as shown in Figure 1, which can be informative for structural characterization. The fragment ion at $[M+NH_4\text{-}17]^+$ at 889.81, arising from the loss of ammonia (NH_3) from the ammoniated triacylglycerol species, was observed and always present in the spectra (Figure 1a). The fragments shown in the spectra with m/z values of 605.54 and 607.55 represent diacylglycerol (DG) fragments 18:0/18:1 and 18:0/18:0. By searching the LIPIDMAPS[28] database, the product ion scan provided the probable structure as TG 18:0/18:0/18:1 for precursor ion 906.83.

High resolution and mass accuracy are required for reliable identification of lipids. Fragment ions of polar headgroups or specific neutral losses from each class of phospholipids and glycerolipids are also very important for identification of each different category of polar lipids. In some cases, MS^3 or MS^4 (multiple steps of mass spectrometry selection with some form of fragmentation occurring in between the stages) may be important in order to obtain reliable identification.



(a)



(b)

Figure 1. Targeted MS/MS analysis of Qq-TOF mass spectrometry. (a) product-ion spectra of m/z 906.83 [TG (18:0/18:0/18:1)+NH_4]$^+$, (b) representation of the two most abundant diacylglycerol fragments at 605.55 and 607.56.

1.1.4 Comparison and overview of lipid extraction techniques

Due to different polarities, lipids dissolve in solvents based on the relative strengths of the interactions between the solvent and either the hydrophobic or the hydrophilic regions of the lipid molecules. For instance, lipids of low polarity, such as triacylglycerols or cholesterol esters, are very soluble in non-polar hydrocarbon solvents. Organic solvents are always used for lipid extraction from tissues, cells, and other biological fluid[29]. Here, the three main lipid extraction methods are discussed, including the Folch method, the Bligh and Dyer extraction, and a newer version of lipid extraction using methyl-*tert*-butyl ether[30].

First, according to the Folch[31] total lipid isolation method, lipids are extracted by homogenizing a tissue with 2:1 chloroform-methanol (v/v) solution, then the homogenate is filtered and the filtrate further purified with a 5 volume addition of water.

Second, a mixture of chloroform:methanol:water in the proportion of 1:2:0.8 and 2:2:1.8 (before and after dilution, respectively) is used for lipid extraction, particularly for phospholipids as described by Bligh and Dyer[31]. This method is recommended for large samples with a high proportion of endogenous water.

Third, for extraction by methyl-*tert*-butyl ether[30] (MTBE), 1.5 mL methanol is added to a 200 µL sample aliquot in a glass tube, and the tube is vortexed. 5 mL of MTBE is added and the mixture placed on a shaker for 1 hr at room temperature. Then, 1.25 mL of MS-grade water is added for full separation of the aqueous layer and organic layer.

There are reviews comparing the extraction efficiency between these distinct organic extraction methods[29]. For instance, a comparison of the Folch versus Bligh and Dyer method was done on marine tissues and meat products[32], which showed that the efficiency of

extraction was determined by the amount and percentage of lipids in the sample. Previous studies have proven that for samples containing more than 2% lipid, the Bligh and Dyer method significantly underestimated the lipid content, and this underestimation increased significantly with increasing lipid content of the sample[33].

For the MTBE method, the lipid layer is on top, which could help minimize carry-over of proteins. Another advantage of the MTBE lipid isolation method is that there is less toxicity in the extraction process by omitting chloroform as an organic solvent. Researchers have also made modifications based on the classifications of the lipids of interest. Different solvents are necessary to extract different lipids optimally. Plus, lipid extraction is carried out differently according to the nature of the biological specimen. For example, plant tissues are normally treated with isopropanol in order to denature the botanical enzymes prior to lipid extraction.

In our experiments, we used a modified Bligh and Dyer method with chloroform:methanol:isopropanol in a 2:1:1.25 ratio for lipid extraction from colostrum or milk samples as opposed to the standard method which uses chloroform:methanol:water in the proportion of 1:2:0.8 and 2:2:1.8 (before and after dilution, respectively). In addition, water was later added to allow for full separation of the aqueous layer and organic layer. Thereafter, 10 mg of the cream layer (the thick white or pale yellow fatty liquid that rises to the top when either colostrum or milk is left to stand, and composed of mainly lipids) was used for lipid extraction, which contained greater than 2% lipid. Therefore, it was necessary to increase the chloroform portion to two fold compared to the original Bligh and Dyer extraction solvent mixture in order to have a fuller recovery of

lipids.

1.1.5 Mass spectrometry as a predictive tool for production-related metabolic diseases (PRMDs) in dairy cows

Over the past 100 years, the United States dairy industry has experienced a sharp reduction in the total number of cows and at the same time a nearly six-fold increase in average milk production per cow, with a substantially greater overall annual milk production. This has led to dairy cows being at an increased risk of developing production-related metabolic diseases (PRMDs). PRMDs are due to certain cows' physiologic inability to cope with the metabolic demands of such high milk production. These health problems occur most often during the transition period (3 weeks pre-partum to 3 weeks after parturition), with the highest clinical incidence occurring within two weeks postpartum. Recently, PRMDs in dairy cows have been broadly defined to include conditions such as hypocalcemia, hypomagnesemia, ketosismana, hepatic lipodosis, abomasal displacement, laminitis, retained placenta, and other disorders that occur during the transition period.

Over the past 20 years, there has been great interest in the impact that PRMDs have on dairy farm profitability, and in understanding what contributing factors and/or predisposition for such diseases exist, to eliminate and minimize economic losses. Although it is known that the highest incidence for most PRMDs occurs within the first 60 days in milk (DIM), the multifactorial disease incidence has not been altered by transition diets, manipulating prepartum dietary cation anion balance[34], and over conditioning avoidance[35]. The nutritional and metabolic status[36,37] of certain animals, combined with a poor

physiologic adaption[38] to the negative energy balance experienced during the rapid rise in milk production with the onset of lactation, is associated with the development of PRMDs[39]. Although PRMDs usually effect a subset of cows within a dairy herd, the insidious nature and economic impact these health problems cause warrants a better means of earlier detection. Accurate biomarkers that could be easily applied to an animal substrate would provide an important diagnostic and management tool to predict individual cow resistance or risk for PRMDs while the animal is clinically normal. This method could be used during the peripartum period to provide an economical method to identify cows that are resistant or susceptible to PRMDs, for retention, breeding, early treatment intervention, or culling decisions to increase profit margins.

Biochemical analytes have been previously explored to predict the development of PRMDs in dairy cows including the measurement of serum non-esterified fatty acids (NEFA)[40], β-hydroxy-butyrate (BHBA) and calcium concentrations[41] during the final week of gestation or the first and second week postpartum, to provide useful information for herd health monitoring and culling risk[42]. The effectiveness of their predictive ability was summarized by Ospina et al (2010) as shown in Figure 2. Additionally, the serum concentrations of NEFA at calving have been shown to be positively correlated with a negative energy balance and the incidence of certain peripartum diseases[43]. Hyperketonemia in the blood during the first week of lactation, on the other hand, appeared to be an important risk factor for the subsequent development of clinical ketosis, metritis, and displaced abomasum[44]. Serum cholesterol concentration was shown to have a greater decrease during the transition period in cows that developed disease[45] PRMDs.

Threshold (mEq/L)	Sensitivity	95% CI for sensitivity	Specificity	95% CI for specificity	LR+[1]
Animals sampled prepartum (n = 1,440)					
0.2	64	49 to 73	48	45 to 51	1.2
0.27[2]	57	42 to 72	62	60 to 65	1.5
0.4	30	17 to 45	82	80 to 84	1.7
0.5	23	12 to 38	89	87 to 91	2.2
Animals sampled postpartum (n = 1,318)					
0.4	95	83 to 99	39	37 to 42	1.6
0.72[2]	80	65 to 91	73	70 to 75	3.0
1.02	59	42 to 74	87	85 to 89	4.6

[1]Likelihood ratio positive.
[2]Highest combined sensitivity and specificity.

(a)

Threshold (mg/dL)	Sensitivity	95% CI for sensitivity	Specificity	95% CI for specificity	LR+[1]
8	76	60 to 88	69	67 to 72	2.5
10[2]	71	55 to 84	80	77 to 82	3.5
12	63	47 to 78	86	84 to 88	4.6
14	51	35 to 67	90	88 to 91	4.9

[1]Likelihood ratio positive.
[2]Highest combined sensitivity and specificity.

(b)

Figure 2[42]. Prediction efficiency using NEFA and BHBA as predictors for transition diseases risk in dairy cows. (a) Information on thresholds from receiver operator characteristic curves for NEFA concentrations as predictors of displaced abomasum. (b) Information on thresholds from receiver operator characteristic curves for BHBA concentrations as predictors of displaced abomasum in animals sampled postpartum (n =1318). Results from Table 2 and Table 4 Ospina et al (2010). BHBA = beta hydroxybutyris acid, NEFA = non-esterified fatty acid. From Ospina et al (2010) used with permission.

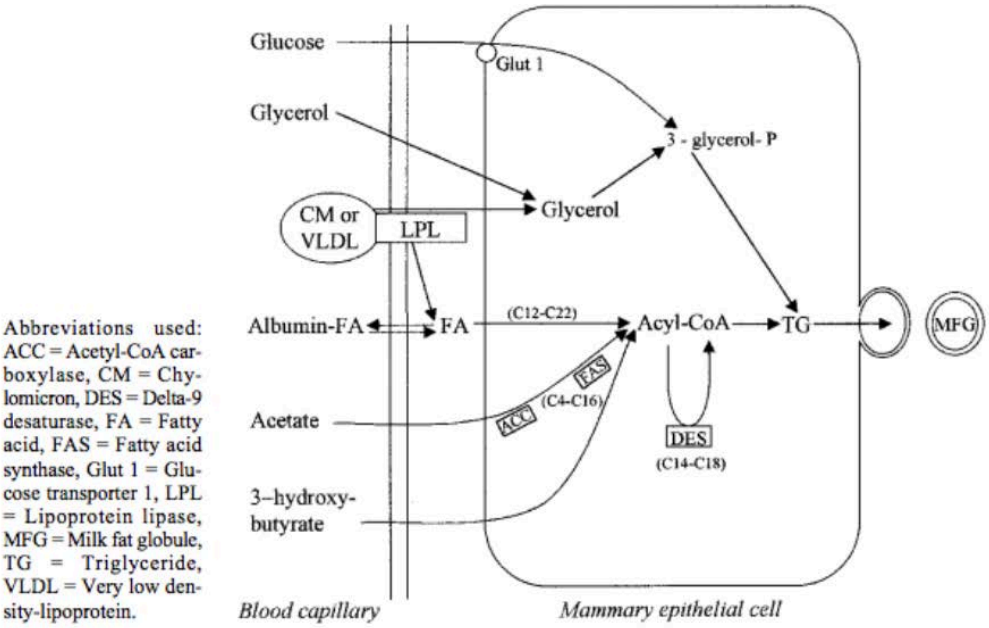


Figure 3. Schematic of milk fat synthesis and secretion in ruminants[46]. From Chilliard et al (2009) used with permission

Therefore, lipids, including NEFA, BHBA, and cholesterol, have shown some promise as biomarkers of PRMDs. The objective of the current research was to detect accurate biomarkers that can predict PRMD resistance or susceptibility, and was based on the presence of lipids in colostrum and early milk secreted by the cow's udder (Figure 3, from Ospina et al (2010), a schematic description of the milk fat synthesis pathway), to define the lipid status broadly through a mass spectrometry based lipidomics approach. With MS techniques, and may potentially be applied as biomarkers. Our findings were promising and may further suggest biological pathways involved in or altered by PRMDs.

1.1.6 Summary

In summary, sample preparation for mass spectrometry-based lipidomics analysis sample is easier compared to sample preparation for proteomics analysis. However, lipidomics fragmentation analysis is not as advanced as proteomics, and it can be challenging to do the lipid MS spectral analysis, to not only discover useful biomarkers, but to also achieve these markers' identification.

In this project, we focused on lipid biomarker discovery to identify peripartum dairy cows at risk for the later development of PRMDs using a lipidomics approach. If such markers could be found, given that they are present prior to development of PRMDs, they may allow therapeutic intervention for the prevention of PRMDs. Lipids in colostrum or milk samples were collected and evaluated for this study after extraction by the modified Bligh and Dyer method. Several individual candidate biomarkers were found in colostrum and milk. Statistical modeling of the individual markers resulted in 3 panels of biomarkers capable of

identifying a very high percentage of at-risk, as well as PRMD resistant animals. One set of biomarkers had 87.5% sensitivity and 100% specificity based on our data set. The sample collection was done by members of Dr. Beverly Roeder's lab. Among the undergraduate students Holly Martin, Kory Brown, Austin Cook, Alice Huang, and Evan Buckmiller from the Department of Biology and Department of Chemistry and Biochemistry (Brigham Young University) made substantial contributions to the collection, sorting, and processing of these samples, including optimizing the various processing protocols. Members of my lab (Dr. Steven Graves) worked on sample processing and instrument analysis and methodology. One of the undergraduate students, Chris Lau, helped write a micro program within Excel that facilitated our processing the MS data analysis. Dr. Dennis Eggett from Department of Statistics, Brigham Young University, helped us with statistical modeling to successfully identify different panels of biomarkers. Graduate students Swati Anand, Komal Kedia and Ying Ding from Dr. Steven Graves's lab advised and assisted with MS fragmentation and lipid identification of the relevant biomarkers. This proved to be a substantial project requiring both my committed efforts as well as the collaborative efforts of many others to accomplish it.

1.2 Mass spectrometry based viral proteomics study

1.2.1 Viral proteomics

Viruses have long been studied for their pathology, associated disease, as model systems for molecular processes and as tools for identifying important cellular regulatory proteins and pathways. Most viral research has been focused on positive and negative gene regulation, repressor operator interactions, DNA replication, transcriptional elongation and termination, etc. However, with the development of high-throughput proteomics methods, the protein composition of virions, the structure and protein interactions of viral proteins, and the effects of viral infection and individual viral proteins on the cellular proteome can be obtained.

Determining the global protein composition of a specific virus particle normally involves multiple steps including pure virus preparation, detergent denaturation, electrophoretic separation, and identification of protein bands by mass spectrometry where mass spectrometry is a valuable tool in structural and functional viral proteomics. Dennis Hruby's lab[47] has obtained a detailed knowledge of the Pox vaccine viral proteome through sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) + LC-ESI-Q-TOF MS, SDS-PAGE +LC-ESI-QIT MS, HPLC+ LC-ESI-QIT MS, LC-ESI-Q-TOF MS, MALDI-TOF/TOF MS analysis methods, and were able to identify 63 proteins. By using SDS PAGE and LTQ- Orbitrap, Kerstin Radtke's group were able to identify a total of 67 structural and nonstructural viral proteins of the Herpes Simplex Virus type 1 (HSV1)[48] and identified 90 novel phosphorylation sites and 10 novel ubiquitination sites on different viral proteins. Therefore, a comprehensive mass spectrometry based viral proteomics study has been proven to be

very effective. By comparison with complementary data from genome sequencing, cryo-EM and image reconstruction, a new dimension to viral protein structure and function can be obtained.

Mass spectrometry based viral protein studies have enabled a more comprehensive characterization of virions and of virus-virus and virus-host interactions involved in infection and pathogenesis. Viral proteomics has been a promising tool to survey antigenicity of the influenza virus[49] and for understanding virus tropism[50]. Moreover, the development of viral proteomics might also provide potent tools for diagnosis, and could generate novel insights into the therapeutics of virus-induced infectious diseases.

1.2.2 Summary

As is known, a virion, i.e. an entire virus particle, consists of an external protein shell called a capsid and having an inner core containing nucleic acid. The core confers infectivity, and the capsid provides specificity to the virus. Extremophilic viruses are the viruses that can tolerate environmental limits beyond what most living organisms endure in terms of temperature, pH, salinity, desiccation, hydrostatic pressure, radiation, anaerobiosis, etc. Thus far, several extremophilic viruses have been characterized. The continuous attempts to isolate and to study viruses that thrive in extreme environments are needed to address such questions as the origin, activity, or importance of the in situ microbial dynamics. However, this topic appears to open a new window on an unexplored part of the viral world.

In this project, three groups were involved, including Dr. Matt Domek's lab efforts in the collection and cultivation of the NS01 virions. Virion imaging and structural modeling studies

were achieved by Dr. David Belnap's lab, and our lab and specifically my efforts, were focused on NS01 protein characterization. In this thesis, I will talk about Dr. Domek's and Dr. Belnap's work for reference and clarity, but discuss my protein characterization studies in much greater detail.

.

Chapter 2 Sequence and structural characterization of Great Salt Lake Bacteriophage NS01, a turreted T7-like virus with flexible HK-97 subunits

2.1 Introduction

Studies of virus populations in the Dead Sea indicate that viruses in extreme environments play a predatory role in recycling organic matter and regulating bacterial host populations where eukaryotic predators are rare [51]. Thus far, six extremophilic viruses have been studied by cryogenic electron microscopy (cryo-EM): SH1 [52], STIV [53], STIV2 [54], CW02 [55], HVTV-1 [56], and HSTV-2 [56]. Four of the extremophilic viruses studied via cryo-EM to date have shown unusual turret appendages at each five-fold vertex; however, the purpose of such turrets remains unknown. Thus, the ubiquity of phages in aquatic environments [57] and the lack of structural data about those in extreme environments suggest the need and potential for more research in understanding extremophilic viruses.

