Master's Thesis

Metabolomics of Lyme Patients Using NMR spectroscopy and Biomarker Analysis

Elanthiraiyan Saravanan

University of Jyväskylä

Faculty of Mathematics and science

Department of Physics

Nanoscience Centre

17/10/2018

PREFACE

The work presented in this thesis was carried out in collaboration with Prof. Perttu Permi from the Department of Biological and Environmental science and Dr. Leona Gilbert, Dr. Nitipon Puttaraksa from the Division of Cell and Molecular Biology at the University of Jyväskylä. The objective of this thesis is to present the metabolic changes in Lyme disease patients and to check the utility of the metabolites as biomarkers for Lyme disease.

I would like to thank my parents and my entire family for their support and encouragement. I would also like to thank my country India, pleasant city Jyväskylä and Finland for providing me a platform to peruse my dreams.

My sincere gratitude to Dr. Leona Gilbert, Prof. Perttu Permi and Dr. Nitipon Puttaraksa for giving me an opportunity to do this multi-disciplinary project and for guiding me thorough out this project. A special thanks to Prof. Hannu Häkkinen for agreeing to be my supervisor in the nick of time.

I am greatly indebted to Prof. Janne Ihalainen, Dr. Jussi Toppari and Dr. Andreas Johansson for their valuable advises and first-rate lectures.

Deepest thanks to my colleagues Kati Karvonen, Arttur Pudas, Kunal Garg, Laura Maahinen and Nykky Jonna for helping me overcome various hurdles that I have faced as a new man in this foreign land.

I cannot proceed further without mentioning solicitude and endearment of my friends Abi, Vaishu, Praveen, Afra, Gomathy, Sailee, Dhanik, Bhavani, Karthick, Kati, Arttur, Chath Saro, Hamalesh and Abishek! I would like to blame them all for my joyful life in this beautiful country.

- Elan

University of Jyväskylä

Faculty of Mathematics and science

Author:	Elanthiraiyan Saravanan
Title of thesis:	Metabolomics of Lyme Patients Using NMR spectroscopy and Biomarker Analysis
Date:	17/10/2018
Department:	Department of Physics
Supervisor(s):	Leona Gilbert (PhD, Docent), Perttu Permi (Professor), Nitipon Puttaraksa (Postdoctoral researcher) and Hannu Häkkinen (Professor).

Abstract:

Lyme disease (LD) is one of the dreadful tick-borne diseases. It is commonly caused by pathogen *Borrelia burgdorferi* and three different species of *Borrelia*. Number of LD cases being reported each year is increasing drastically in various parts of the world. Although history of LD dates back to late 1970s, LD diagnosis and treatment is yet a challenging task. The existing, Centers for Disease Control and Prevention (CDC) and Food and Drug Administration (FDA) approved serological diagnostic methods of LD diagnosis show low assay sensitivity. The sensitivity of the serological based LD diagnostic techniques is less than 69 % for early or acute LD diagnosis. Therefore, there is an immense need for more sensitive diagnostic method.

LD pathogens interact with various types of cells within the body such as neurons, muscle fibers, etc. The pathogens primarily disturbs the cellular pathways which causes fluctuations in the levels of metabolites, including the levels of amino acids. We used nuclear magnetic resonance spectroscopy (NMR) which is one of the promising techniques for omics studies, to detect and quantify metabolites in LD patient sera. We observed a strict elevation in the level of lactate and decrease in the level of glucose, choline and alanine in majority of acute and late LD patients sera used in this study. Statistical modeling and results suggest that based on the identified features acute and late LD patients can be distinguished from the healthy donors with sensitivity 80 % and specificity 90 - 100 %. Notably, 89 % of the CDC negative lyme samples were correctly classified based on the identified features. Our results suggest that metabolic profiling of LD patients could be a better alternate for both acute and late LD diagnosis with improved sensitivity, compared to expensive serological techniques.

Key words: Metabolomics, Biomarkers, Lyme disease, NMR spectroscopy

TABLE OF CONTENTS

ABBREVIATIONS

AUC	Area under curve
BMRB	Biological magnetic resonance bank
CDC	Centers for Disease Control and Prevention
CPMG	Carr-Purcell-Meiboom-Gill spin echo sequence
ELISA	Enzyme-linked immunosorbent assay
EUCALB	European Concerted Action On Lyme Borreliosis
FDA	Food and Drug Administration
FID	Free induction decay
HMDB	Human metabolome database
IDSA	Infectious Diseases Society of America
IgG	Immunoglobulin G
IgM	Immunoglobulin M
LDL	Low density lipoprotein
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
OspC	Outer surface protein C
PLS-DA	Partial least square-discriminate analysis
ROC	Receiver operating characteristic curve
SVM	Support vector machine
TSP	Trimethyl-silypropanoic acid
TTTA	Two tier testing algorithm
VLDL	Very low density lipoprotein
WB	Western blot

1. INTRODUCTION

1.1 Lyme disease

In the year 1977, Dr. Alan Steree and colleagues first recorded Lyme disease (LD) as an infectious illness [1]. The major causative organisms of LD are *Borellia burgdorferi, Borellia garinii, Borellia afzelli* [1] and recently discovered *Borellia myonii* [2]. *Borellia* is transmitted to humans by ticks from animals such as deer, birds, rats, etc. *Borellia* is also called as spirochete because of its spiral morphology. The spirochetes measure 10-30 μm in length and 0.18-0.25 μm in width [3]. The parasite is known to be pleomorphic as it has an ability to change its morphology [4]. These parasites adapt to host's body temperature and ambient conditions like many other bacteria and viruses [4]. Wherever, humans are exposed to Lyme vectors the chances of acquiring LD is higher. Particularly, people who live in close proximity to forest areas with high deer population are more likely affected by LD. There are three different stages of LD namely acute LD, early disseminated LD and late LD [5]. Acute LD or early localized LD is the stage one LD. If LD is not diagnosed and treated at the early stage or initial phase of infection, then it leads the patient into severe phases. The symptoms of LD in patients includes headache, fatigue, fever and chills, muscle aches, joint aches, swollen lymph nodes, neck stiffness, nausea, vomiting, erythema migrans (EM) etc., [2,6]. EM rashes are commonly called as 'bull's-eye' rash [2,6]. EM rash can be observed in 80 % of acute LD or early localized LD patients within a period of 7 to 14 days after the tick bite [5]. The symptoms that occur in the advanced stages include facial nerve weakness and paralysis (Bell's palsy), intermittent-joint pain, inflammation in the brain, heart problems and spinal cord, etc., [7–10]. Van Dam and colleagues, suggested that this diversity in symptoms and severity of the infection depends on the species of *Borellia* [11]. LD patients infected by *Borellia burgdorferi* commonly suffer from joint pain and arthritis associated illness, LD patients infected by *Borellia garinii* commonly suffer from neurological symptoms and *Borellia afzelli* infected LD patients shown skin disorders [1,11]. Studies with mice models propose that severity in critically ill LD patients is due to auto-immune condition where the host immune system functions abnormally and begin to damage the native cells of host itself [1].