Bacteriophage NS01 is a tailed, dsDNA extremophilic virus isolated from the Great Salt Lake and infects *Salinivibrio costicola*. Another halovirus, CW02, was recently reported to infect a *Salinivibrio costicola*-like bacterium, SA50, and characterized by cryo-EM [55]. CW02 is a dsDNA, tailed virus that belongs to the *Caudovirales* order and *Podoviridae* family of phages, as defined by its short, non-contractile tail and icosahedral head [55]. Interestingly, CW02 contains flexible turret structures at each 5-fold axis, much like other extremophilic viruses such as SH1 [52], STIV [53], and STIV2 [54].

The structural and sequence analysis of NS01 revealed its relationship to CW02 and other dsDNA, tailed phages. Negative stain microscopy images suggest that NS01 is morphologically similar to viruses of the *Podoviridae* family and that it also has turrets at the

5-fold vertices similar to those found in SH1 [52], STIV [53], and STIV2 [54] (Figure 2A). PSI-BLAST sequence analysis also showed a relationship between NS01 and Enterobacteria phage T7 [58] and T7-like phages such as Vibrio phage ICP3 [59], Vibrio phage VpV262 [60], Enterobacteria phage N4 [61], and Salinivibrio phage CW02 [55]. A 22 Å cryo-EM reconstruction of NS01 revealed T=9 icosahedral symmetry in the capsid head, prominent turret structures at each 5-fold vertex, and showed that the HK97-like fold fit within the capsid density, supporting the observation that dsDNA, tailed phages contain a common ancestral HK97-like fold.

2.2 Materials and methods

Cultivation and purification: Cultivation and purification were done in the laboratory of Professor Matt Domek. In brief, host bacteria and bacteriophage were isolated from the Great Salt Lake, Utah, USA. PCR was applied for amplification of host bacteria and approximately 1350 base pairs were sequenced (Idaho State University, Molecular Research Core Facility, Pocatello, Idaho, USA) and queried against the GenBank database using the BLAST search tool. The isolate was found to share 99% sequence identity with *Salinivibrio costicola* subsp. costicola strain ATCC 33508 (NR_027590.1). NS01 was further isolated by use of a plaque assay of filtered GSL water. The bacteriophage was further purified by isopycnic ultracentrifugation (64% CsCl w/v) at ~125,000 x g for 24 hours at 4 ^{o}C.

Genome isolation and sequencing, and sequence analysis: Genomic DNA was sequenced at the Brigham Young University DNA sequencing Center using a Roche Genome Sequencer FLX Instrument and employing the GS FLX Titanium Sequencing XLR70 kit (Roche Diagnostics

Corporation, Indianapolis, Indiana, USA) by Dr. Belnap's lab. Putative open reading frames (ORFs) of the NS01 genome were determined using GeneMarkS [62] and numbered according to whole-genome homology with bacteriophage PA11 [63].

Electron microscopy and image reconstruction, and structural modeling: For negative staining and cryo-EM, purified NS01 was prepared for microscopy as previously described (33) and was done in Dr. David Belnap's lab. Structural studies were accomplished by Professor David Belnap's lab. All surface renderings of the reconstruction were done at 1-σ contour level, which was defined by the sum of the average and standard deviation of map densities. Individual protein subunits were manipulated by eye to obtain a final fitting.

NS01 Desalting: 100 kDa filter (Sartorius Stedim, Bohemia, New York, USA) was applied for NS01 virion desalting. Virion in solution was loaded onto a 100 kDa filter and centrifuged at 1000 min^{-1}. Double distilled water was used for virion buffer exchange for 3 times and further transferred to a new 100 kDa filter for complete desalting. Virions were collected by centrifuging at highest speed through the filter with a new centrifuge tube.

Protein separation by SDS-PAGE and in-gel digestion: NS01 virions were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to separate individual proteins. The gel of 4-15% polyacrylamide, containing 10 wells (Bio-Rad Laboratories, Inc., USA) was stained using Coomassie blue reagent, after which individual gel bands were excised then shredded by spinning gel fragments through micropipette tips. Proteins were

prepared for analysis by mass spectrometry using a modified filter-aided sample preparation (FASP) protocol, adapted from the protocol of Dr. Mann's group[64].

Briefly, the gel fragments were destained using a 1:1 mixture of acetonitrile:8M urea, 0.1M Tris-HCl (pH 8.5). The denatured proteins were subsequently reduced with 0.1M dithiothreitol in UA buffer (8M urea, 0.1M Tris-Cl, pH 8.5), then carboamidomethylated with 50 mM iodoacetamide in UA buffer. Each solution was washed by filtration through a 10 kDa molecular weight cutoff filter (Sartorius Stedim, Bohemia, New York, USA). The urea and iodoacetamide were then replaced by 50 mM ammonium bicarbonate. Proteins were finally digested by sequencing grade trypsin (Promega, Madison, WI, USA) overnight at 37 ^{o}C and then acidified by addition of formic acid (Rockford, IL, USA) to a 1% v/v concentration in the final solution.

Mass spectrometry: Acidified peptide samples were loaded onto a 75 micron x 15 cm nanoAquity C18 column (Waters Corporation, Milford, Massachusetts, USA) and eluted by a 300min binary gradient of Optima grade Solvent A (5% acetonitrile, 0.1 % formic acid) and Optima grade Solvent B (0.1% formic acid, 99.9% acetonitrile) (Thermo Scientific, San Jose, California, USA) at a rate of 325 nL/min composed of the following steps: From a baseline at 98% Solvent A, a 208-minute linear gradient to 63% A , a 23-minute gradient to 1.5% A, followed by a 20-minute return to a 98% A baseline for 27 minutes.

Column effluent was directed to a nanoESI spray source on a LTQ-Orbitrap XL (Thermo Scientific, Waltham, Massachusetts, USA). Data dependent acquisition was performed by

coupling a 60,000 resolution survey scan in the Orbitrap with up to the top ten subsequent CID MS/MS scans, acquired in the ion trap (signal threshold of 500, normalized collision energy of 35%, isolation width of 2 m/z, with a two-count dynamic exclusion in a three-minute window). A lock mass of the cyclomethicone N5 ions generated in the electrospray process from ambient air (m/z 371.101230) was used as an internal mass calibration standard[65].

Data analysis: MS/MS spectra were converted to Mascot generic format by the MsConvert program (http://proteowizard.sourceforge.net/pubs.shtml). Mascot searching was queried against a database of ORFs derived from the NS01 genome. The peptide search space was expanded by concatenating the NS01 database with a comprehensive database of the *Escherichia coli* ATCC 8739 proteome (GenBank accession no. YP_001723016.1) and a reverse decoy version of the same database. Results were filtered to provide a 1% false discovery rate (FDR) confidence at the peptide level.

2.3 Results and discussion

Proteins in the head structure module. We identified gp88 as the major capsid protein because of its high abundance in the NS01 SDS-PAGE analysis and LC-MS of the gel band (Figure 4). PSI-BLAST of gp88 resulted in matches with P22-like coat proteins in Clostridium phage phiCP26F and Clostridium phage phiCP13O with E-values ≤ 1e-18. Pairwise sequence analysis of gp88 and the major capsid proteins of T7, HK97, and P22 resulted in only 19%, 15%, and 20% identity, respectively. However, secondary structure predictions by the Phyre

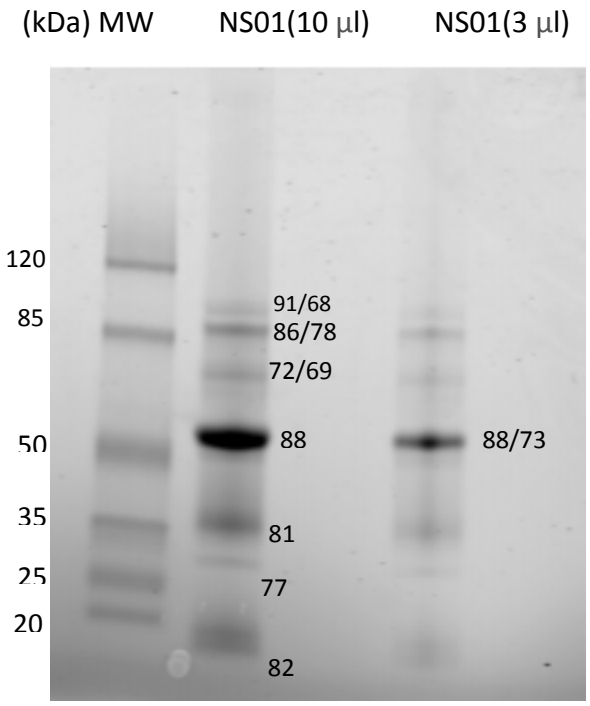


Figure 4. Protein composition of mature wild-type NS01 particles determined by SDS-PAGE and identified by LC-MS/MS. 4 pairs of gene products have similar sizes and were not separated in the gel including gp91/68, gp86/78, gp72/69, gp88/73. See Table 2 for properties of gene products. Lane MW contains molecular size markers. MW: molecular weight marker in kDa; NS01 (10 µL): 10µL of samples were loaded onto the gel; NS01 (4 µL): 4µL of samples were loaded onto the gel.

Protein Fold Recognition Server [66] predicted the structure of gp88 to most closely resemble the major capsid protein fold of prochlorococcus2 cyanophage p-ssp7, T7, P22, and HK97 with 100%, 100%, 99.5%, and 98.1% confidence, respectively. These results are consistent with the observation that the HK97 fold is more conserved than is its sequence among T7-like phages.

Proteins in the tail structure module. Putative tail structure proteins were predicted to be leftwards of the head structure module in the genome, characteristic of T7-like phages[60]. A PSI-BLAST search suggested that gp79 and gp86 are putative tail proteins.

In addition, several large ORFs are found in the NS01 genome at locations reasonably predicted to encode tail-associated proteins. For example, trimeric structures of coiled-coils and β-helices are a common motif in tail-associated proteins, such as the tail-spike protein of P22 (homotrimers of β-helices)[67] and tail fibers of bacteriophage T7 (triple-stranded coiled-coils)[68]. Thus, the presence of coiled-coils and β-helices in candidate ORFs within the putative tail structure module was examined. The BetaWrapPro program predicted right-handed beta-helix motifs in gp81, gp82, and gp85, and the COILS program predicted a coiled-coil motif in gp78.

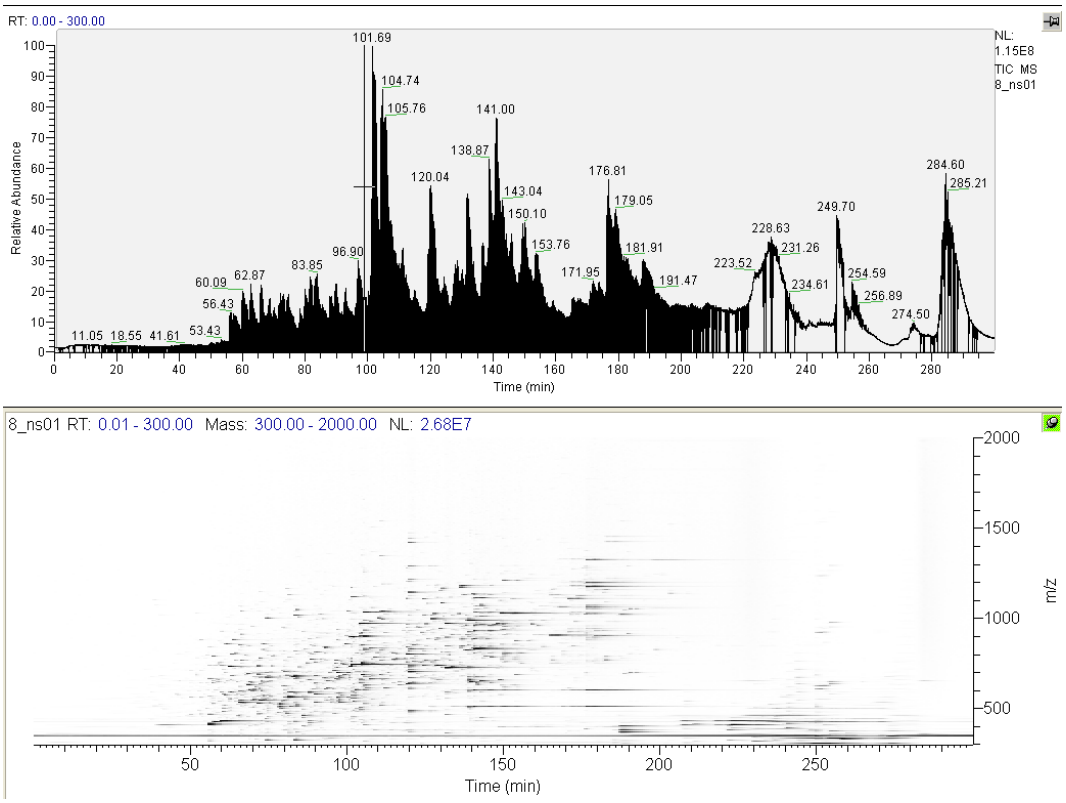


Figure 5. HPLC Orbitrap-LTQ chromatography (top) and ion map (bottom) of gp88.

Mascot Search Results

Protein View

Match to: **gene_88** Score: **1744**
|NS01|377_aa|-|56340|57473 **>empty-fasta-def-line**
Found in search of 8_ns01.mgf

Nominal mass (M_r): **40561**; Calculated pI value: **5.10**
NCBI BLAST search of gene_88 against nr
Unformatted sequence string for pasting into other applications

Variable modifications: Carbamidomethyl (C),Phospho (Y),Phospho (ST),Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **91%**

Matched peptides shown in **Bold Red**

```
     1 MSFTTDNSPF AGGKPNAQNF AGTSFTADPS LAAPQRAGNG KDAANSNWIA
    51 TIYSKEVLME FRKTSVVESI TNNDYYGEIS DYGDSVIIVK EPEIDIVDYA
   101 RGQKLESQAI DADSLKIVLD QAKAFQFQVD DIEKKLAHNN WQSLASNRAA
   151 YKLKDSFDLA ILNYIADNMT VKNIIASTTG ETAIAGATSG SAAFTTLAAA
   201 GANALKIHEA PSASEVSSGV ALTPVNLLNK FNLKLDLADV PEDGRWVVVD
   251 PEFIEVAMRE DSNVLNRDYN DGMRSIKNGL MNLSGIRGLQ MYKTNNSPKL
   301 VTSDRPQTDG KLNDAGRIIM AGHSSAVATV SAIVKTESFR SQQTFADVVR
   351 GLHVYGRAVI RPESLTGAAV TYQPIDA
```

Figure 6. Mascot result output of gp88, the major capsid protein, which had a score of 1744 and a sequence coverage of 91%.

ORF	Unique peptides	% Coverage	Putative function & structure prediction
gp91	18	37.48	possible N4-like portal protein
gp86	15	40.96	tail fiber
gp72	21	63.24	NA
gp88	28	91 (10 uL)	coat protein
gp77	5	48.79	NA
gp88	20	73.74 (3 uL)	coat protein
gp78	16	54.29	coiled-coils (*COILS*)
gp68	19	36.12	NA
gp69	14	37.62	NA
gp81	10	27.2	beta-helix (BetaWrapPro)
gp82	5	28.39	beta-helix (BetaWrapPro)
gp73	4	13.59	NA

Table 2. NS01 protein identification results from Mascot found by running against the NS01 ORF database. Trypsin was used for a bottom-up proteomics approach. Unique peptide identification and the percentage of amino acid sequences observed were included in the table as a reference for the identification confidence.

Chapter 3 Discovery of colostrum and milk biomarkers predictive of production-related metabolic disease risk through a lipidomics approach

3.1 Abstract

Production-related metabolic diseases (PRMDs) are a set of debilitating disorders occurring in certain early lactation dairy cows. Animals could potentially be rescued prior to clinical signs of disease if it were known that they were at risk. In order to identify valuable lipid biomarkers, predictive of PRMDs in asymptomatic postpartum dairy cows, samples of colostrum (20 cases, 22 controls) and post-partum day 4 milk (16 cases, 20 controls) were collected and analyzed for lipid composition and specific lipid abundance using mass spectrometry. A modified Bligh and Dyer lipid extraction method, followed by direct injection electrospray ionization-mass spectrometry (ESI/MS) approach, was used for this lipidomics study. Many lipids were differentially expressed in both colostrum and milk as determined in a comparison between cows that later developed PRMDs to cows having no clinical signs of health problems. After statistical modeling of these several candidate markers, three panels of biomarkers were developed. A combination of 2 milk lipid biomarkers including a diglyceride (DG 18:0/18:0) and a triglyceride (TG 18:0/18:0/18:1) provided predictions with 75.0% sensitivity at 90.0% specificity. A combination of 3 colostrum lipid biomarkers including TG 16:0/18:1/18:3, DG 16:0/16:0 and an unclassified lipid $C_{40}H_{60}NO$ provided predictions with 90.0% sensitivity at 86.4% specificity. Furthermore, a combination of 7 colostrum and milk biomarkers, including calculated differences between 'shared' markers found to be significantly different in both colostrum and milk, provided a predictive sensitivity of 87.5% at a specificity of 100%. Therefore, three panels of useful,

predictive, lipid biomarkers have been discovered in peripartum colostrum and milk sampling via lipidomics that can identify at risk cows days to weeks prior to onset of PRMDs.

Keywords: Production related metabolism; lipidomics; mass spectrometry; colostrum; milk; biomarker

3.2 Introduction

Production-related metabolic diseases (PRMDs) in dairy cows have been and continue to be a costly problem for the dairy industry. Different phenotypes of PRMDs have been documented, including hypocalcemia (commonly known as milk fever), displaced abomasum (DA), ketosis, fatty liver syndrome, and retained placenta amongst the most common disorders. These health problems may present with decreased milk production, altered milk composition, reduced reproductive capacity, shortened life expectancy and lower cull value, resulting in both animal and economic loss. Interventions for PRMDs are currently available after onset of clinical signs, but result in additional economic losses associated with medication and labor costs. Identifying animals that will have these complications early enough to prevent onset of PRMDs has not been possible previously.

PRMDs are multifactorial and the causes are complex. The different phenotypes of PRMDs have been documented based on the clinical signs of disease manifested. However, although a variety of health problems have been identified, they are inter-related physiologically, and not distinct, independent entities[69,70]. Acute milk fever results from hypocalcemia, characterized by reduced ionized blood calcium levels, and must be treated as a medical emergency[71]. Displaced abomasum, in contrast, has been shown to be preceded by ketosis, and is often accompanied by subclinical hypocalcemia[72]. Ketosis is a metabolic process that occurs when the body does not have enough glucose to maintain normal cellular energy production[73]. In consequence, stored fats are broken down and mobilized to meet energy needs, resulting in elevated fatty acids and a build-up of organic acids called ketone bodies within the circulation. Fatty liver syndrome in dairy cows has been shown to occur when the

synthesis of triacylglycols (TGs) is higher than their export and appears to be associated with hormone disregulation[74] and not surprisingly, disease states associated with ketosis may have varying degrees of hepatic lipidosis. Any of the aforementioned PRMDs may also occur with concurrent mastitis, metritis, retained placenta, or other health problems, begging the question as to whether or not affected animals may have a genetic predisposition for this problem. Since PRMDs are not limited to a single phenotype, and represent an array of disorders having inter-related causes, including dietary and environmental management practices, biomarkers that can identify the downstream responses to production demands in dairy cows associated with these metabolic disorders would be a valuable tool to detect the animals at risk, thereby allowing for appropriate interventions.

Lipid metabolism has been found to be abnormal in certain metabolic diseases in dairy cows[75]. For example, the liver triglyceride to glycogen ratio has been used to predict susceptibility of cows to ketosis.[75] Increase in plasma non-esterified fatty acids (NEFA) and β-hydroxybutyrate (BHBA) are both significantly associated with development of peripartum diseases.[42] Milk fatty acids, such as C18:1 cis-9, have been proposed as possible biomarkers able to diagnose elevated concentrations of plasma NEFA early in dairy cows.[76] However, there has not been a comprehensive study of lipid biomarkers with the intent of their anticipating PRMDs. Most current research has focused on dietary management and how diet and/or environmental interactions (e.g., over-conditioning at calving, low nutrient intake, environmental stress) can lead to peripartum diseases. Therefore, it is still highly desirable to discover novel biomarkers that are able to predict PRMDs with adequate

sensitivity and specificity.

In other previous studies, lipids have been shown to be related to metabolic disorders, including ketosis[77]and fatty liver [78] which are also associated with development of DA and milk fever. The majority of PRMDs occurs after calving, during early lactation, and are believed to be due to difficulty adapting to the high demands of lactation, resulting in physiological imbalances in susceptible cows.[79] Since colostrum and milk are both rich in lipids, and readily accessible for collection, we hypothesized that lipid expression differences in colostrum and milk samples between PRMD susceptible and resistant dairy cows would exist and could be potential predictive biomarkers.

'Shotgun', i.e., in-depth or comprehensive, lipidomics can survey thousands of unique lipids in a single biological specimen. One such lipidomic approach using direct injection, electrospray ionization coupled with highly mass accurate, tandem-mass spectrometers (ESI/MS/MS) represents a powerful tool for cataloguing, quantifying and chemically characterizing lipids[20] in tissue, cell or body fluids. Lipidomics can complement peptidomic and proteomic methods. Indeed, colostrum and milk samples can be readily fractionated into lipid- and protein-rich layers. Lipidomics, especially direct injection lipidomics, is substantially less involved than top-down, global proteomic methods, which typically require enzyme digestion and multiple separation steps prior to MS. Consequently, lipidomics is well suited for the discovery of lipid biomarkers. Early, predictive biomarkers could help detect a cow's propensity for PRMDs in advance, to allow for prevention or early stage treatment

interventions. Susceptible cows could be treated with extra nutrients, minerals, or other supplements, along with a health-promoting environment, to reduce animal losses[80] while maintaining a physiologically compatible, albeit lower level of milk production. PRMD susceptible cows could be later reintroduced to the herd after their health was stable, and lactation parameters met economic production criteria.

3.3 Materials and lipid standards

Reproducibility and quantitative comparisons using direct injection lipidomics are dramatically improved with the introduction of known, consistent amounts of one or more synthetic lipid standards to the specimen to be analyzed. The synthetic lipid standard used in these studies was archeol (2, 3-diphytanyl-sn-glycerol), purchased from Avanti Polar Lipids, Inc. (Alabaser, AL, USA). Chloroform (HPLC grade) was purchased from Avantor Performance Materials, Inc. (Center Valley, PA, USA). Methanol (Optima LC/MS) was purchased from Fisher Scientific (Fair Lawn, NJ, USA). Isopropyl alcohol (suitable for LC and UV-spectrophotometry) and ammonium acetate were purchased from Mallinckrodt Chemicals, Inc. (Phillipsburg, NJ, USA)

3.4 Sample collection and analysis

Sample collection:

Randomly chosen, age-, lactation-, and parity-matched clinically normal primiparous (n = 101) and multiparous (n=109) Holstein cows at Brigham Creek Dairy (Elberta, UT) were studied. Cows received the same dry cow total mixed ration prepartum and were fed the

same fresh cow, early lactation diet the first week postpartum.

Composite aliquots of colostrum secreted on the day of parturition and milk produced on the fourth day of lactation were collected from the four quarters of each cow's udder as it pooled in a milking claw and was sampled via the exit milk tube. Samples were collected in 50 mL conical vials, then buried immediately and completely in crushed ice for transport from Brigham Creek Dairy, Elberta, UT, to the lab at Brigham Young University, Provo, UT. Each sample was divided into 10 mL aliquots before being flash frozen in liquid N_2 in a Dewar flask and stored in a -80°C freezer until later analysis.

Animal health records were monitored through 60 days in milk (DIM) for evidence of PRMDs or other complication. The specific phenotypes considered were ketosis, left displacement abomasum (LDA), milk fever (MF), fatty liver, and/or hind limb weakness attributed to obturator nerve paresis (OP) (Table 3).

Lipid extraction:

Colostrum and milk samples were processed for lipidomic analysis in the same way. Selected samples were taken from the -80°C freezer and thawed completely at room temperature. Lipid and the aqueous protein fractions of colostrum (or milk) were separated by centrifugation (Sorvall Legend XTR Centrifuge, Thermo Scientific) for 20 min at 650 x g at 4°C. After separation, 10 mg of the upper, lipid containing layer of the sample was taken and dissolved in 3.8 mL of a 2:1:1.25 (v/v/v) solution of chloroform:methanol:isopropanol[81].

After shaking the mixture for 30 s until complete lipid dissolution, 1.2 mL of double distilled deionized water was added and the mixture was shaken vigorously and then allowed to sit for organic and aqueous layer separation. After incubating at 37^{o}C overnight for complete lipid extraction, the bottom organic phase was taken and diluted 500 times with a 2:1:1.25 (v/v/v) solution of chloroform:methanol:isopropanol containing 15 mM of ammonium acetate and containing the lipid standard archaeol (6 nM) just prior to instrumental analysis.

Lipid profiling by time-of-flight mass spectrometry:

Mass spectrometric analysis was performed on an electrospray ionization source time-of-flight mass spectrometer (6230 ESI-TOFMS, Agilent Technologies). The ionization voltage was set to 3.5 kV, gas pressure to 15 psi and the source was controlled by Agilent MassHunter Workstation Data Acquisition software (Agilent Technologies, Santa Clara, CA, USA). All lipid samples were infused at the flow rate of 10 µL/min by a syringe pump (New Era Pump Systems, Inc, Farmingdale, NY, USA) and analyzed in positive ion mode for 3 min with MS data collected over the range from m/z 100 to 1500. Technical replicates (n = 2) were performed for each sample and the values averaged in order to minimize instrumental variability.

Data analysis:

Instrument software (Agilent MassHunter Qualitative Analysis B.07.00 software, Agilent Technologies, Santa Clara, CA, USA) was used to generate a peak list with the abundance of each lipid (ion counts) recorded for each specimen. The abundances of the 2 instrumental

replicates were exported as CSV files using Agilent MassHunter. Two columns are generated in each file, including a peak list with m/z values and the corresponding abundances. The peaks were aligned between each run by a home-made macro program within an Excel file through Visual Basic Programming Language according to the m/z values and the associated abundances recorded. Accuracy of peak alignment by mass between samples was further checked manually. The intensity of the lipid standard in each run was determined and used for data normalization. Then, the normalized abundances of the 2 instrumental replicates were averaged.

Statistical analysis:

In order to find useful biomarkers that distinguish PRMD-resistant and susceptible animals, an initial, two-tailed Student's t-test was carried out on the normalized abundances for the two comparison groups. Those lipid molecules having a p-value less than 0.05 were compiled and submitted for further discriminant statistical analysis. Colostrum and milk samples were analyzed separately. Furthermore, the 'shared' significant m/z values between colostrum and milk, i.e., those species that were significantly different in both milk and colostrum, were selected, and the differences between the normalized abundances for each peak were calculated by subtraction, and the differences further considered in the statistical analyses.

Lipid compounds found in significantly different, quantitative abundances based on the Student t-test, were included in a second data set that was submitted to linear

discriminative analysis to model combinations or panels of milk and/or colostrum biomarkers (SAS 9.3, SAS Institute Inc., Cary, NC, USA). Linear discriminant analysis has been described previously[42]. In brief, each statistically different lipid was evaluated in a step-wise discriminant statistical analysis with it being the independent variable of lipid abundance in samples and HSC ranking (heath score): 0 (healthy) or 1 (PRMD treated/culled/died within 2 months postpartum) being the dependent variables. A significance level of $p < 0.05$ was considered significant for all tests. Discriminant analysis, grouping variable HSC, was performed for each colostrum (CS) and milk (MK) measure. Calculated differences between 'shared' markers were obtained through significant m/z values of normalized abundance of colostrum minus normalized abundance of milk. A two-sample t-test was carried out to determine significant differences in the measurable abundance of lipid species found in colostrum and milk. Further statistical analysis was performed using a general linear model procedure (PROC GLM) with SAS (version 9.3) to determine differences between PRMD positive and negative cows' lipid species found in colostrum, milk, or shared in both substrates.

Additionally, after reviewing the results in milk, it was observed that all lipid biomarkers of a particular m/z range were increased in milk of animals later developing PRMDs whereas all the lipid biomarkers in a higher m/z range were reduced in disease susceptible animals. To test the significance of this, a 2x2 contingency table was created that looked at the classification of diseased animals with the consistency of an elevated or decreased abundance observed for lipid biomarkers based on their m/z category. A Fisher Exact test

was performed and a p-value <0.05 was considered significant.

Lipid identification:

Three panels of high performing predictive biomarkers were generated by statistical analysis. Each marker in the three panels was then chemically characterized by ESI-tandem mass spectrometry (MS/MS). Targeted MS/MS was applied to extracts in an effort to identify or substantially characterize each useful biomarker. A combination of 'exact' mass studies, together with MS/MS fragmentation studies using collisionally-induced dissociation, were used to chemically characterize candidate biomarkers. Fragmentation data was acquired on both a QSTAR Pulsar I quadrupole orthogonal time-of–flight mass spectrometer through an IonSpray Source (Applied Biosystems, Foster City, CA, USA) and on an Agilent 6530 accurate-mass quadrupole/time-of flight mass spectrometry (Agilent Technologies, Santa Clara, CA, USA). Specific colostrum and milk samples were selected for characterization based on a higher abundance of the targeted lipid peak of interest being present and the sample was extracted using 2:1:1.25 chloroform:methanol:isopropanol with 15 mM ammonium acetate present.

One MS instrument, the QSTAR, had the following settings: voltage potential of 4800 V for the ion spray needle coupled with a flow rate of 4 µL/min using the syringe pump. The drying gas used was N_2 and the collision gas was nitrogen. Optimal collision energies varied depending on the size, structure, and abundance of the lipids, but were typically between 15 and 40 eV. Analyst QS software was used for controlling the instrument, and a 50 - 2000

mass/charge range was used for each scan.

For the second instrument, the Agilent Q-TOF, direct infusion with the syringe pump was used for sample injection at a flow rate of 10 µL/min. The ion selection quadrupole Q1 was operated under the unit resolution settings, and fragments were detected within the m/z range of 100 - 1000 using positive ion mode. Each MS/MS spectrum was acquired for 30 s with a collision energy range of 15 - 40 eV. An isolation window of 1.3 was used.

Chromatograms were extracted from ESI total ion chromatograph (TIC) at a MS/MS level with Agilent MassHunter Qualitative Analysis B.07.00 (Agilent Technologies, Santa Clara, CA, USA). The exported product ion mode was broadly used for lipids identification. Predicted identities of target lipids were searched for using the on-line reference site LIPID MAPS[82] and the Elemental Composition Calculator programed by Frank Antolasic was used in conjunction with the experimentally determined accurate mass of lipids after determination of their detected adduct. Fragmentation information was further manually determined in product ion mode through review of neutral loss species, or scanned fragment information. The LIPID MAPS MS fragment prediction tool (http://www.lipidmaps.org/tools/index.html) was also applied to determine the product ion peak lists, which often represented sn1 and sn2 acyl losses mainly for glycerolipids.

3.5 Results

Descriptive Data:

Day 1 colostrum and day 4 milk secretions were collected from 101 primiparous and 109 multiparous asymptomatic postpartum Holstein cows. Approximately 10% of the cows from this larger cohort developed PRMDs within 24 weeks after calving. Therefore, their samples were selected along with healthy matched controls, and submitted to lipidomic analysis. The 'disease' cows were diagnosed with PRMDs, that included abnormalities such as ketosis, left displaced abomasum (LDA), hypocalcemia (commonly known as milk fever (MF)) and obturator nerve paresis (OP) as shown in Table 3. All cows enrolled in the study were monitored for the development of PRMDs and their health history records were tracked for 8 weeks postpartum in order to select healthy control and PRMD samples for this study. Controls were matched by age, parity, and lactation to animals that developed PRMDs and sorted according to similar date of calving. The selected samples are described in the demographics table (Table 3) that includes cow ID number, the diagnosis group (disease or control), and phenotype. Cow 19010, 21389 and 23567 did not have milk collected on day 4 after calving due to illness or no longer being in the herd. Cow 23608 and 23971 did not have milk samples due to specimen loss or it being compromised. Cow 23374 was diagnosed with suspected milk fever, but was not included in the final data analysis due to uncertainty of the health record diagnosis.

Significant lipid biomolecules between PRMD resistant and susceptible cows through lipids profiling approach:

A peak list of m/z values, i.e., a list of lipid species observed in the MS, with their corresponding abundances (ion counts > 600) were exported from the Qq-TOF mass

spectrometer using Agilent MassHunter Qualitative Analysis B.07.00 software (Agilent Technologies, Santa Clara, CA, USA). A total of 1800-2000 ion peaks (ion counts > 600), representing potentially that number of different lipids, were detected in MS runs of separately analyzed colostrum and milk samples. After peak alignment among all the colostrum or milk samples based on the m/z, normalization was carried out by using the ratio of the peak abundance to the abundance of the internal standard. A Student's t-test was applied within Excel for both colostrum and milk lipids. Some potential markers with statistically significant differences (p-value <0.05) between diseases and controls are listed in Table 3 and Table 4. In total, 61 statistically significant, quantitatively different lipids were discovered in colostrum (Table 3), and 77 significant, quantitatively different lipids were discovered in milk (Table 4). Of these, 31 lipids were found to be significantly different between cases and controls in both colostrum and milk. Additionally, for these 'shared' markers, differences between colostrum and milk values of the same peak were calculated by using normalized abundances of that lipid in colostrum minus its normalized abundance in milk. P-values of shared differences were calculated with SAS 9.3 as provided in Table 5. Further statistical modeling suggested how predictive these markers are for PRMD resistant and susceptible dairy cows as shown in Figure 11.