1.1.1 Global impact

In USA, Europe, South Africa, Brazil and in parts of Asia, LD is one of the dreadful vector borne diseases. In US states and territories 402,502 LD cases were reported in the period 2014 - 2016 [12] and according to scientists from Centers for Disease Control and Prevention (CDC) more than 300,000 LD cases are reported each year in US alone [13]. LD endemic countries across the world are shown in Figure 1. In different parts of the world, different species of *Borellia* remain to be a common causative of LD. *Borellia burgdorferi* is common in America whereas *B.garinii* and *B.afzelli* are common in Europe and Asia [14]. Similarly, corresponding vectors or carriers of *Borellia* differ from one region to another [1]. For example *Ixodes*

scapularis and *Ixodes pacificus* are the major vector in USA, and *Ixodes ricinus* is the major vector in Europe, *Ixodes persulcatus and Ixodes ricinus* are the major vector in Asia [1].

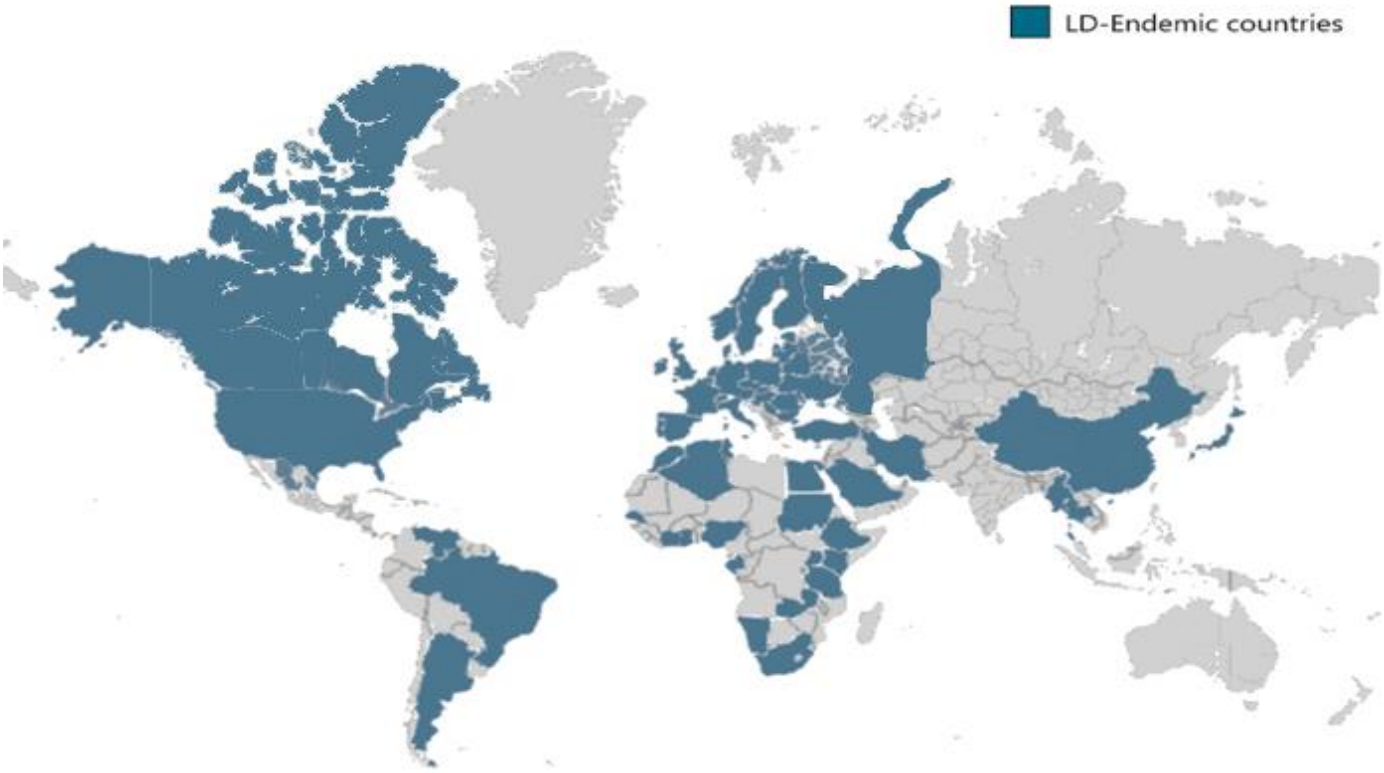


Figure 1: Lyme disease endemic countries. Parts of the world where LD is endemic with moderate and high occurrence are shaded blue. [Adapted from https://bewareofthebugs.com/diseases/lyme-disease/]. In countries such as Pakistan, Thailand, parts of Asian-Russia etc.,no documented occurrence found [http://www.cvbd.org/en/occurrence-maps/world-map/].

1.1.2 Lyme diagnosis and treatment

Lyme disease patients are treated with antibiotics such as amoxicillin, doxycycline and cefuroxime axetil in the initial phase of infection for a period of two to three weeks [13] and treated with cetriaxone, penicillin, cefuroxime axetil, amoxicillin or doxycycline in disseminated stage for 10 to 30 days [13]. Modes of drug administration are oral and intravenous. Though these antibiotic treatments are recommended by Infectious Diseases Society of America (IDSA) and European Concerted Action on Lyme Borreliosis (EUCALB, www.eucalb.com) [15] the duration of the treatment is still debated and it is a controversial part. There are numbers of therapeutic agents for LD, but the real challenge lies in diagnosis. Particularly early diagnosis is very challenging. LD testing methods currently available are expensive. According to 2014 statistics for LD in United States of America 2.4 million specimens tested with an approximate cost of 492 million US dollars which is roughly 4,800 US dollars per sample [5]. Serological based diagnosis is the widely practiced method for LD diagnosis [16]. Two-tier testing Algorithm (TTTA) is one such serological based technique in which two set of assays performed consecutively [5]. The decision tree of TTTA is presented in Figure 2. TTTA involves serological assays such as Enzyme Immunoassay (EIA) and Western Blot (WB). Individuals suffering from LD symptoms for a period $\leq$ 30 days are tested for IgM and IgG and those with prolonged

symptoms > 30 days are tested only for IgG [5]. TTTA is Food and Drug Administration (FDA) approved diagnostic method.