D or C	Cow ID	Diagnosis	Lipids
C	19048	-	csmk
C	20505	-	csmk
C	20804	-	csmk
C	20873	-	csmk
C	21132	-	csmk
C	21155	-	cs
C	21247	-	csmk
C	21859	-	csmk
C	22219	-	csmk
C	22598	-	csmk
C	22877	-	csmk
C	23609	-	csmk
C	23772	-	csmk
C	26556	-	csmk
C	26558		csmk
C	26678	-	csmk
C	26776	-	csmk
C	26852	-	csmk
C	26998	-	csmk
C	29241	-	csmk
C	29552	-	csmk
C	29554	-	cs
C	29610	-	cs
D	14112	Died	csmk
D	16320	MF	csmk
D	17829	OP, Died	csmk
D	17841	MF, Died	csmk
D	19010	MF	cs
D	20594	LDA	csmk
D	20712	LDA	csmk
D	21389	LDA	csmk
D	22377	LDA	csmk
D	23567	MF, LDA	cs
D	23608	LDA	cs
D	23762	LDA	csmk
D	23971	LDA	cs
D	25249	LDA	csmk
D	25853	LDA	csmk
D	26035	MF, OP	csmk
D	26832	LDA	csmk
D	29337	LDA	csmk
D	29551	LDA	csmk

D	29685	LDA	csmk
SuspD	23374	SuspectMF	Csmk

Table 3. Demographics. Postpartum first day colostrum and milk secreted on day 4 were collected from 210 Holstein cows; cows that developed PRMDs were selected as diseased animals and compared with age, parity, and lactation matched controls. D: disease; C: control; MF: hypocalcemia, commonly known as milk fever; LDA: left displaced abomasum; OP: obturator nerve paresis; CS: colostrum; MK: milk

Candidate No.	m/z value	normalized value of diseases	normalized value of controls	Abundance comparison	p value
1	344.2247	0.114880157	0.103689575	Higher in dieases	0.042061688
2	388.2525	0.121962693	0.112940105	Higher in dieases	0.048555005
3	489.1043	0.373807522	0.290967829	Higher in dieases	0.049572488
4	570.4639	0.176409178	0.11189889	Higher in dieases	0.013504484
5	586.5356	0.202742707	0.265637153	Higher in controls	0.001414111
6	598.5014	0.112665915	0.080800463	Higher in diseases	0.028964416
7	612.5527	0.065970917	0.078148394	Higher in controls	0.043833247
8	615.5605	0.069239485	0.087939768	Higher in controls	0.042757696
9	648.463	0.085478463	0.115099741	Higher in controls	0.049597223
10	652.5473	0.09833478	0.07354011	Higher in diseases	0.04724037
11	654.5636	0.939107997	0.652840941	Higher in diseases	0.014671469
12	659.5223	0.198002601	0.142805881	Higher in diseases	0.004243638
13	668.5827	0.16539349	0.11621013	Higher in diseases	0.015227126
14	675.4999	0.126403559	0.099737468	Higher in diseases	0.034354145
15	680.5771	0.513222007	0.349466169	Higher in diseases	0.004172134
16	682.5948	2.590566861	1.404894949	Higher in diseases	0.000949216
17	687.5507	0.380657952	0.224279411	Higher in diseases	0.000405532
18	694.5999	0.107000465	0.067609268	Higher in diseases	0.003174857
19	695.5915	0.19157094	0.138180817	Higher in diseases	0.007715938
20	696.6045	0.225131134	0.148052417	Higher in diseases	0.002541267
21	703.5253	0.262044497	0.163091281	Higher in diseases	0.001391699
22	704.5477	0.134370476	0.081080697	Higher in diseases	0.000546421
23	706.587	0.217696672	0.140102698	Higher in diseases	0.002698514
24	708.6098	0.732302369	0.366041823	Higher in diseases	0.000466794
25	710.6251	1.197685611	0.752962244	Higher in diseases	0.002817283
26	723.6208	0.399721453	0.242845266	Higher in diseases	0.001139869
27	731.5633	0.12838516	0.084643368	Higher in diseases	0.001933664
28	732.5673	0.113258213	0.082269331	Higher in diseases	0.005490254
29	734.6126	0.098106733	0.07821988	Higher in diseases	0.039988195
30	736.6403	0.249638522	0.1680125	Higher in diseases	0.003092746
31	749.6357	0.126106942	0.068235491	Higher in diseases	0.000335962
32	751.6513	0.196540202	0.133759701	Higher in diseases	0.002955134
33	752.6595	0.139024989	0.105137977	Higher in diseases	0.011054544
34	848.7654	0.934633969	0.749696345	Higher in diseases	0.01734462
35	850.7813	2.586551283	2.001595098	Higher in diseases	0.007887614
36	855.7364	0.491215538	0.395923832	Higher in diseases	0.006170212
37	858.7473	0.101043456	0.074631085	Higher in diseases	0.000650882
38	861.574	0.070961192	0.099041479	Higher in controls	0.035043823
39	862.788	0.166454098	0.11855104	Higher in diseases	0.00492844
40	864.788	0.32486829	0.263356503	Higher in diseases	0.007491634
41	872.745	0.222708437	0.18695274	Higher in diseases	0.023286159
42	873.7445	0.154204103	0.12767435	Higher in diseases	0.016994617
43	874.7789	0.477500575	0.329773294	Higher in diseases	0.000405247
44	876.7964	1.366009113	0.692794525	Higher in diseases	0.00025238
45	878.8096	0.961036568	0.493672525	Higher in diseases	0.000466523
46	881.7594	0.304220425	0.175428923	Higher in diseases	6.91957E-05
47	883.759	0.201953914	0.112400998	Higher in diseases	0.000112302
48	889.7908	0.142787896	0.118691043	Higher in diseases	0.015486554
49	890.8055	0.151286664	0.110704439	Higher in diseases	0.0015826
50	891.8077	0.338952477	0.272686318	Higher in diseases	0.007426845
51	897.7294	0.157819943	0.098184814	Higher in diseases	0.00036568
52	899.7562	0.139790521	0.08871563	Higher in diseases	0.000225419
53	900.7894	0.220946001	0.139311432	Higher in diseases	0.000267302
54	902.8102	0.365998427	0.175549359	Higher in diseases	0.000451383
55	904.8244	0.333156789	0.156091233	Higher in diseases	0.00073496
56	906.8353	0.180777646	0.101178543	Higher in diseases	0.000967386
57	917.8214	0.169726223	0.098529751	Higher in diseases	0.000225545
58	919.8341	0.123850821	0.066611016	Higher in diseases	0.000146124
59	964.5856	0.073206947	0.105163006	Higher in controls	0.015808127
60	965.6969	0.107055016	0.06447254	Higher in diseases	0.000659915
61	967.7134	0.122811632	0.10158746	Higher in diseases	0.036742958

Table 4. Significant PRMD predictive lipid biomarkers ($p<0.05$) in colostrum. P values were obtained by Student's t-test between PRMD susceptible and resistant dairy cows. The means of normalized abundance in 20 diseased animals and 22 controls are shown in the 3rd and 4th columns.

Candidate No.	disease	disease mean	control mean	Abundance comparison	p value
1	572.4828	0.227073934	0.489156442	Higher in controls	5.66205E-05
2	600.5146	0.426188238	0.918376635	Higher in controls	1.72877E-05
3	626.5314	0.332706621	0.530873278	Higher in controls	0.000134814
4	628.5465	1.404315699	2.637108898	Higher in controls	4.38457E-06
5	633.5	0.150864782	0.265523759	Higher in controls	9.266E-06
6	640.5417	0.105486833	0.135917624	Higher in controls	0.005901871
7	642.5606	0.270175612	0.451766164	Higher in controls	2.83684E-07
8	652.5463	0.163679734	0.21762311	Higher in controls	0.001433185
9	654.5626	1.572666631	2.098405246	Higher in controls	0.000953975
10	656.5783	4.802102104	6.779852288	Higher in controls	2.41044E-05
11	657.5816	1.821030483	2.549534246	Higher in controls	2.62771E-05
12	658.5805	0.415087477	0.561397172	Higher in controls	4.48915E-05
13	659.5156	0.270610645	0.332875614	Higher in controls	0.005245116
14	661.5323	0.450185483	0.620125673	Higher in controls	3.67155E-05
15	668.5788	0.314661478	0.390475848	Higher in controls	0.003101227
16	669.5772	0.310755852	0.47647484	Higher in controls	1.96141E-06
17	670.5907	0.530011614	0.733158168	Higher in controls	1.14043E-05
18	675.4994	0.142120983	0.197464108	Higher in controls	0.00216583
19	677.507	0.295841154	0.439742727	Higher in controls	0.002393067
20	684.6074	4.038547322	5.375889332	Higher in controls	0.000453034
21	689.5597	0.404386212	0.525474832	Higher in controls	0.000420062
22	695.5945	0.288641146	0.348551776	Higher in controls	0.010188417
23	697.6055	0.762463331	0.974135232	Higher in controls	0.000186872
24	705.5422	0.280096199	0.389001747	Higher in controls	0.008195311
25	712.6371	1.282073361	1.879963512	Higher in controls	5.97024E-05
26	717.583	0.154938615	0.219021411	Higher in controls	8.00007E-05
27	721.5006	0.193633957	0.23423158	Higher in controls	0.048132457
28	725.6337	0.552109984	0.70004443	Higher in controls	0.001084341
29	732.5781	0.132098686	0.16261439	Higher in controls	0.027660775
30	733.5725	0.117869841	0.172986401	Higher in controls	0.002006596
31	734.6183	0.116226648	0.156062153	Higher in controls	0.000653951
32	738.6543	0.744455276	1.048604892	Higher in controls	0.000161202
33	740.6681	0.493078794	0.834138952	Higher in controls	0.000370689
34	752.6554	0.229681858	0.274217616	Higher in controls	0.007499157
35	753.6635	0.1902872	0.265708582	Higher in controls	5.5045E-05
36	754.6695	0.123341686	0.185374768	Higher in controls	1.45157E-05
37	764.6702	0.234524451	0.302551159	Higher in controls	0.000746323
38	766.6855	0.478029349	0.723050046	Higher in controls	0.000156398
39	768.6997	0.396400708	0.590469763	Higher in controls	0.020185394
40	780.6912	0.094034212	0.132513766	Higher in controls	2.56998E-05
41	792.7012	0.221344994	0.282647947	Higher in controls	0.001632091
42	794.7174	0.537540377	0.700470268	Higher in controls	0.01869102
43	808.7291	0.111769356	0.145186647	Higher in controls	0.000464602
44	809.7327	0.095715993	0.126344779	Higher in controls	0.01393217
45	810.7431	0.111801112	0.139741421	Higher in controls	0.028215881
46	848.7645	1.205927939	0.97061175	Higher in diseases	0.001509529
47	850.7802	3.14024801	2.310057231	Higher in diseases	2.60282E-05
48	855.7365	0.458887525	0.352796522	Higher in diseases	6.91148E-06
49	862.7773	0.261855371	0.206431713	Higher in diseases	0.003570806
50	863.7788	0.262347251	0.229356579	Higher in diseases	0.005640039
51	864.7913	0.444669242	0.343999739	Higher in diseases	0.000307567
52	865.7945	0.287828953	0.231822591	Higher in diseases	0.000241857
53	872.7541	0.240567243	0.203867439	Higher in diseases	0.02493066
54	874.7794	0.884013552	0.601752056	Higher in diseases	0.000311042
55	876.7962	3.975003341	2.283182118	Higher in diseases	4.45627E-05
56	878.8092	2.900494826	1.794288368	Higher in diseases	2.01532E-05
57	879.8124	1.398439996	0.894483247	Higher in diseases	2.12595E-05
58	881.7614	0.612866418	0.381728559	Higher in diseases	5.98877E-06
59	883.7632	0.430026039	0.285768862	Higher in diseases	7.37893E-06
60	888.789	0.123029972	0.086810486	Higher in diseases	0.000846913
61	889.7915	0.192554315	0.14362613	Higher in diseases	0.000244128
62	890.8051	0.317900205	0.212723122	Higher in diseases	0.000149677
63	891.8074	0.453012105	0.312917793	Higher in diseases	1.28318E-05
64	896.7527	0.090403386	0.075289411	Higher in diseases	0.022673036
65	897.7299	0.303077206	0.222820825	Higher in diseases	0.012642883
66	898.76	0.232897889	0.176361653	Higher in diseases	0.006328352
67	900.7909	0.393967161	0.269274785	Higher in diseases	0.000288913
68	902.8106	1.199197503	0.637454941	Higher in diseases	2.18259E-05
69	904.8246	1.308415831	0.719666079	Higher in diseases	1.31168E-05
70	906.836	0.588400876	0.365295263	Higher in diseases	1.85972E-05
71	908.8049	0.157094249	0.110560115	Higher in diseases	8.6087E-05
72	909.7825	0.221899985	0.133238648	Higher in diseases	5.60766E-06
73	915.8031	0.107663063	0.075385692	Higher in diseases	0.000332921
74	917.8208	0.374602543	0.223599465	Higher in diseases	5.61349E-05
75	919.8333	0.29246751	0.187698404	Higher in diseases	4.94772E-05
76	939.6842	0.188264489	0.240903499	Higher in controls	0.000158301
77	967.7069	0.16795484	0.19725702	Higher in controls	0.005952086

Table 5. Significant PRMD predictive lipid biomarkers ($p < 0.05$) in milk. P-values were obtained by Student's t-test between PRMD susceptible and resistant dairy cows. The means of the normalized abundance in 16 diseased animals and 20 controls are shown in the 3rd and 4th columns.

Obs	biomolecules	Mean_C	Mean_D	Difference	StdErr	DF	tValue	Probt
1	DF652_5473	-0.1448	-0.06356	-0.08124	0.01650	34	-4.92	<.0001
2	DF654_5636	-1.4531	-0.6123	-0.8407	0.1475	34	-5.70	<.0001
3	DF659_5223	-0.1912	-0.06697	-0.1243	0.02416	34	-5.14	<.0001
4	DF668_5827	-0.2751	-0.1447	-0.1304	0.02803	34	-4.65	<.0001
5	DF675_4999	-0.09726	-0.01074	-0.08652	0.02147	34	-4.03	0.0003
6	DF695_5915	-0.2121	-0.09033	-0.1217	0.02325	34	-5.23	<.0001
7	DF732_5673	-0.07994	-0.01389	-0.06605	0.01811	34	-3.65	0.0009
8	DF734_6126	-0.07823	-0.01643	-0.06180	0.01415	34	-4.37	0.0001
9	DF752_6595	-0.1702	-0.08774	-0.08243	0.01767	34	-4.67	<.0001
10	DF848_7654	-0.2434	-0.2847	0.04131	0.09265	34	0.4	0.6585
11	DF850_7813	-0.3721	-0.5994	0.2273	0.2663	34	0.85	0.3993
12	DF855_7364	0.03917	0.02702	0.01215	0.04909	34	0.25	0.8059
13	DF862_788	-0.09118	-0.09356	0.002384	0.02681	34	0.09	0.9297
14	DF864_788	-0.08544	-0.1160	0.03054	0.03793	34	0.81	0.4263
15	DF872_745	-0.01984	-0.01910	-0.00074	0.02608	34	-0.03	0.9777
16	DF874_7789	-0.2784	-0.4047	0.1263	0.07273	34	1.74	0.0915
17	DF876_7964	-1.6041	-2.5649	0.9608	0.3163	34	3.04	0.0046
18	DF878_8096	-1.3085	-1.8992	0.5907	0.2005	34	2.95	0.0058
19	DF881_7594	-0.2066	-0.2989	0.09229	0.04264	34	2.16	0.0375
20	DF883_759	-0.1734	-0.2197	0.04625	0.03018	34	1.53	0.1346
21	DF889_7908	-0.02712	-0.04889	0.02177	0.01526	34	1.43	0.1629
22	DF890_8055	-0.1029	-0.1628	0.05986	0.02474	34	2.42	0.0210
23	DF891_8077	-0.04577	-0.1122	0.06643	0.03648	34	1.82	0.0774
24	DF897_7294	-0.1247	-0.1399	0.01526	0.03527	34	0.43	0.6680
25	DF900_7894	-0.1311	-0.1644	0.03338	0.03462	34	0.96	0.3417
26	DF902_8102	-0.4627	-0.8160	0.3533	0.09900	34	3.57	0.0011
27	DF904_8244	-0.5646	-0.9608	0.3962	0.09890	34	4.01	0.0003
28	DF906_8353	-0.2650	-0.4005	0.1355	0.03998	34	3.39	0.0018
29	DF917_8214	-0.1262	-0.1980	0.07175	0.02924	34	2.45	0.0194
30	DF919_8341	-0.1202	-0.1625	0.04225	0.02184	34	1.93	0.0614
31	DF967_7134	-0.09657	-0.04388	-0.05269	0.01299	34	-4.06	0.0003

Table 6. Statistical analysis of 'shared' peak differences, i.e. for lipids significantly different between diseased and control animals in both colostrum and milk. The differences of shared biomarkers were calculated by using the normalized abundance of colostrum minus the normalized abundance of milk. Mean_C: mean differences of controls; Mean_D: mean differences of diseases; 31 shared peaks were found as shown, p values named as Probt in the rightmost column were obtained through SAS 9.3.

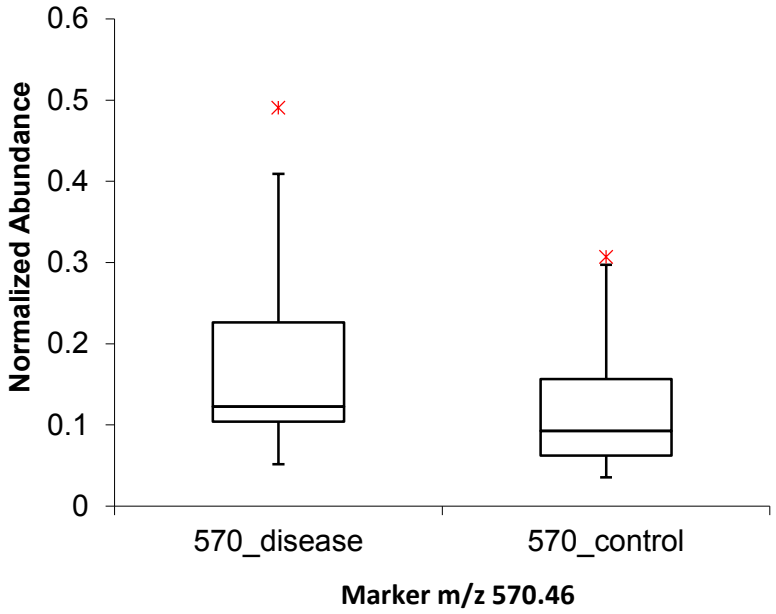
Figure 7. Representative mass spectra of PRMDs lipid marker at m/z 682.59 from 6 colostrum samples. The peak at 670.70 m/z is the ammoniated internal standard archaeol. The upper three samples are diseased cows (14112, 17841, 20712), and the bottom three are control cows (20873, 22219, 21859). Reproducible internal standard was shown in randomized MS runs. As observed in these 6 colostrum samples' mass spectra, m/z 682.59 is more highly expressed in cows which later developed PRMDs compared to control animals that remained healthy.

Statistical modeling of predictive PRMD lipid biomarker panels:

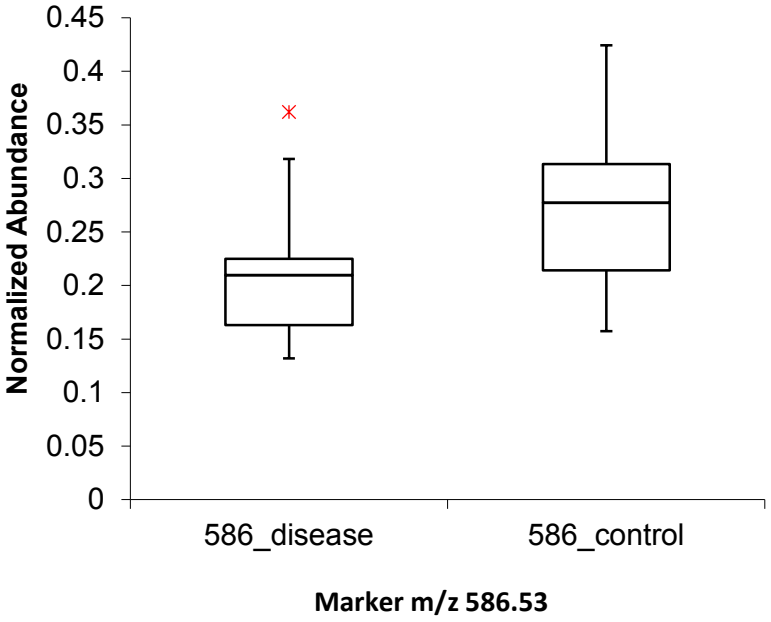
The best colostrum biomarker panel as found for these data using linear discriminative analysis contained three lipids as shown in Figure 8. The panel yielded 90.0% sensitivity and 86.4% specificity. The best milk biomarker panel contained two lipids as shown in Figure 9, which provided a sensitivity of 75.0 % at a specificity of 90.0%. In addition, an optimized combined colostrum and milk biomarker panel included 7 lipid markers: 2 lipids were from the colostrum data set, 2 lipids were from the milk data set, and 3 markers represented calculated differences between colostrum and milk for the 'shared' marker set. The combined biomarker panel yielded 87.5% sensitivity and 100.0% specificity (Figure 10).

	PRMD susceptible		PRMD resistant		
Colostrum Biomarker	mean	SE	mean	SE	P value
570.4639	0.1764	0.0248	0.1118	0.0144	0.0135
586.5356	0.2027	0.0121	0.2656	0.0152	0.0014
855.7473	0.1010	0.0058	0.0746	0.0049	0.0006

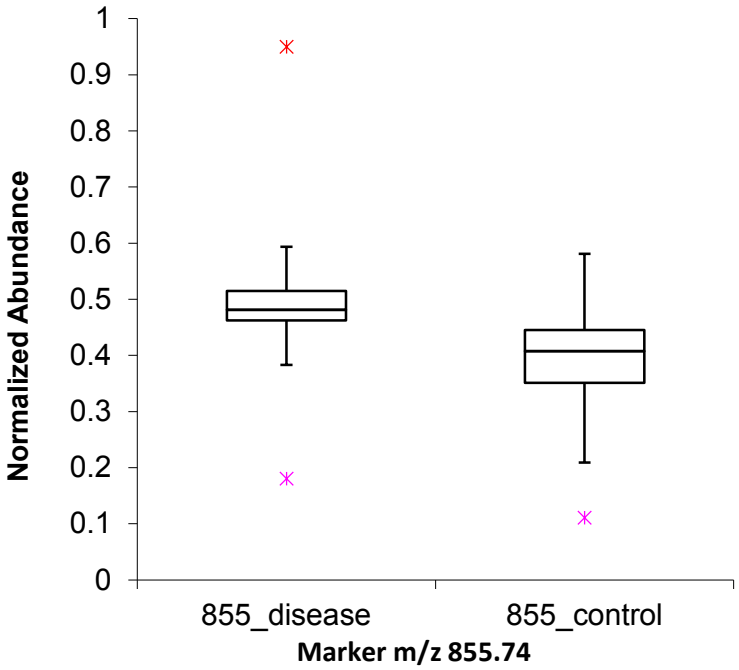
(a)



(b)



(c)

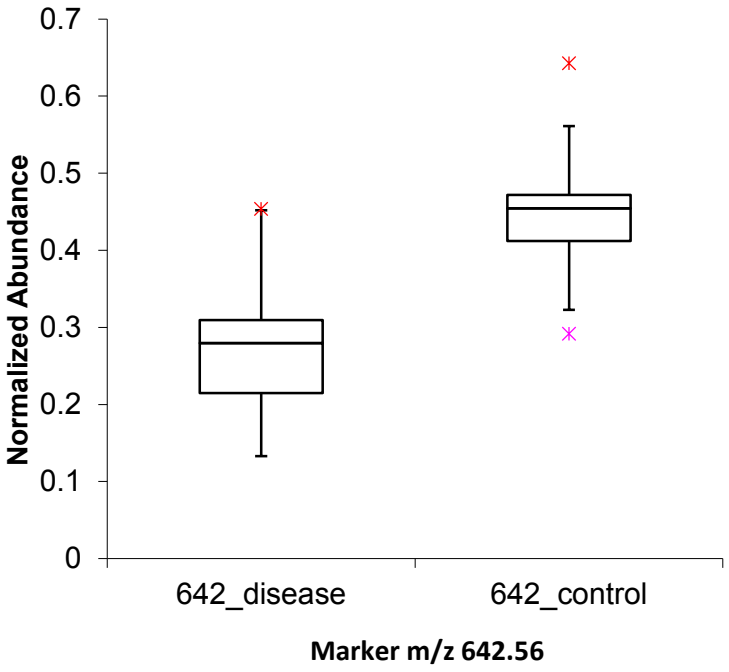


(d)

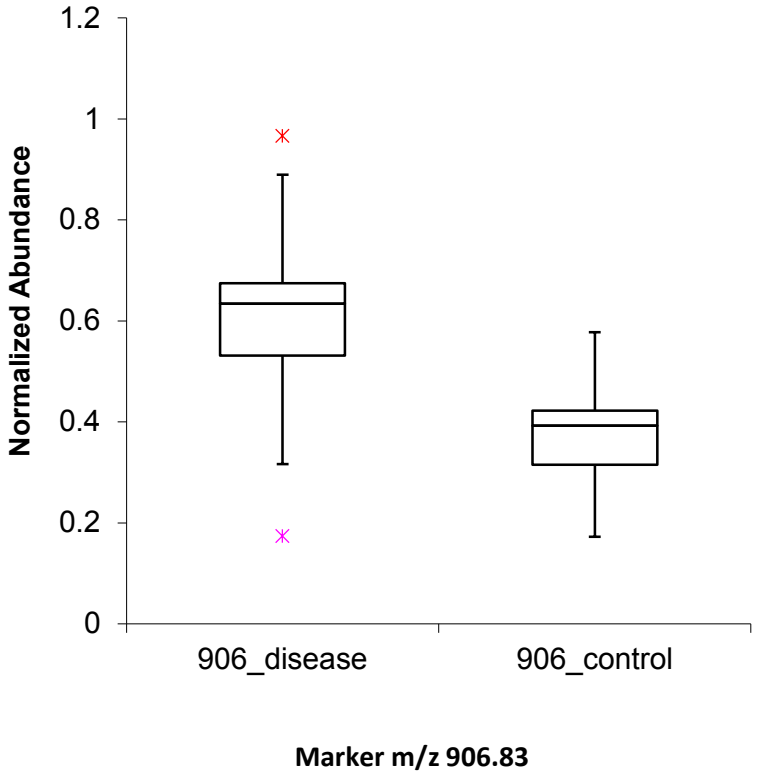
Figure 8. (a) Comparison of the panel of 3 colostrum lipid markers at m/z 570.4639, 586.5356 and 855.7473 of PRMD susceptible (n = 20) and PRMD resistant (n = 22) cows. SE: standard error (b) Boxplot of colostrum marker m/z 570.46. (c) Boxplot of colostrum marker m/z 586.53. (d) Boxplot of colostrum marker m/z 855.74. The y-axis is the normalized abundance. The boxplot data were from 20 cows that developed PRMDs and 22 healthy controls.

	PRMD susceptible		PRMD resistant		
Milk Biomarker	mean	SE	mean	SE	P value
642.5606	0.2701	0.0219	0.4517	0.0197	2.8368E-07
906.8360	0.5884	0.0451	0.3652	0.0217	1.8597E-05

(a)



(b)

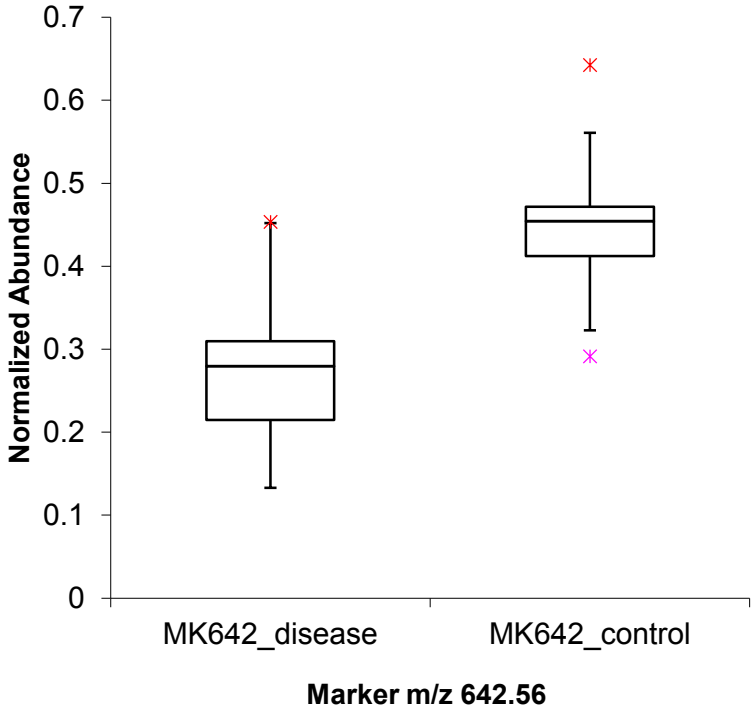


(c)

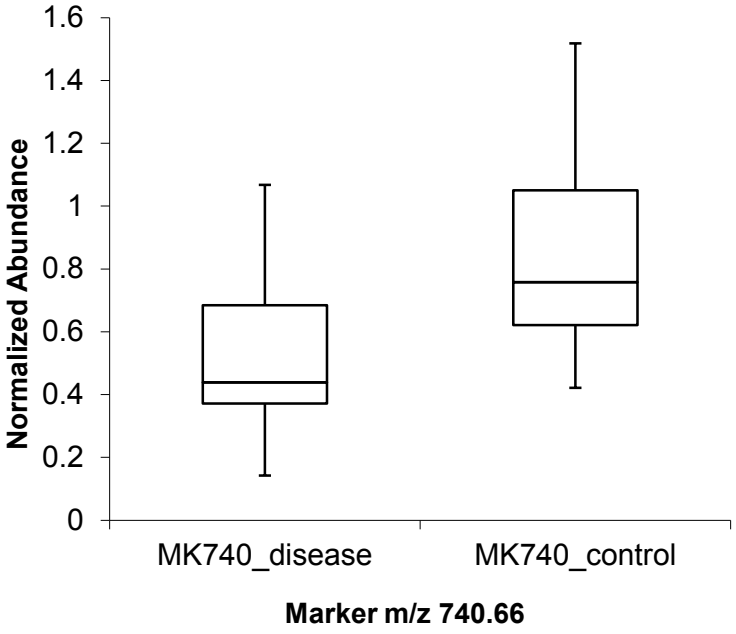
Figure 9. (a) Comparison of the 2 milk lipid markers at m/z 642.5606 and 906.836 yielding the optimum panel of PRMD susceptible (n = 16) and PRMD resistant (n = 20) cows. (b) Boxplot of milk marker m/z 642.56. (c) Boxplot of milk marker m/z 906.83. The y-axis represents normalized abundance. Boxplot was obtained from 16 cows that developed PRMDs and 20 healthy controls.

	PRMD susceptible		PRMD resistant		
Comibined Biomarker	mean	SE	mean	SE	P value
MK_642.5606	0.2701	0.0219	0.4517	0.0197	2.8368E-07
MK_740.6681	0.4930	0.0594	0.8341	0.0670	3.7069E-04
CS_344.2247	0.1148	0.0059	0.1036	0.0026	4.2062E-02
CS_682.5948	2.5905	0.3287	1.4048	0.1624	9.4922E-04
DF_652.5473	-0.0635	0.0124	-0.1447	0.0108	<0.001
DF_906.8353	-0.4005	0.0355	-0.2649	0.0217	6.1400E-02
DF_919.8341	-0.1624	0.0193	-0.1202	0.0119	1.8000E-03

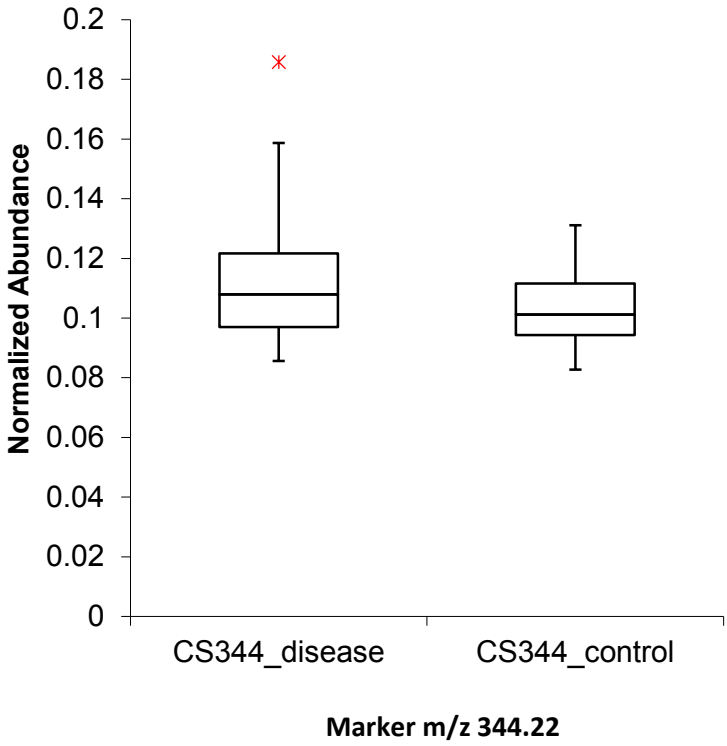
(a)