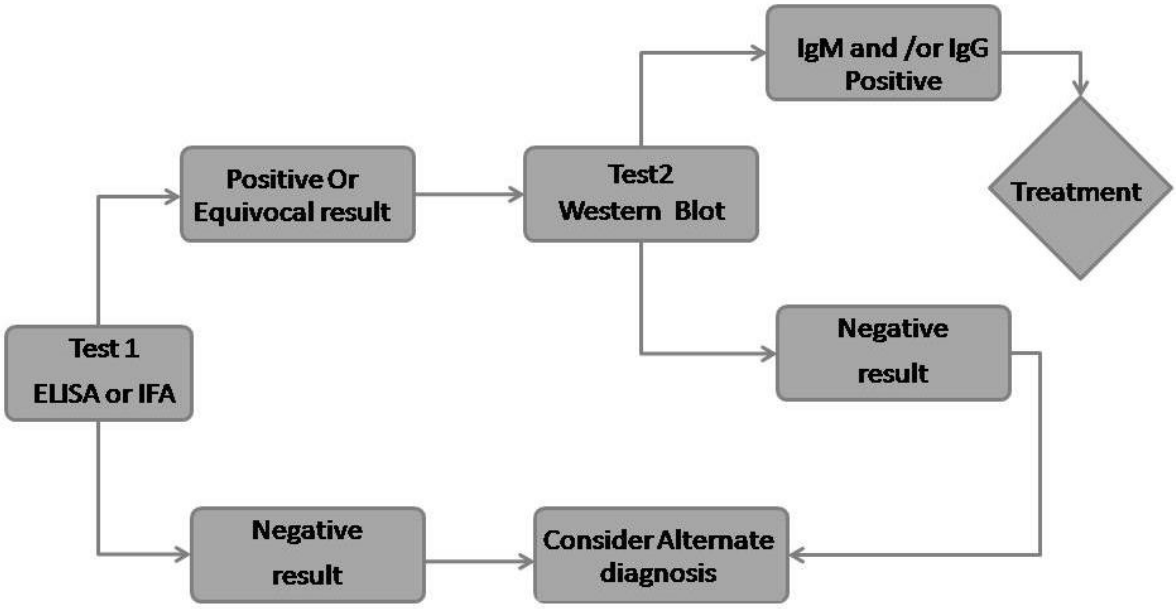


Figure 2: Schematic of Two Tier Testing Algorithm (TTTA). The schematic shows the two tier testing decision tree, patients with clear symptoms of LD such as EM rash are first subjected to Enzyme Immunoassay (EIA) or Immunofluorescence Assay (IFA). If the test results are negative then the patients are advised to consider alternate diagnosis and if the results are positive or equivocal then based on the count of days the patient had LD symptoms IgM or IgG western blot (WB) is done. If the results of WB are positive then the patients are treated with antibiotics.
[Adapted from https://www.cdc.gov/lyme/healthcare/clinician_twotier.html]

Various diagnostic kits based on TTTA are commercially available [16]. Although the result heterogeneity exists among the kits, they are used in clinical diagnosis [5]. Recently, novel approaches have been developed to detect 10 unique *Borrellia burgdorferi* antigens including previously reported OspC and FlaB. If two out of 10 antigens are present in the sample then the patient is considered to be LD positive [17]. Molecular technique namely iPCR (immuno-Polymerase chain reaction) was developed and thought to be used for LD diagnosis but it is not approved by Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) [17]. In 2015, Molins and his colleagues proposed 44 lipids and lipophilic as bio signatures of early LD [18]. Based on these features they tried to distinguish LD patients from healthy donors and negative Lyme samples and made a comparison with serological assays [18].

1. 2 Metabolism and metabolomics

By standard definition, metabolism is the process of building up or breaking down of compounds in living organisms. Every single organism from unicellular to multi-cellular performs numerous biochemical reactions in order to survive. Organic molecules such as amino acids, nucleotides, small proteins (usually less than < 300 kDa), fatty acids, triglycerides, signaling molecules involved in this process of metabolism are called metabolites. They can be found in all types of body fluids such as urine, semen, saliva, sweat, etc., [19]. The

metabolic profile of individuals differs from one another based on the factors such as living environment, food habits, genetic background and pathological stimuli [20]. The metabolic profile of a person reflects his or her physiological condition and so the studies about metabolites have become common in biology and medicine [20,21]. Metabolomics and metabonomics are the general terms used interchangeably to refer the study about metabolic changes in an organism [20]. The most important tools commonly used in metabolomics are liquid chromatography-mass spectrometry (LC-MS), gas chromatography-mass spectrometry (GC-MS) and nuclear magnetic resonance (NMR) spectroscopy [22,23]. NMR-based metabolic profiling is a sample non-destructive technique [19,24], but compared to LC-MS, NMR spectroscopy technique is less sensitive. Yet, NMR technique is suitable for biomarker identification because of its remarkable ability to probe novel metabolites [23]. Beckonert's article that was published in 2007 is the most cited article for metabolomics studies [23].

1.2.1 Biomarkers

Diagnosis based on metabolites and molecular features dates back to 1500 BC. During this period the physicians from south-east Asia and Egypt used simple technique to diagnose diabetes which is considered to first clinical test for diabetes [25]. In modern science a biomarker is used for disease diagnosis, to study the effect of therapeutic agents on patients and disease prognosis [23]. A biomarker is an indicator of disease, physiological and biological condition of the living organism [26]. It is usually a significant feature such as chemicals, metabolites and proteins which are present in the body fluids of the diseased that can be easily acquired and tested. Gene mutations, RNA transcripts and cell counts can also be biomarkers [23]. Ability of a biomarker is typically evaluated by its sensitivity and specificity [23], which are given by following relation (1.1) and (1.2),

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of positives}} \tag{1.1}$$

$$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of negatives}} \tag{1.2}$$

A cost-effective way of checking the utility of a biomarker is Receiver Operating Characteristic (ROC) curve analysis. In ROC curve specificity of a biomarker (in x-axis) is plotted against sensitivity (in y-axis) and from the area under curve (AUC) value of the generated curve, the reliability of biomarker in clinical diagnosis is graded. If AUC value of the ROC curve is 1 then the biomarker is considered to be excellent and if AUC is 0.5 or less, then the biomarker is considered to be a random classifier. Such metabolite cannot be used for any practical purpose. A scale for assessment of biomarker based precisely on its AUC is given in Table 1 [27].