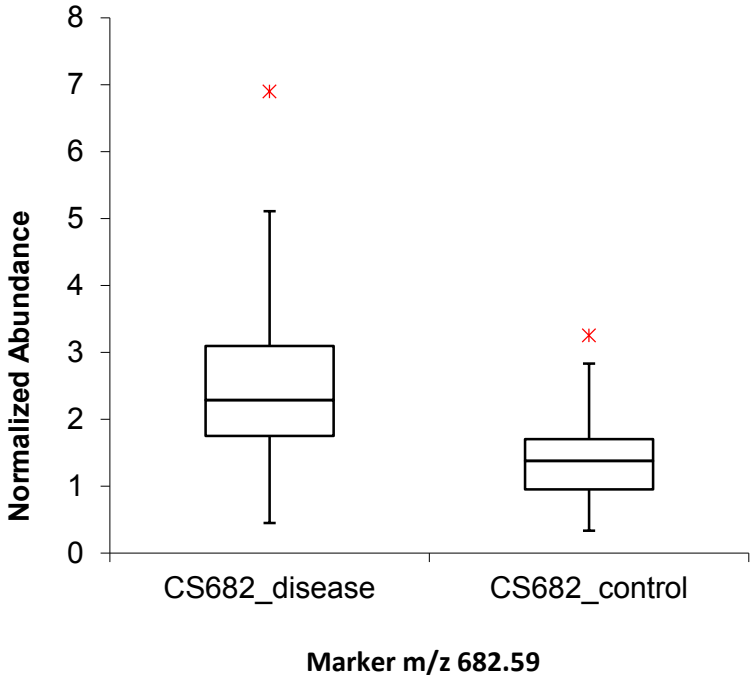
(b)



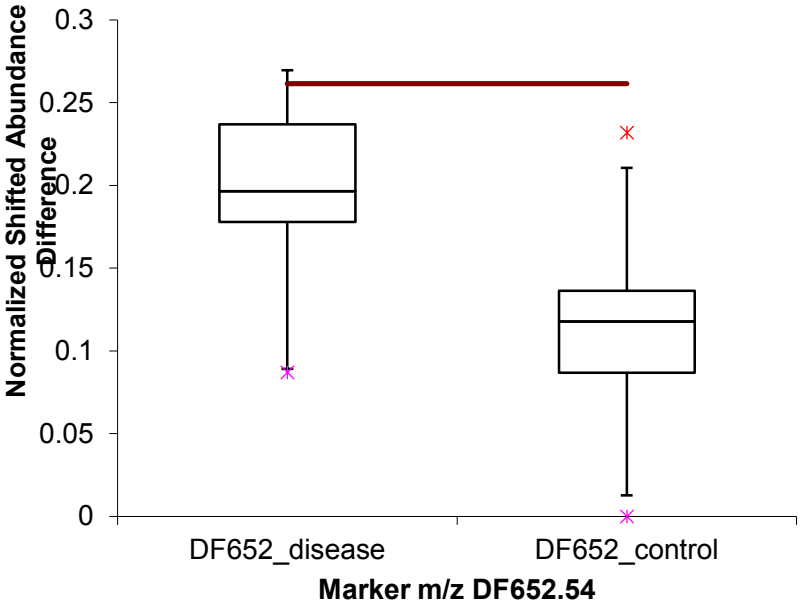
(c)



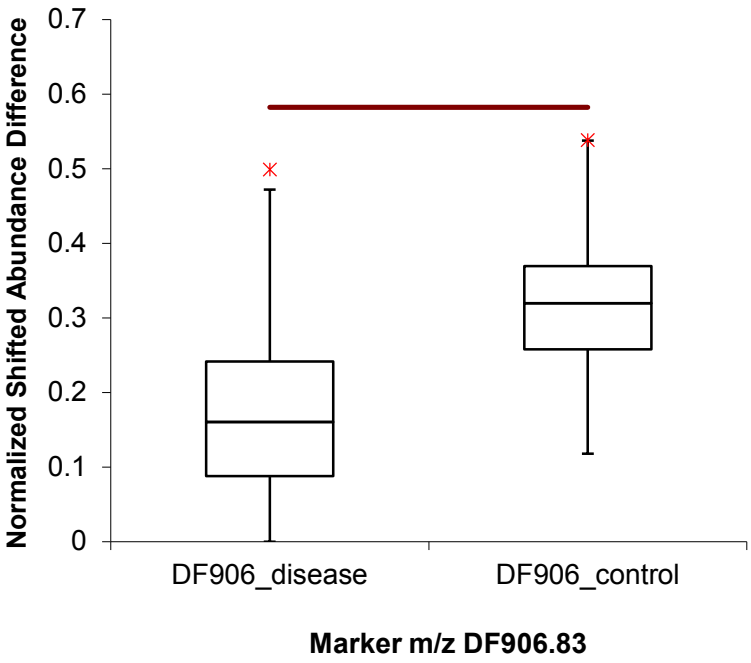
(d)



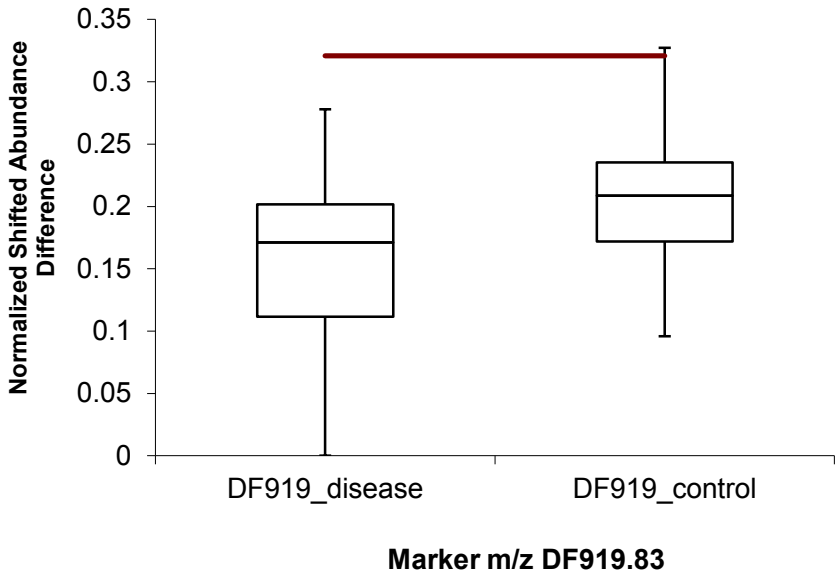
(e)



(f)



(g)



(h)

Figure 10. (a) Representation of a panel of 7 predictive PRMD lipid biomarkers. MK: milk; CS: colostrum; DF: difference for shared markers; DF = normalized abundance of CS – normalized abundance of MK; Milk lipids m/z 642.5606 and 740.6681 and colostrum lipids m/z 344.2247 and 682.5948 were selected for the panel. Additionally, three calculated differences for shared markers (different in both colostrum and milk) with statistical differences between animals that later developed PRMDs and cases were included. Lipids with m/z 652.5473, 906.8353 and 919.8341 shown as negative mean values indicated the normalized abundance of the biomarker in milk was higher than in colostrum. (b) Boxplot of milk lipid m/z 642.56. (c) Boxplot of milk lipid m/z 740.66. (d) Boxplot of colostrum lipid m/z 344.22. (e) Boxplot of colostrum lipid m/z 682.59. (f) Boxplot of normalized shifted abundance differences of colostrum and milk marker m/z 652.54. Y-axis (normalized DF abundance) was shifted by 0.2615. (g) Boxplot of normalized abundance differences of colostrum and milk marker m/z 906.83. Y-axis was shifted by 0.5821. (h) Boxplot of normalized abundance differences of colostrum and milk marker m/z 919.83. Y-axis of normalized DF abundance was shifted by 0.3208.

Characterization of PRMD biomarker candidates:

Lipid marker structures were identified by means of targeted MS/MS analyses on the QSTAR and/or QqTOF using collisionally-induced dissociation (CID). The markers that made up the three panels were submitted to this further characterization. This represented 10 unique lipid biomarkers that needed to be chemically characterized. Of these, 5 markers were successfully identified as triacylglycerols (TG), including m/z 855.7473 which represented the protonated TG (16:0/18:1/18:3) determined based on 2 abundant fragments at m/z 573.49 and m/z 599.48. The marker m/z 906.836 was first determined to be an ammoniated TG because of a peak at $[M+NH_4 -17]^+$. Utilizing this same approach, marker m/z 740.6681 was characterized as $[TG (12:0/14:0/16:0)+NH_4]^+$. Identification studies on the marker m/z 919.8341 yielded two possibilities, including an oxidized TG (18:0/18:0/19:1)+OH^+ or an oxidized TG (18:0/18:1/19:0)+OH^+. As for the marker m/z 682.5948, a NH_3 neutral loss was observed in the spectra. And the exact mass of marker m/z 664.56 suggested its elemental composition to be $C_{41}H_{76}O_6$, which indicates strongly that the marker is a TG. Three of the 10 markers were categorized as diacylglycerols (DG) through the same approach with determinations based on the fragments and neutral losses in the fragmentation spectra. The fragmentation studies identified these 3 markers to be DG (16:0/16:0), DG (18:0/18:0) and DG (18:2/19:0).

The elemental composition of the last two lipid markers m/z 570.4639 and m/z 344.2247 were determined as $[C_{40}H_{56}O+NH_4]^+$ and $[C_{21}H_{26}O_3+NH_4]^+$. However, these two markers were not classified into a specific lipid group due to lack of identifiable headgroups or recognizable constituent species in the fragmentation data.

Predictors	Elemental composition	Identification	Abundance	P Value

570.4639	$C_{40}H_{56}O+NH_4$	NA	Higher in diseases	0.013504484
586.5356	$C_{35}H_{68}O_5+NH_4$	DG(16:0/16:0)	Higher in controls	0.001414111
855.7473	$C_{55}H_{98}O_6+H$	TG(16:0/18:1/18:3)	Higher in Diseases	0.000650882

Table 7. Identification of predictive PRMD lipid markers that were part of the optimized panel for colostrum. These provided 90.0% sensitivity at 86.4% specificity. This panel predicted 19 out of 22 control cows and 18 out of 20 cows with later PRMDs.

Predictors	Elemental composition	Identification	Abundance	P Value

642.5606	$C_{39}H_{76}O_5+NH_4$	DG(18:0/18:0)	Higher in controls	2.83684E-07
906.836	$C_{57}H_{108}O_6+NH_4$	TG(18:0/18:0/18:1)	Higher in diseases	1.85972E-05

Table 8. Characterization of predictive PRMD milk biomarkers. An optimized predictor panel of milk lipids provided 75.0% sensitivity at 90.0% specificity, by predicting 12 out of 16 cows that later developed PRMDs and 18 out of 20 cows that remained healthy that were used as controls.

Predictors	Elemental	Identification	Abundance	P Value

	composition			
MK642.5606	$C_{39}H_{76}O_5+NH_4$	DG(18:0/18:0)	Higher in controls	2.83684E-07
CS682.5948	$C_{41}H_{76}O_6+NH_4$	TG	Higher in diseases	0.000949216
CS344.2247	$C_{21}H_{26}O_3+NH_4$	NA	Higher in diseases	0.042061688
DF906.8353	$C_{57}H_{108}O_6$ $+NH_4$	TG(18:0/18:0/18:1)	Higher changes in diseases	0.0614
DF919.8341	$C_{58}H_{110}O_6$+OH	TG (18:0/18:0/19:1)+OH; TG (18:0/18:1/19:0)+OH	Higher changes in diseases	0.0018
DF652.5473	$C_{40}H_{74}O_5+NH_4$	DG(18:2/19:0)	Higher changes in controls	<.0001
MK740.6681	$C_{45}H_{86}O_6+NH_4$	TG(12:0/14:0/16:0)	Higher in controls	0.000370689

Table 9. Characterization of predictive PRMD biomarkers combining colostrum, milk and shared markers. An optimized predictor panel using a combination of colostrum, milk and the differences between milk and colostrum for 'shared' lipids found in both of these secretions showing significant differences in both colostrum and milk, provided 87.5% sensitivity at a specificity of 100.0%, by predicting 14 out of 16 cows that later developed PRMDs and 20 out of 20 cows that remained healthy and served as controls. The difference (DF) between the normalized abundance of the biomarker for lipids with m/z 652.5473, 906.8353 and 919.8341 indicated it was higher in milk than in colostrum.

3.6 Discussion

Production-related metabolic diseases (PRMDs) have one of the highest incidences of any disease in dairy cows worldwide, e.g., on average milk fever occurs in ~7% of all animals after calving and displaced abomasum in ~5% of cows in well managed dairy herds, which requires appropriate management and treatments that may include surgical intervention or even loss of the animal depending on the severity of the disease at the time of diagnosis. A condition like left displaced abomasum (LDA) can require increased costs for labor, drugs and/or surgery, as well as reduced milk production, fertility issues, and possibly death of the cow. Prevention of PRMDs is more efficient and much cheaper than any treatment for the aforementioned diseases. Thus, a useful and easy way of predicting cows at risk for PRMDs in advance of clinical signs is necessary and would be significant in helping preserve dairy cow health to enhance overall herd production and to lower economic losses.

Previous attempts to predict cows at risk for PRMDs have focused on circulating levels of lipids in blood plasma or serum. Animals with an elevated serum NEFA concentration (more than 0.3 mEq/L) 14 to 2 days prior to calving or animals with an elevated serum BHBA concentration (more than 10 mg/dL) and NEFA concentration (more than 0.6 mEq/L) 3 to 14 days postpartum, have been shown to be at an increased risk of transition diseases in dairy cows (see Figure 2 for a summary of specificity and sensitivity using NEFA and BHBA as predictors for transition diseases).[42] However, the sensitivity and specificity of both of these markers are inadequate to be useful, are labor intense to collect venipuncture derived samples, and perform much less well than the biomarkers described here. We recognize that more samples from geographically distinct regions are needed for analysis in order to validate our results, to be confident that the observed biomarkers from our study are

present in other cows from unrelated herds. The goal being to identify early, more efficient and hence valuable lipid biomarkers in substrates that are readily available, easy to collect such as colostrum or milk, and that are predictive of imminent PRMDs in asymptomatic postpartum dairy cows.

A comprehensive lipidomics approach employing ESI-MS was for the first time applied to lipid quantification and the selection of potential PRMD markers in day 1 postpartum colostrum and day 4 postpartum milk. Here, 10 mg of the cream layer of colostrum or milk was used for lipid extraction, and the resulting extract diluted 1000 times before loading the specimen onto the ESI-MS. This dilution allowed most of the observed lipids to fall within the linear dynamic concentration range of the instrument, i.e. the range over which ion signal is directly proportional to the analyte concentration. Most of the lipids in colostrum or milk are neutral lipids such as glycerolipids, and consequently a concentration of chloroform twice that of methanol was used for the extraction step in order to obtain the broadest representation of lipids in the organic phase. Statistical linear discriminative analysis was then used to create different panels of biomarkers having greater predictive sensitivity and specificity than individual markers. After developing optimal predictive biomarker panels, targeted MS/MS analysis was performed to further chemically characterize the 10 relevant lipid biomarkers obtained from statistical remodeling.

As shown in Table 4, among the 61 statistically significant biomarker candidates in colostrum, 55 of them were present in higher concentration based on their normalized abundances in dairy cows that subsequently developed PRMDs. The other 6 lipids were present in higher levels in healthy controls. As previously demonstrated, a number of

compounds, most notably IgG, nutrients (e.g., vitamins, lipids) and other serum proteins, are secreted as colostrum accumulates in the mammary gland several weeks before calving under the influence of various lactogenic hormones, including prolactin, 17β-estradiol, insulin, etc[83,84]. The differences found in the aforementioned lipid biomarker candidates that were secreted in colostrum and/or milk from cows that either developed PRMD or remained healthy, may have occurred during the period prior to calving. Until now, there have been few if any research studies carried out on day 1 colostrum as part of lipid biomarker discovery.

There was a finding in the post-partum day 4 milk specimens, lipid biomarkers that suggested a significant change in the biological composition of triglycerides in milk between animals destined to develop PRMDs and those animals that remained healthy with normal production status. Among the 77 statistically significant biomarker candidates in milk listed in Table 5, those having m/z values from 572.48 to 810.74 (n = 45) were all higher in controls, whereas those having m/z values from 848.76 to 919.83 (30 of 32) were higher in animals that later developed disease. The remaining 2 statistically significant biomarker candidates having m/z values of 939.68 and 967.71 were both higher in controls. Overall these lipid 'class' distinctions in milk were significant ($p=2.3x10^{-19}$). All of these milk lipid biomarkers appear to be triglycerides (TG). It has been previously established that the mammary gland of the cow synthesizes *de novo* fatty acids with an even number of carbons of the 4:0-14:0 acids together with about half of the 16:0 from acetate and β-hydroxybutyrate that accounts for approximately 60 and 45% of the fatty acids present in milk on a molar and weight basis, respectively[85]. The remaining 40-60% of milk fatty acids are longer-chain 16:0 and

predominantly C_{18} fatty acids derived mainly from dietary lipids, but are also from lipolysis of adipose tissue triacylglycerols, which make their way into the circulation, mainly as plasma NEFA and triglyceride-rich lipoproteins[85]. Thus, fatty acids incorporated into milk fat TGs are provided either from mammary *de novo* synthesis or from the uptake of preformed fatty acids from the peripheral circulation. Figure 3 illustrates the milk fat TG synthesis pathway. Even though we did not determine the structure of all the milk lipid biomarkers, those lipids with m/z values from 848.76 to 919.83 appear to represent longer chain TG based on their m/z values and on the analysis of selected lipids in this range. Indeed, those milk biomarkers that were included in the optimal panel, then fragmented and characterized having m/z values of 906.83 and m/z 919.8341 were found to be TGs having longer chain fatty acid components. These were higher in clinically normal cows that later developed PRMDs (see Table 9). Milk lipids in the range of m/z 510 to ~810 are likely to be TGs composed of shorter chain fatty acids. For example, the biomarker at m/z 740.668 was characterized as a TG having shorter chain fatty acids present, i.e., 12:0/14:0/16:0 and was found in higher levels in controls. Since the dairy cows that developed PRMDs and the matched control group cows selected for this study were provided with the same feed, the differences in the lipid biomarker concentrations between affected and healthy animals cannot be attributed to diet.