Table 1. Scale for assessment of biomarker

AUC	Biomarker utility
0.9 - 1.0	Excellent
0.8 - 0.9	Good
0.7 - 0.8	Fair
0.6 - 0.7	Poor
0.5 - 0.6	Fail

Modern approach of identifying and quantification of metabolites for diagnostic purposes is currently used in more than 200 clinical tests. For example, clinical tests such as detecting urinary nitrate in diagnosing patients with bacterial infections, bilirubin to monitor the liver function and many other are based on quantification or semi quantification of metabolites and biomarkers [23]. Some of the considerable merits of metabolite biomarkers compared to protein biomarkers are non-invasive sample collection, speed and quantitative accuracy [27]. A list of metabolites, their respective functions and role as biomarker is presented in Table 2.

Table 2. List of metabolites, functions and their role as biomarker [[28]**, HMDB]**

Metabolites	Function	Disease associated with metabolites and their role as biomarker
α glucose	Primary source of energy	-
β glucose	Primary source of energy	-
Lactate	Several bio-chemical process and play major role in brain metabolism.	Early stage biomarker for cancer [29]
Choline	Precursor of acetylcholine and play essential role in lipid metabolism	Biomarker for acute coronary syndrome [30] and cancer diagnosis [31]
Creatine	Helps in storage and transmission of energy from high energy phosphate compounds.	Biomarker for mitochondrial disease [32,33], hepatocellular necrosis and renal disease [32]
Glutamine	Most abundant but non-essential amino acid. Studies suggest that they are essential for immune system, inter-organ nitrogen transport, precursor for neurotransmitter. [34]	Cancer diagnosis [35–37]

Glutamate	Neural communication and considered to play role in cognitive functions of the brain.	Indicator of chronic kidney disease [35]
Glycine	Associated with synthesis of phospholipids, collagen and in energy release.	Cancer diagnosis [38,39]
Leucine	One of the three Branched chain amino acids (BCAA) involved in muscle stress and energy metabolism. It stimulates insulin release.	Cancer diagnosis [40–42]
Valine	One of the three branched chain amino acids (BCAA) involved in muscle, stress and energy metabolism.	Cancer diagnosis and its deficiency is the sign of neurological defects [38,41,43]
Alanine	Alanine is one of precursor and also a regulator in glucose metabolism. It is also an essential amino acid for lympocyte reproduction.	Cancer diagnosis [35,39]
LDL	Carrier of cholesterol, triglycerides and fatty acids.	Cardiovascular disease [44] and atherosclerotic kidney disease [45]
VLDL	Carrier of cholesterol, triglycerides and fatty acids.	Bio marker for nonalcoholic steatohepatitis [46]
Citric acid	Release of stored energy through TCA cycle or citric acid cycle.	The evaluation of plasma citric acid is scarcely used in the diagnosis of human diseases.

1.2.2 Software packages for metabolomics

SIMCA-P, SAS, metaP-server, MeltDB, MetaboAnalyst are some of the software available for metabolomics data analysis and interpretation. Among those, MetaboAnalyst is commonly used by academic researchers in metabolomics community [47]. MetaboAnalyst is a web based freeware developed mainly for supporting quantitative metabolomics. It is a user friendly software, and it is one of the preferred software for univariate, multivariate and pathway analyses [47]. The process of analysis using MetaboAnalyst involves five following

steps (1) data formatting (2) data processing (3) data normalization (4) signature feature identification (signature feature identification and building a model in case of biomarker analysis) (5) Choosing the suitable algorithm [47]. In order to reduce the over fitting problems in ROC curve analysis described in section 1.2.1, MetaboAnalyst tool allows the user to build a model, based on own judgments of the users.

1.3 Nuclear Magnetic Resonance spectroscopy

1.3.1 General information

Nuclear magnetic resonance (NMR) spectroscopy is an advanced technique that works on the principles of classical physics and quantum mechanics. In the year 1946, NMR spectroscopy was first developed by research groups at Stanford and M.I.T. in USA and the use of the proton NMR for quantitative analysis was first reported by Jung nickel and Forbes few years later in 1963 [48]. Since then this technique has gained its importance in various fields of science especially, medical science. NMR spectroscopy is similar to the well-known magnetic resonance imaging (MRI) technique. The fundamental principle behind this staggering NMR technique is detection of nuclear spin states of the atoms in a molecule placed in a strong magnetic field, with the help of radio frequency. A simplified pictorial representation of NMR set up is shown in Figure 3.

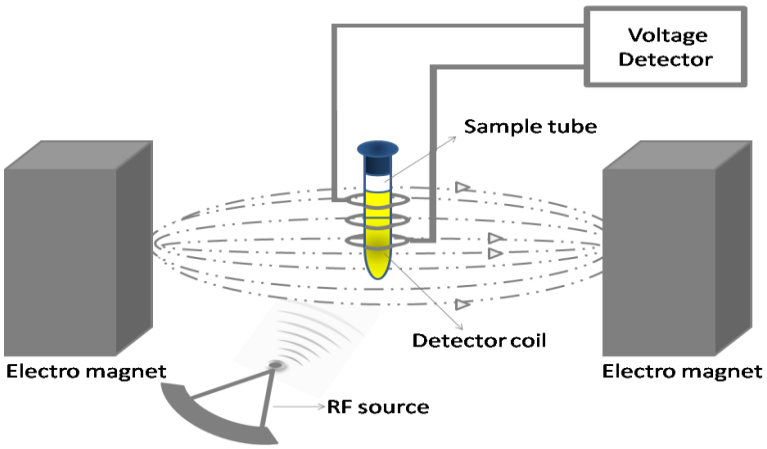


Figure 3: Simplified representation of NMR spectroscopy setup. NMR spectroscopy setup comprises strong electromagnets, magnetic field strength ranging from 2.35 Tesla to 21.2 Tesla, radio frequency (RF) transmitter, detector coil positioned in between the magnets, rotating sample holder, etc., as depicted in this figure. [Adapted from https:// www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/nmr/nmr1.htm].

This technique is commonly used to accurately determine the complex structure of organic molecules and to study complex protein structures. Physical phenomena involved in NMR technique are usually found to be complicated which actually makes it more interesting. Despite its complexity, it is one of the effective techniques for structural biology and metabolomics [22,49,50]. The cost of the NMR spectrometer depends primarily on the strength of the magnet. For instance NMR spectrometer 2.35 Tesla magnet is less expensive compared to the one with 21.2 Tesla. The strength of the magnet is the major factor that influences the sensitivity of the technique. In order to use NMR spectroscopy as an analytical tool for a study, it is very important to understand the physical phenomena involved in it which includes nuclear spins, magnetic

resonance, etc. In experimental point of view, basic understanding about ideas, such as pulsed NMR and mathematical tools namely the Fourier transform [51], are encouraged.

1.3.2 Magnetic resonance

Atomic nuclei are charged particles and so a spinning nucleus can be compared to spinning bar magnet for easy understanding. When a spinning nucleus is placed in a strong magnetic field, upon irradiation with suitable frequency the nucleus gains energy and switches to higher energy state. This process of switching or flipping from one energy state to another is termed as nuclear magnetic resonance [48]. In a time interval T the nucleus relaxes to its low energy state or stable state (usually in many microseconds). Faraday's law of induction clearly states that a change in the magnetic field linked to a coil induces an electron motive force (emf) in it. Thus, flipping of the spin systems/nucleus placed in the strong magnetic field induces voltage in the surrounding detector coil. When the excitation using radio frequency is turned off, the induced voltage decays with respect to time which is known as free induction decay (FID). The detected voltage is amplified and processed. The FID signal is a very complex time domain data, it is Fourier transformed to a comparatively simple data/ spectra and the process is depicted in Figure 4. The energy required to switch the orientation of the nucleus from lower energy state to higher energy state depends on the strength of the magnetic field B_0. It is given by equation 1.3 [48],

$$\Delta E = \frac{\gamma h B_0}{2\pi}, \tag{1.3}$$

Where h is the Planck's constant (6.63×10^{-34} J.s), γ is the gyromagnetic ratio which is a constant number for specific nuclei. The gyromagnetic ratio γ is given by equation 1.4,

$$\gamma = \frac{2\pi\mu}{hI}, \tag{1.4}$$

where μ is the magnetic moment and I is the spin number, energy required to flip the nucleus (according to Niels Bohr condition, proposed in the year 1913) is given by equation 1.5,

$$\Delta E = h\nu. \tag{1.5}$$

From (1.3) and (1.5), transition frequency ν that is required to initiate the nuclear transition can be given as follows

$$\nu = \frac{\gamma B_0}{2\pi}. \tag{1.6}$$

The equation (1.6) is commonly called as Larmor equation. The nuclear precession rate is given by equation 1.7,

$$\omega_0 = 2\pi v, \tag{1.7}$$

where v is the transition frequency. ω_0 is also termed as Larmor frequency or angular Larmor frequency.

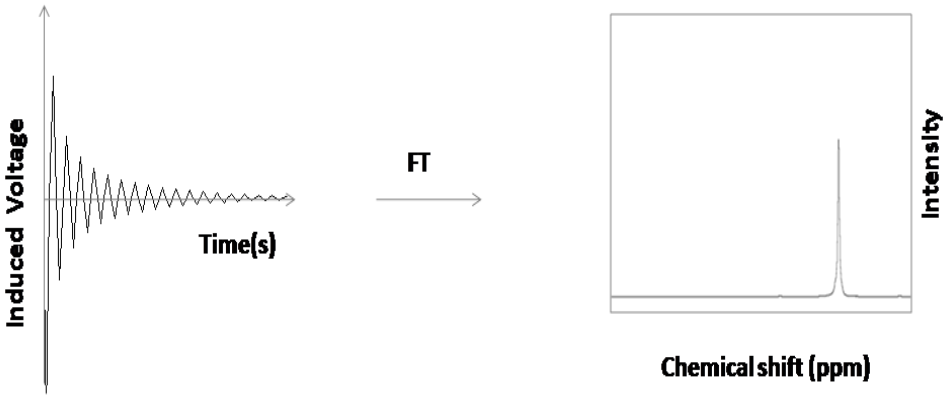


Figure 4: An example of the Fourier transformation (FT) of free induction decay. FID-time domain data on left and the Fourier transformed FID on right.

1.3.3 Nuclear spin and Zeeman Effect

Nuclear spin (*I*) is represented by a value which can be an integer or a half integer. The nuclei of a hydrogen atom ^{1}H (Protium), naturally abundant form of hydrogen, have only one proton and it has spin ½. Nucleus with odd number of nucleons such as ^{1}H, ^{13}C isotope of carbon etc., are capable of spinning. Nucleus with odd number of protons, when placed in a strong magnetic field can exist in different energy states or spin states due to Zeeman effect which is the splitting of energy levels in the presence of a non-fluctuating magnetic field as depicted in Figure 5. The number of such different states, a nucleus can exist is given by, 2*I*+1 rule. For example, *I* of the ^{1}H is ½ and it can exist in two different energy states as per the 2*I*+1 rule, which are -½ and +½. The nuclei with higher energy précis against the direction of the magnetic field whereas the nuclei with lower energy précis along the direction of the polarizing magnetic field**.** The transition frequency v of ^{1}H in a 14.1 T magnetic field is 600 MHz. As the strength of the magnetic field increases, the transition frequency required to flip the nuclei from lower energy state to higher energy state also increases, as presented in figure 5.

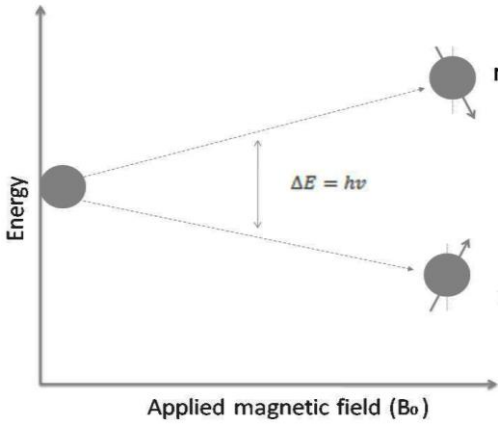


Figure 5: Zeeman Effect. A splitting of energy levels in the presence of non-fluctuating magnetic field is depectited in this figure. The nucleus in lower energy state spins along the direction of the magnetic field (on a precision axis) and the nucleus in higher energy state spins against the direction of the magnetic field (on a precision axis). [Adapted from Thomas 1998].