The long chain fatty acid C18:1 cis-9 has been previously proposed as a possible biomarker able to diagnose elevated concentrations of plasma NEFA[76]. As for hyperketonemia, 90% of non-hyperketonemic controls showed a milk fat C18:1 cis-9-to-C15:0 ratio of 40 or lower, whereas 70% of cows suffering from hyperketonemia

showed milk fat C18:1 cis-9-to-C15:0 ratios exceeding 40, which is consistent with our findings of potential different trends between longer chain TGs and shorter chain TGs[86]. Perhaps the higher expressed longer chain TGs in diseased animals is an early compensatory response to reduced maternal short chain fatty acid production *de novo* by the udder.

As to the reduced levels of short chain fatty acid TGs seen in the PRMD susceptible animals, this might be a foreshadowing of those particular cows having limited ability to generate adequate amounts of their own TGs which compromises the health of the animal as production demands remain very high at the onset of lactation.

However, in colostrum the results were quite different. The statistically significant biomarkers in both the short chain fatty acid and longer chain fatty acid TGs were both elevated in the colostrum of animals later developing PRMDs. Clearly colostrum production and milk production are representative of different stages in lactation. It has been observed that the colostrum secreted within 24 hr after calving has a distinct fat composition compared with the secretion produced on day 4 after calving in dairy cows[87]. Nevertheless, in colostrum all lipid biomarkers representative of both short chain and long chain fatty acid TGs are significantly increased in those animals that later developed PRMDs. Given the timing of colostrum versus milk production, the data strongly suggests that there are preexisting problems prior to calving in those animals that later develop PRMDs that lead to increased TGs in colostrum, and that the demands of milk production as early as day 4 postpartum deplete endogenously produced TGs quickly in animals destined to develop PRMDs. Collectively, the data suggest that there are profoundly compromised biological pathways that lead to or reflect PRMDs, but these remain to be identified.

Of those markers that were part of the optimized panel, 8 out of 10 lipid biomarkers were successfully identified. Even though it was challenging to confidently determine the lipid class of colostrum biomarker m/z 344.2247, based on the most likely elemental composition determined as $C_{21}H_{26}O_3$+NH_4 predicted by exact mass studies, it would suggest that it is an oxidized lipid. This marker showed a higher normalized intensity in the colostrum of case animals compared with healthy controls.

The colostrum, milk and optimized mixed panels of lipid markers described here may be useful biomarkers for later routine application. All have shown predictive abilities with higher than 75.0% sensitivity and specificity. The one panel optimized using linear discriminative analysis demonstrated that there was complete separation between cows destined to develop PRMDs and those that had an uncomplicated postpartum course. This panel combined biomarkers from colostrum, milk and those lipids that were found in both colostrum and milk that were significantly different in both specimens (see Figure 11). Clearly, further confirmation needs to occur with a more focused effort on the development of a specific risk index, i.e., the quantitative likelihood of PRMDs based on specific quantities of the several biomarkers as part of a panel. Ideally, a single collection of colostrum and milk with lipidomic measurement of targeted lipids could be converted into percent likelihood of an animal developing a PRMD. Also, additional studies will need to be done surveying broader geographic regions and different dairy breeds to confirm the effectiveness of the current biomarkers across all dairy cows.

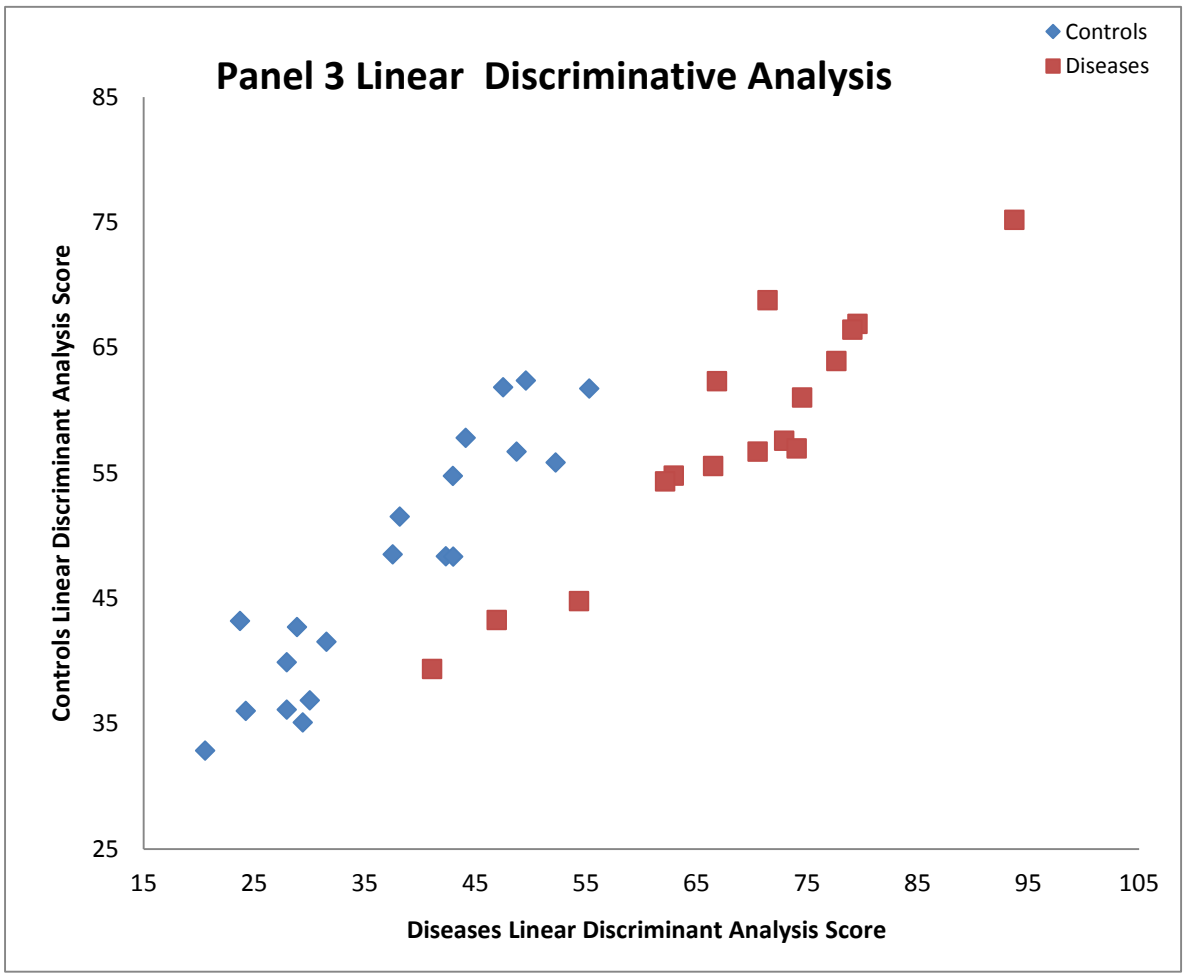


Figure 11. Linear discriminative analysis of the optimal lipid biomarkers demonstrating separation between modeled biomarker values for animals developing PRMDs later (red) and healthy control animals (blue). The x-axis and y-axis represent the discriminant function scores.

Chapter 4 Conclusions and future work

4.1 Summary of Bacteriophage NS01 proteomics study

Using the combination of SDS-PAGE and HPLC-MS-MS, bacteriophage NS01 proteins were successfully characterized with high probability scores and good coverage according to the output results from the search engine Mascot. The protein characterization approach was a multi-step process including NS01 purification from salts, SDS gradient gel protein separation, excised gel protein digestion with trypsin, digested protein analyzed by HPLC-MS-MS, and further data processing with software. SDS-PAGE combined with HPLC-MS-MS techniques helped overcome MS ion suppression issues and characterized proteins with high confidence (p value <0.01). The hardest part of this project, however, was purification step, which required a complete buffer exchange of the bacteriophage before loading it onto the SDS-PAGE. This study expanded our knowledge regarding the extremophilic virus NS01 allowing it to be categorized as a turreted T7-like virus with flexible HK-97 subunits according to the major capsid protein gp88 secondary structure predications as well as the putative tail protein gp86.

However, there still exist some aspects of the research that can be improved. For example, two-dimensional gel electrophoresis (2-DE) could be used to separate proteins not only based on their molecular weights but also on their isoelectric points, allowing for many more bacteriophage proteins to be isolated potentially and studied. See our gel picture Figure 4. There were 4 gel bands identified as gp91/68, gp86/78, gp72/69 and gp88/73 that because of their similar sizes were not fully separated. However, if 2-DE had been applied, protein isoelectric points may have helped to fully clarify more of the bacteriophage

proteome.

4.2 Summary of biomarker study for PRMDs through a lipidomics approach

This project worked on finding and identifying predictive biomarkers for PRMDs present in colostrum and day 4 postpartum milk using mass spectrometry through a lipidomics approach. It successfully discovered several individual candidate lipid markers predictive of PRMDs as well as three panels of biomarkers resulting from statistical modeling. The three panels of biomarkers provided adequate sensitivity and specificity, all higher than 75.0%, which would be generally acceptable for use in a real life setting.

Even though this project has shown promising results from the current three panels of lipid markers and the individual biomarker candidates in those panels, improvements still need to be made.

For example, internal lipid standards from diverse representative lipid groups such as synthetic TGs could be spiked into the 2:1:1.25 chloroform:methanol: isopropanol extraction solution as part of the lipid extraction step. If applicable, a neutral loss scan can be used for DG and TG detection and product fragments applied for lipid quantification similar to multiple reaction monitoring. In addition, further research can be carried out on lipids outside the panels to chemically characterize them more completely, which may provide more of an indication of the underlying biological pathway disturbances corresponding to metabolic disorders in affected dairy cows.

Furthermore, in order to be an accepted dairy application, additional blinded studies carried out at several geographic locations involving other breeds of dairy cows need to be

carried out with the same method to validate the effectiveness of the three panels of biomarkers. This may also require development of different biomarkers unique to different breeds. Also, specific milk or colostrum biomarkers might be evaluated for their ability to predict specific disease phenotypes. With even more research, one could test the response of animals to nutritional or other interventions to determine if biomarkers normalize with treatment and health improvement allowing for treatment assessment.

References

1. Han, X.; Rozen, S.; Boyle, S. H.; Hellegers, C.; Cheng, H.; Burke, J. R.; Welsh-Bohmer, K. A.; Doraiswamy, P. M.; Kaddurah-Daouk, R., Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. *PLoS One* 2011, *6* (7), e21643.

2. (a) Fox, T. E.; Bewley, M. C.; Unrath, K. A.; Pedersen, M. M.; Anderson, R. E.; Jung, D. Y.; Jefferson, L. S.; Kim, J. K.; Bronson, S. K.; Flanagan, J. M.; Kester, M., Circulating sphingolipid biomarkers in models of type 1 diabetes. *J Lipid Res* 2011, *52* (3), 509-17; (b) Fox, T. E.; Han, X.; Kelly, S.; Merrill, A. H., 2nd; Martin, R. E.; Anderson, R. E.; Gardner, T. W.; Kester, M., Diabetes alters sphingolipid metabolism in the retina: a potential mechanism of cell death in diabetic retinopathy. *Diabetes* 2006, *55* (12), 3573-80.

3. Fan, M.; Sidhu, R.; Fujiwara, H.; Tortelli, B.; Zhang, J.; Davidson, C.; Walkley, S. U.; Bagel, J. H.; Vite, C.; Yanjanin, N. M.; Porter, F. D.; Schaffer, J. E.; Ory, D. S., Identification of Niemann-Pick C1 disease biomarkers through sphingolipid profiling. *J Lipid Res* 2013, *54* (10), 2800-14.

4. Loomba, R.; Quehenberger, O.; Armando, A.; Dennis, E. A., Polyunsaturated fatty acid metabolites as novel lipidomic biomarkers for noninvasive diagnosis of nonalcoholic steatohepatitis. *J Lipid Res* 2015, *56* (1), 185-92.

5. Pan, Y.; Zhou, H.; Mahsut, A.; Rohm, R. J.; Berejnaia, O.; Price, O.; Chen, Y.; Castro-Perez, J.; Lassman, M. E.; McLaren, D.; Conway, J.; Jensen, K. K.; Thomas, T.; Reyes-Soffer, G.; Ginsberg, H. N.; Gutstein, D. E.; Cleary, M.; Previs, S. F.; Roddy, T. P., Static and turnover

kinetic measurement of protein biomarkers involved in triglyceride metabolism including apoB48 and apoA5 by LC/MS/MS. *J Lipid Res* 2014, *55* (6), 1179-1187.

6. Gorden, D. L.; Myers, D. S.; Ivanova, P. T.; Fahy, E.; Maurya, M. R.; Gupta, S.; Min, J.; Spann, N. J.; McDonald, J. G.; Kelly, S. L.; Duan, J.; Sullards, M. C.; Leiker, T. J.; Barkley, R. M.; Quehenberger, O.; Armando, A. M.; Milne, S. B.; Mathews, T. P.; Armstrong, M. D.; Li, C.; Melvin, W. V.; Clements, R. H.; Washington, M. K.; Mendonsa, A. M.; Witztum, J. L.; Guan, Z.; Glass, C. K.; Murphy, R. C.; Dennis, E. A.; Merrill, A. H., Jr.; Russell, D. W.; Subramaniam, S.; Brown, H. A., Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J Lipid Res* 2015, *56* (3), 722-36.

7. Barber, M. N.; Risis, S.; Yang, C.; Meikle, P. J.; Staples, M.; Febbraio, M. A.; Bruce, C. R., Plasma lysophosphatidylcholine levels are reduced in obesity and type 2 diabetes. *PLoS One* 2012, *7* (7), e41456.

8. Demirkan, A.; van Duijn, C. M.; Ugocsai, P.; Isaacs, A.; Pramstaller, P. P.; Liebisch, G.; Wilson, J. F.; Johansson, A.; Rudan, I.; Aulchenko, Y. S.; Kirichenko, A. V.; Janssens, A. C.; Jansen, R. C.; Gnewuch, C.; Domingues, F. S.; Pattaro, C.; Wild, S. H.; Jonasson, I.; Polasek, O.; Zorkoltseva, I. V.; Hofman, A.; Karssen, L. C.; Struchalin, M.; Floyd, J.; Igl, W.; Biloglav, Z.; Broer, L.; Pfeufer, A.; Pichler, I.; Campbell, S.; Zaboli, G.; Kolcic, I.; Rivadeneira, F.; Huffman, J.; Hastie, N. D.; Uitterlinden, A.; Franke, L.; Franklin, C. S.; Vitart, V.; Consortium, D.; Nelson, C. P.; Preuss, M.; Consortium, C. A.; Bis, J. C.; O'Donnell, C. J.; Franceschini, N.; Consortium, C.; Witteman, J. C.; Axenovich, T.; Oostra, B. A.; Meitinger, T.; Hicks, A. A.; Hayward, C.; Wright, A. F.; Gyllensten, U.; Campbell, H.; Schmitz, G.; consortium, E., Genome-wide

association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet* 2012, *8* (2), e1002490.

9. Wang, C.; Kong, H.; Guan, Y.; Yang, J.; Gu, J.; Yang, S.; Xu, G., Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis. *Anal Chem* 2005, *77* (13), 4108-16.

10. Lee, S. J.; Zhang, J.; Choi, A. M.; Kim, H. P., Mitochondrial dysfunction induces formation of lipid droplets as a generalized response to stress. *Oxid Med Cell Longev* 2013, *2013*, 327167.

11. Pettegrew, J. W.; Panchalingam, K.; Hamilton, R. L.; McClure, R. J., Brain membrane phospholipid alterations in Alzheimer's disease. *Neurochem Res* 2001, *26* (7), 771-82.

12. Mielke, M. M.; Lyketsos, C. G., Alterations of the sphingolipid pathway in Alzheimer's disease: new biomarkers and treatment targets? *Neuromolecular Med* 2010, *12* (4), 331-40.

13. Gonzalez-Dominguez, R.; Garcia-Barrera, T.; Gomez-Ariza, J. L., Metabolomic study of lipids in serum for biomarker discovery in Alzheimer's disease using direct infusion mass spectrometry. *J Pharm Biomed Anal* 2014, *98*, 321-6.

14. Han, X.; D, M. H.; McKeel, D. W., Jr.; Kelley, J.; Morris, J. C., Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer's disease: potential role in disease pathogenesis. *J Neurochem* 2002, *82* (4), 809-18.