1.3.4 Nuclear spin population distribution

In practice, the sample to be analyzed has the number of nuclei approaching the Avogadro number and except for a small population of the nucleus, the magnetization of the other nuclei cancels out each other. Therefore the detected output is actually the net effect of a small population, which is then amplified and processed to obtain necessary information. In the absence of magnetic field, all nucleus précis on different axis and when the magnetic field is applied, the nucleus précis on parallel axes. The population distribution of the nucleus obeys Boltzmann distribution [48], refer to Figure 6. The distribution of the nucleus population in an undisturbed condition is given by

$$\frac{N_{upper}}{N_{lower}} = e^{-h\nu/kT}, \tag{1.8}$$

where N_{upper} is the population of the nuclei in higher energy state and N_{lower} is the population of the nuclei in lower energy state, k is the Boltzmann constant, and T is the absolute temperature in Kelvin [48]. From equation (1.8), for protons placed in magnetic field of 18.8 T (transition frequency is 800 MHz) at thermal equilibrium at 298 Kelvin (room temperature) the ratio of the population of nuclei is 0.999872 [48].

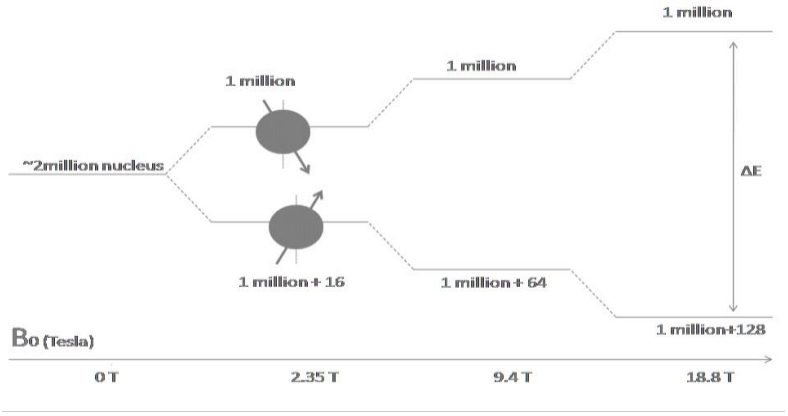


Figure 6: Magnetic field strength dependence of nuclear energy levels. This figure shows the increase in detectable nuclear population, with respect to magnetic field strength. The energy required to flip the nucleus ΔE also increase as a function of magnetic field strength [Adapted from Thomas1998].

1.3.5 Chemical shift

The resonance frequency of the spin system depends on the local chemical environment of nucleus as the electron shielding of the nucleus is influenced by surrounding atoms. This effect of local environment on spin system is reflected in detected signals which helps us to know the nearby nuclei and the chemical groups. These information can be acquired directly when the time domain data is Fourier transformed to frequency domain data. The frequency domain data is expressed in Hz, which is usually a more of complex number, making a resonance peak hard to be recognized and compared. In addition to this, the resonance frequency of a compound is dependent on the applied magnetic field, see equations (1.6) and (1.7) i.e. the resonance frequency of a nucleus will vary when measured in a different magnetic strength, making it even more hard to recognize and compare data measured with different NMR spectrometers. For these reasons, a magnetic strength independent factor, the chemical shift, usually represented as δ, is used. It is more convenient to compare the peak positions with their respective chemical shifts expressed in 0 to 20 ppm and it given by equation 1.9,

$$\delta = \frac{v_{sample} - v_{ref}}{v_{spectrometer}} \times 10^6, \tag{1.9}$$

where v_{ref} is resonance frequency of a reference standard, v_{sample} is resonance frequency (in Hz) of the analyte and $v_{spectrometer}$ is the spectrometer frequency. The principles and parameters discussed about NMR so far in this work, is only a pinch of whole concepts associated with NMR spectroscopy. More detailed information can be found in the text of Thomas L.James [48,52].

1.3.6 Signal to noise ratio

Signal to noise ratio (S/N) in NMR spectra can be improved by signal averaging as a factor of √n where n is the number of signal averaged, according to *ceteris paribus* approach [48,53]. It can be done with ease by exciting the spin systems with RF at different time periods and detecting the FID consecutively for n number of times. This process is commonly addressed as pulsing which is one of the important area of NMR spectroscopy. NMR spectroscopists can design the pulse sequence based on the sample type and other requirements. These pulse sequences can also act as filters when designed accordingly. Some pulse sequences are capable of attenuating the signals from the nuclei of unwanted entities such as large proteins, by altering the spin orientation and time taken for spin lattice relaxation of those spin systems. As stated in section 1.3.2, when the effect of RF pulse is removed, the spin system relaxes to thermal equilibrium. The time taken by the spin system for longitudinal relaxation represented as T_1 and for transverse relaxation represented T_2, depends on type of the nucleus and its local magnetic environment.

1.3.7 1D proton NMR data analysis and statistical methods

In order to lower the complexity in NMR data processing and compound identification various tools and software such as Chenomx, Mesternova (Mnova), AMIX, TopSpin were developed and are commercially available. With the help of these software packages, peak picking and identification of the metabolites can be done automatically or semi-automatically, as some of these software are linked with databases such as HMDB (the Human Metabolome Data Base), BMRB (Biological Magnetic Resonance data Bank) etc. Since no single review article can be found regarding the utility and performance of the software, no software can be considered superior over another. However, from the opinions of the users it can inferred that each of the software has its own merits and demerits for example Mnova and Top spin seems to be the most commonly used software packages for 1D NMR data processing (Phase correction, baseline correction), whereas Chenomx is usually used to perform quantitative analysis on 1D NMR spectra. The processed information is then studied using statistical tools. Generally, the role of the statistical approaches such as Principle component analysis (PCA), Partial Least Square-Discriminate Analysis (PLS-DA), etc., is to reduce the complexity of a data making it more understandable inorder to study the relationship between the samples. Partial Least Square-Discriminate Analysis (PLS-DA) is a supervised classification method [27]. Both PCA and PLS-DA identify the variables or principle components that can be used to classify the data sets, by identifying the direction along which the variance is high for any given data set. PLS-DA is preferred over PCA in quantitative, semi or relative quantitative analysis. The detailed explanation of this method is discussed elsewhere [54,55].