15. Ide, Y.; Waki, M.; Hayasaka, T.; Nishio, T.; Morita, Y.; Tanaka, H.; Sasaki, T.; Koizumi, K.; Matsunuma, R.; Hosokawa, Y.; Ogura, H.; Shiiya, N.; Setou, M., Human breast cancer tissues

contain abundant phosphatidylcholine(36ratio1) with high stearoyl-CoA desaturase-1 expression. *PLoS One* 2013, *8* (4), e61204.

16. Love, R. R.; Newcomb, P. A.; Wiebe, D. A.; Surawicz, T. S.; Jordan, V. C.; Carbone, P. P.; DeMets, D. L., Effects of tamoxifen therapy on lipid and lipoprotein levels in postmenopausal patients with node-negative breast cancer. *Journal of the National Cancer Institute* 1990, *82* (16), 1327-32.

17. Sutphen, R.; Xu, Y.; Wilbanks, G. D.; Fiorica, J.; Grendys, E. C., Jr.; LaPolla, J. P.; Arango, H.; Hoffman, M. S.; Martino, M.; Wakeley, K.; Griffin, D.; Blanco, R. W.; Cantor, A. B.; Xiao, Y. J.; Krischer, J. P., Lysophospholipids are potential biomarkers of ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2004, *13* (7), 1185-91.

18. Ried, J. S.; Baurecht, H.; Stuckler, F.; Krumsiek, J.; Gieger, C.; Heinrich, J.; Kabesch, M.; Prehn, C.; Peters, A.; Rodriguez, E.; Schulz, H.; Strauch, K.; Suhre, K.; Wang-Sattler, R.; Wichmann, H. E.; Theis, F. J.; Illig, T.; Adamski, J.; Weidinger, S., Integrative genetic and metabolite profiling analysis suggests altered phosphatidylcholine metabolism in asthma. *Allergy* 2013, *68* (5), 629-36.

19. Luster, A. D.; Tager, A. M., T-cell trafficking in asthma: lipid mediators grease the way. *Nature reviews. Immunology* 2004, *4* (9), 711-24.

20. Han, X.; Gross, R. W., Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom Rev* 2005, *24* (3), 367-412.

21. Taguchi, R.; Houjou, T.; Nakanishi, H.; Yamazaki, T.; Ishida, M.; Imagawa, M.; Shimizu, T., Focused lipidomics by tandem mass spectrometry. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 2005, *823* (1), 26-36.

22. Balazy, M., Eicosanomics: targeted lipidomics of eicosanoids in biological systems. *Prostaglandins & Other Lipid Mediators* 2004, *73* (3-4), 173-80.

23. Houjou, T.; Yamatani, K.; Imagawa, M.; Shimizu, T.; Taguchi, R., A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2005, *19* (5), 654-66.

24. Ekroos, K.; Chernushevich, I. V.; Simons, K.; Shevchenko, A., Quantitative profiling of phospholipids by multiple precursor ion scanning on a hybrid quadrupole time-of-flight mass spectrometer. *Anal Chem* 2002, *74* (5), 941-9.

25. Kayganich-Harrison, K. A.; Rose, D. M.; Murphy, R. C.; Morrow, J. D.; Roberts, L. J., 2nd, Collision-induced dissociation of F2-isoprostane-containing phospholipids. *J Lipid Res* 1993, *34* (7), 1229-35.

26. Argoudelis, C. J.; Perkins, E. G., Determination of double bond position in mono-unsaturated fatty acids using combination gas chromatography mass spectrometry. *Lipids* 1968, *3* (4), 379-81.

27. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & Cellular Proteomics : MCP* 2012, *11* (11), 1475-88.

28. Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S., LIPID MAPS online tools for lipid research. *Nucleic Acids Research* 2007, *35* (Web Server issue), W606-12.

29. Reis, A.; Rudnitskaya, A.; Blackburn, G. J.; Mohd Fauzi, N.; Pitt, A. R.; Spickett, C. M., A comparison of five lipid extraction solvent systems for lipidomic studies of human LDL. *J Lipid Res* 2013, *54* (7), 1812-24.

30. Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D., Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res* 2008, *49* (5), 1137-46.

31. Folch, J.; Lees, M.; Sloane Stanley, G. H., A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* 1957, *226* (1), 497-509.

32. Iverson, S. J.; Lang, S. L.; Cooper, M. H., Comparison of the Bligh and Dyer and Folch methods for total lipid determination in a broad range of marine tissue. *Lipids* 2001, *36* (11), 1283-7.

33. Perez-Palacios, T.; Ruiz, J.; Martin, D.; Muriel, E.; Antequera, T., Comparison of different methods for total lipid quantification in meat and meat products. *Food Chemistry* 2008, *110* (4), 1025-9.

34. Block, E., Manipulation of dietary cation-anion difference on nutritionally related production diseases, productivity, and metabolic responses of dairy cows. *J Dairy Sci* 1994, *77* (5), 1437-50.

35. Rukkwamsuk, T.; Wensing, T.; Geelen, M. J., Effect of overfeeding during the dry period on the rate of esterification in adipose tissue of dairy cows during the periparturient period. *J Dairy Sci* 1999, *82* (6), 1164-9.

36. Grummer, R. R., Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. *Journal of Animal Science* 1995, *73* (9), 2820-33.

37. Curtis, C. R.; Erb, H. N.; Sniffen, C. J.; Smith, R. D.; Kronfeld, D. S., Path analysis of dry period nutrition, postpartum metabolic and reproductive disorders, and mastitis in Holstein cows. *J Dairy Sci* 1985, *68* (9), 2347-60.

38. de Vries, M. J.; Veerkamp, R. F., Energy balance of dairy cattle in relation to milk production variables and fertility. *J Dairy Sci* 2000, *83* (1), 62-9.

39. Cameron, R. E.; Dyk, P. B.; Herdt, T. H.; Kaneene, J. B.; Miller, R.; Bucholtz, H. F.; Liesman, J. S.; Vandehaar, M. J.; Emery, R. S., Dry cow diet, management, and energy balance as risk factors for displaced abomasum in high producing dairy herds. *J Dairy Sci* 1998, *81* (1), 132-9.

40. Nachtomi, E.; Eger, S.; Amir, S.; Schindler, H., Postpartum Nonesterified Fatty-Acids Concentration in Blood-Plasma of Dairy-Cows Fed Different Energy-Levels Prepartum. *Nutr Rep Int* 1986, *34* (4), 521-527.

41. Cavestany, D.; Blanc, J. E.; Kulcsar, M.; Uriarte, G.; Chilibroste, P.; Meikle, A.; Febel, H.; Ferraris, A.; Krall, E., Studies of the transition cow under a pasture-based milk production system: metabolic profiles. *Journal of Veterinary Medicine. A, Physiology, Pathology, Clinical Medicine* 2005, *52* (1), 1-7.

42. Ospina, P. A.; Nydam, D. V.; Stokol, T.; Overton, T. R., Evaluation of nonesterified fatty acids and beta-hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. *J Dairy Sci* 2010, *93* (2), 546-54.

43. Adewuyi, A. A.; Gruys, E.; van Eerdenburg, F. J., Non esterified fatty acids (NEFA) in dairy cattle. A review. *The Veterinary Quarterly* 2005, *27* (3), 117-26.

44. Duffield, T. F.; Lissemore, K. D.; McBride, B. W.; Leslie, K. E., Impact of hyperketonemia in early lactation dairy cows on health and production. *J Dairy Sci* 2009, *92* (2), 571-80.

45. Kaneene, J. B.; Miller, R.; Herdt, T. H.; Gardiner, J. C., The association of serum nonesterified fatty acids and cholesterol, management and feeding practices with peripartum disease in dairy cows. *Preventive Veterinary Medicine* 1997, *31* (1-2), 59-72.

46. Chilliard, Y.; Ferlay, A.; Faulconnier, Y.; Bonnet, M.; Rouel, J.; Bocquier, F., Adipose tissue metabolism and its role in adaptations to undernutrition in ruminants. *The Proceedings of the Nutrition Society* 2000, *59* (1), 127-34.

47. Yoder, J. D.; Chen, T. S.; Gagnier, C. R.; Vemulapalli, S.; Maier, C. S.; Hruby, D. E., Pox proteomics: mass spectrometry analysis and identification of Vaccinia virion proteins. *Virol J* 2006, *3(10)*, 10.

48. Bell, C.; Desjardins, M.; Thibault, P.; Radtke, K., Proteomics analysis of herpes simplex virus type 1-infected cells reveals dynamic changes of viral protein expression, ubiquitylation, and phosphorylation. *J Proteome Res* 2013, *12* (4), 1820-9.

49. Morrissey, B.; Downard, K. M., A proteomics approach to survey the antigenicity of the influenza virus by mass spectrometry. *Proteomics* 2006, *6* (7), 2034-41.

50. Ravichandran, V.; Major, E. O., Viral proteomics: a promising approach for understanding JC virus tropism. *Proteomics* 2006, *6* (20), 5628-36.

51. Oren, A.; Bratbak, G.; Heldal, M., Occurrence of virus-like particles in the Dead Sea. *Extremophiles : life under Extreme Conditions* 1997, *1* (3), 143-9.

52. Jaalinoja, H. T.; Roine, E.; Laurinmaki, P.; Kivela, H. M.; Bamford, D. H.; Butcher, S. J., Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus. *Proceedings of the National Academy of Sciences of the United States of America* 2008, *105* (23), 8008-13.

53. Khayat, R.; Tang, L.; Larson, E. T.; Lawrence, C. M.; Young, M.; Johnson, J. E., Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102* (52), 18944-9.

54. Happonen, L. J.; Redder, P.; Peng, X.; Reigstad, L. J.; Prangishvili, D.; Butcher, S. J., Familial relationships in hyperthermo- and acidophilic archaeal viruses. *Journal of Virology* 2010, *84* (9), 4747-54.

55. Shen, P. S.; Domek, M. J.; Sanz-Garcia, E.; Makaju, A.; Taylor, R. M.; Hoggan, R.; Culumber, M.; Oberg, C.; Breakwell, D. P.; Prince, J. T.; Belnap, D. M., Sequence and Structural Characterization of Great Salt Lake Bacteriophage CW02, a Member of the T7-like Supergroup. *Journal of Virology* 2012, *85(15),* 7907-17.

56. Pietila, M. K.; Laurinmaki, P.; Russell, D. A.; Ko, C. C.; Jacobs-Sera, D.; Butcher, S. J.; Bamford, D. H.; Hendrix, R. W., Insights into head-tailed viruses infecting extremely halophilic archaea. *Journal of Virology* 2013, *87* (6), 3248-60.

57. (a) Suttle, C. A., The Significance of Viruses to Mortality in Aquatic Microbial Communities. *Microbial ecology* 1994, *28* (2), 237-243; (b) Wommack, K. E.; Colwell, R. R., Virioplankton: viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews : MMBR* 2000, *64* (1), 69-114.

58. Liu, Q.; Richardson, C. C., Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 1993, *90* (5), 1761-5.

59. Seed, K. D.; Bodi, K. L.; Kropinski, A. M.; Ackermann, H. W.; Calderwood, S. B.; Qadri, F.; Camilli, A., Evidence of a dominant lineage of Vibrio cholerae-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *mBio* 2011, *2* (1), e00334-10.

60. Hardies, S. C.; Comeau, A. M.; Serwer, P.; Suttle, C. A., The complete sequence of marine bacteriophage VpV262 infecting vibrio parahaemolyticus indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology* 2003, *310* (2), 359-71.

61. Choi, K. H.; McPartland, J.; Kaganman, I.; Bowman, V. D.; Rothman-Denes, L. B.; Rossmann, M. G., Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *Journal of Molecular Biology* 2008, *378* (3), 726-36.

62. Besemer, J.; Lomsadze, A.; Borodovsky, M., GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* 2001, *29* (12), 2607-18.

63. Kwan, T.; Liu, J.; Dubow, M.; Gros, P.; Pelletier, J., Comparative genomic analysis of 18 Pseudomonas aeruginosa bacteriophages. *Journal of Bacteriology* 2006, *188* (3), 1184-7.

64. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* 2009, *6* (5), 359-62.

65. Foster, D. B.; Liu, T.; Rucker, J.; O'Meally, R. N.; Devine, L. R.; Cole, R. N.; O'Rourke, B., The cardiac acetyl-lysine proteome. *PLoS One* 2013, *8* (7), e67513.

66. Kelley, L. A.; Sternberg, M. J., Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* 2009, *4* (3), 363-71.

67. Steinbacher, S.; Baxa, U.; Miller, S.; Weintraub, A.; Seckler, R.; Huber, R., Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors. *Proc Natl Acad Sci U S A* 1996, *93* (20), 10584-8.

68. Steven, A. C.; Trus, B. L.; Maizel, J. V.; Unser, M.; Parry, D. A.; Wall, J. S.; Hainfeld, J. F.; Studier, F. W., Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. *J Mol Biol* 1988, *200* (2), 351-65.

69. Goff, J. P.; Horst, R. L., Physiological changes at parturition and their relationship to metabolic disorders. *Journal of Dairy Science* 1997, *80* (7), 1260-1268.

70. Yamamoto, M.; Nakagawa-Ueta, H.; Katoh, N.; Oikawa, S., Decreased concentration of serum apolipoprotein C-III in cows with fatty liver, ketosis, left displacement of the abomasum, milk fever and retained placenta. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science* 2001, *63* (3), 227-31.

71. DeGaris, P. J.; Lean, I. J., Milk fever in dairy cows: a review of pathophysiology and control principles. *Vet J* 2008, *176* (1), 58-69.

72. Coppock, C. E., Displaced abomasum in dairy cattle: etiological factors. *J Dairy Sci* 1974, *57* (8), 926-33.

73. Baird, G. D., Primary ketosis in the high-producing dairy cow: clinical and subclinical disorders, treatment, prevention, and outlook. *J Dairy Sci* 1982, *65* (1), 1-10.

74. Charlton, M., Nonalcoholic fatty liver disease: a review of current understanding and future impact. *Clin Gastroenterol Hepatol* 2004, *2* (12), 1048-58.

75. Grummer, R. R., Etiology of lipid-related metabolic disorders in periparturient dairy cows. *J Dairy Sci* 1993, *76* (12), 3882-96.

76. Jorjong, S.; van Knegsel, A. T.; Verwaeren, J.; Lahoz, M. V.; Bruckmaier, R. M.; De Baets, B.; Kemp, B.; Fievez, V., Milk fatty acids as possible biomarkers to early diagnose elevated concentrations of blood plasma nonesterified fatty acids in dairy cows. *J Dairy Sci* 2014, *97* (11), 7054-64.

77. Holtenius, P.; Holtenius, K., New aspects of ketone bodies in energy metabolism of dairy cows: a review. *Zentralblatt fur Veterinarmedizin. Reihe A* 1996, *43* (10), 579-87.

78. Vazquez-Anon, M.; Bertics, S.; Luck, M.; Grummer, R. R.; Pinheiro, J., Peripartum liver triglyceride and plasma metabolites in dairy cows. *J Dairy Sci* 1994, *77* (6), 1521-8.

79. Andersen, J. B.; Madsen, T. G.; Larsen, T.; Ingvartsen, K. L.; Nielsen, M. O., The effects of dry period versus continuous lactation on metabolic status and performance in periparturient cows. *Journal of Dairy Science* 2005, *88* (10), 3530-41.

80. Bobe, G.; Young, J. W.; Beitz, D. C., Invited review: pathology, etiology, prevention, and treatment of fatty liver in dairy cows. *Journal of Dairy Science* 2004, *87* (10), 3105-24.

81. Bligh, E. G.; Dyer, W. J., A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* 1959, *37* (8), 911-7.

82. (a) Gorban, A. M.; Izzeldin, O. M., Fatty acids and lipids of camel milk and colostrum. *International Journal of Food Sciences and Nutrition* 2001, *52* (3), 283-7; (b) Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R.; Shimizu, T.; Spener, F.; van

Meer, G.; Wakelam, M. J.; Dennis, E. A., Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* 2009, *50 Suppl*, S9-14.

83. Shao, Y.; Wall, E. H.; McFadden, T. B.; Misra, Y.; Qian, X.; Blauwiekel, R.; Kerr, D.; Zhao, F. Q., Lactogenic hormones stimulate expression of lipogenic genes but not glucose transporters in bovine mammary gland. *Domestic Animal Endocrinology* 2013, *44* (2), 57-69.

84. Godden, S., Colostrum management for dairy calves. *The Veterinary clinics of North America. Food Animal Practice* 2008, *24* (1), 19-39.

85. Mansson, H. L., Fatty acids in bovine milk fat. *Food & Nutrition Research* 2008, *52*.

86. Jorjong, S.; van Knegsel, A. T.; Verwaeren, J.; Bruckmaier, R. M.; De Baets, B.; Kemp, B.; Fievez, V., Milk fatty acids as possible biomarkers to diagnose hyperketonemia in early lactation. *J Dairy Sci* 2015, *98(8), 5211-5221*

87. Contarini, G.; Povolo, M.; Pelizzola, V.; Monti, L.; Bruni, A.; Passolungo, L.; Abeni, F.; Degano, L., Bovine colostrum: changes in lipid constituents in the first 5 days after parturition. *J Dairy Sci* 2014, *97* (8), 5065-72.