1.4 Serum separation / preparation

Serum is the body fluid separated from blood which does not contains fibrinogen or other clotting factors. One of the major differences between the blood serum separation and blood plasma preparation procedure is the

usage of anticoagulant in the process of plasma preparation [www.proimmune.com]. From 8 ml of blood, 3.2 ml serum can be acquired. The step by step process is depicted in the Figure 7. The tube containing whole blood is incubated in an upright position at room temperature for 45 minutes. Then, centrifuged at 3000 rpm for 10 minutes. Finally, the supernatant (serum) is carefully collected at room temperature in a separate tube [www.proimmune.com]. Serum aliquots are stored separately at – 80 °C in cryo-vials and labeled accordingly.

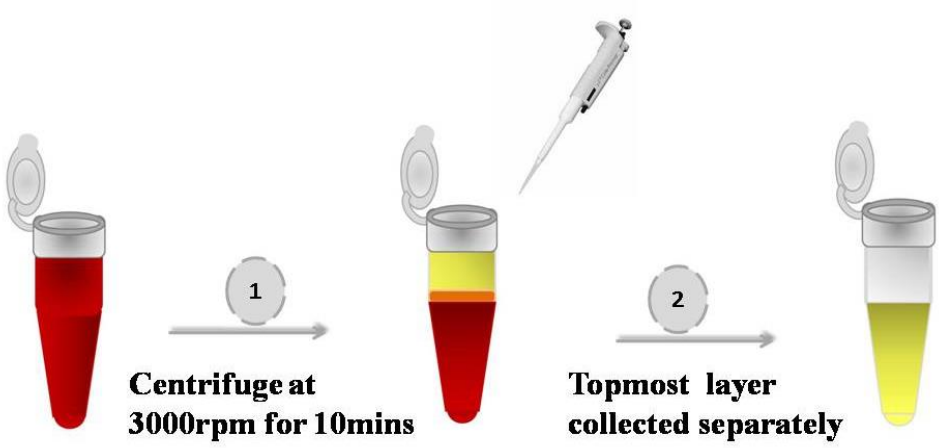


Figure 7: Serum separation method. The two step process of serum preparation is depicted in this figure. The orange layer in between the top layer and bottom most layer is their interphase that need not be disturbed while serum collection. [www.proimmune.com].

2. AIM OF THE STUDY

Lyme pathogens interfere in cellular functions which results in fluctuations in the level of metabolites. These metabolic changes are useful information about LD patients and some of the metabolites can possibly be the biomarkers of LD. In this master thesis project, specific aims were

a) To study the metabolic changes in LD patients using NMR spectroscopy.

b) To perform biomarker analysis.

3. MATERIAL AND METHODS

Human sera sample collection was approved by the Federal Institute for Drugs and Medical Devices, Germany (project no. 95.10-5661-7066); and Western Institutional Review Board, United States of America (USA) (USMA201441, WIRB® protocol #20141439). Necessary permission were acquired from all patients enrolled in this study. The samples were categorized into acute LD, late LD, negative Lyme based on the CDC criterion. Ten samples from healthy donors and ten samples from each of the other three categories, in total 40 samples were used in this study [Refer appendix I for sample description].

3.1 Sample preparation

All serum samples were prepared by standard serum separation protocol prior to this study for other characterization purposes. NMR tubes of 5 mm outer diameter and 7 inches in length were purchased from Wilmad, model 541-PP-7, and used for all NMR measurements. TSP (3-Trimethylsilyl propionic-2, 2, 3, 3-d4 acid sodium salt), used for spectral referencing was purchased from Acros Organics. Deuterium oxide, deuteration degree min. 99.9 % for NMR spectroscopy was purchased from Merck and used for buffer preparation and to dissolve TSP. Sodium phosphate buffer (125 mM) was prepared to maintain a constant pH for all the samples throughout the course of experiments. The pH of the buffer was measured using the pH meter and adjusted to pH 6.0, by adding 1M potassium hydroxide drop wise. Few minutes prior to measurement the sera were thawed and mixed well before transfering them to the NMR tube. 100 µl of serum was suspended in 400 µl of 125 mM sodium phosphate buffer and 3 µl of 1 % TSP solution was added to the mixture. The resultant mixture was subjected to measurement using 800 MHz Bruker AVANCE III HD NMR spectrometer, equipped with cryogenically cooled ^{1}H, ^{13}C, ^{15}N triple-resonance probehead at Nanoscience Center, JYU.

3.2 NMR-data acquisition parameters

A special pulse program with T_2 filtering was employed to collect ^{1}H spectrum. The purpose of T_2 filtering was to attenuate the signal from high molecular weight compounds mainly proteins. The T_2 filter employes Carr-Purcell-Meiboom-Gill (CPMG) sequence that is a spin-echo sequence typically utilized for measuring T_2 relaxation times but can be applied to attenuate the signals from proteins. The sequence of CPMG is as follows –RD-90°-(t-180°-t)$_n$-ACQ, where RD is relaxation delay, 90° and 180° represents 90° RF, 180° RF pulses, t is the spin-echo delay, n represents the number of loops and ACQ stands for acquisition [19]. In addition, during the RD period a weak RF field at the frequency of water protons was applied. This actively suppress the water magnetization and improves the overall sensitivity as higher receiver gain can be used during signal detection. Some of the essential parameters employed are listed in the Table 3. Shimming was done for every sample to increase the accuracy, precision, and reproducibility.

Table 3. NMR Data acquisition parameters

Acquisition mode:	DQD
Time domain points (TD):	65000
Number of dummy scans :	4
Number of scans (ns):	64
Spectral range:	0-20 ppm
Receiver gain:	32
Sample temperature:	298 K
Acquisition time	2.04 seconds
FID resolution:	0.489064 Hz
Filter width:	4032000
D1	4.0 seconds

3.3 Data processing and analysis

^{1}H NMR spectra of the samples were phase corrected and calibrated manually using the TopSpin 3.2 software. A brief table of data processing and analysis is presented appendix I. Exponential weighting function with line broadening 0.3 Hz was used for all the spectra prior to Fourier transform. Baseline correction was done automatically using Polynomial of Degree ABSG in TopSpin 3.2. The integral area of the reference TSP peak at δ 0.0 was calibrated manually for all spectra. Calibration of integral area was done by comparing the TSP peak in one of the sample for which integral area was calibrated to 0.01 with that of the TSP peak in all other samples. Integral values of metabolites were obtained with respect to calibrated TSP peaks. The relative integral values of 15 metabolites of interest obtained carefully from all the spectra. Mean integral values of these individual metabolites were then calculated and histograms are prepared using Origin Pro to study the metabolic differences. The integral values of the metabolites were formatted and labeled as per the requirements mentioned on MetaboAnalyst website [www.metaboanalyst.com]. This online software MetaboAnalyst, was used to compare the samples and their categories. ROC curve based model evaluation (Tester) tool, available in MetaboAnalyst software was used to build and test biomarker models. PLS-DA algorithm was used in class prediction. As over fitting is very common in this type analysis, metabolites were not selected based on their AUC and fold change values that were automatically generated by the online freeware. Instead, they were selected based on the following four criterions 1) sum spectra; 2) role of the metabolite and its association with LD symptoms, given in Table 2; 3) metabolite that vary in majority of the patient population; 4) multivariate exploratory ROC analysis tool available in MetaboAnalyst [Refer appendix III]. The features were chosen to build model and biomarker analysis was performed.

4. RESULTS

Sample preparation is one of the crucial parts in metabolomic study because of the diversity in the composition of human sera and varying ionic concentrations. Test experiments were conducted in order to standardize the sample preparation protocol and for instrument optimization. Human serum purchased from Sigma Aldrich (Male, USA-origin, AB plasma, Lot# SLBN8825V) was used for the trial experiments. Since the sera samples of LD patients were limited and demanding, it was necessary to estimate the minimum quantity of serum that can be used in this study without compromising with the instrument sensitivity. The sensitivity of the NMR spectroscopy technique is different from the sensitivity discussed about biomarker analysis. Test experiments were conducted with four different ratios of serum and buffer as mentioned in Table 4, and results are shown in Figures 8 and 9.

Table 4. Buffers and serum sample ratio tested for sample volume optimization.

Sample name	Dilution factor	Total sample volume (500 µl)		Scan time in minutes	pH	Internal reference	Measurement temperature in K
		Buffer vol. in µl	Serum vol.in µl				
T1	1:2	250	250	1	6.0	-	298
T2	1:5	400	100	1	6.0	-	298
T3	1:10	450	50	1	6.0	-	298
T4	1:20	475	25	1	6.0	-	298

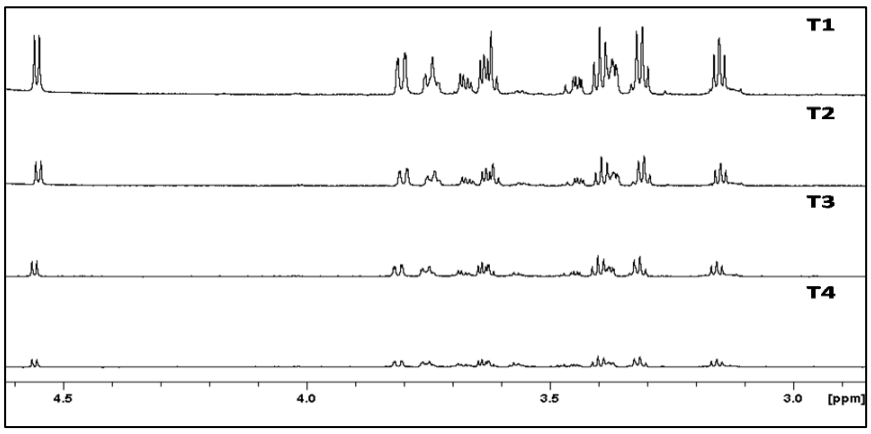


Figure 8: ^{1}H NMR spectra of human serum (δ 1-4.7). ^{1}H NMR spectra of the human serum in four different buffer to serum volume ratio is show in this figure. T1, T2, T3, and T4 spectra represents the four different dilutions of human serum. Each spectra is placed on top of each other for comparison. (Spectra was not calibrated since no standard or reference used in these trials).

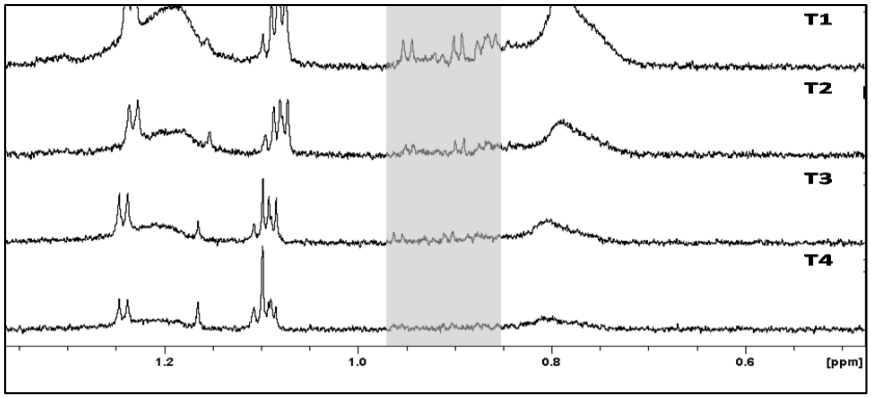


Figure 9: ^{1}H NMR spectra of human serum (δ 0-1.7). These spectra were acquired in ~ 1 minute quick scan. The proton peaks of amino acids such as valine, isoleucine, and leucine are shaded grey. The signal to noise ratio of the spectra are low. In other words the spectra are noisy, because of the less number of scans and measurement time.

By comparing the spectra shown in Figure 8 and 9, 1:5 serum buffer ratio was found to be suitable for our experiments. In other words 100 µl of serum sample were found to be sufficient for our study as there is lesser probability of losing information about the metabolites present, despite using lower sample concentration. In Figure 9 (scaled to show δ 0-1.7 ppm) the amino acids and small proteins peaks that were naturally present in lesser concentration can be seen in all four spectra. The peak intensity was near the noise level, especially in T2, T3, and T4 spectra due to lesser sample volume compared to T1. This was circumvented by signal accumulation i.e. by increasing the number of scans per FID, but S/N can be improved only to the square root of number of scans. For example, 100 times longer measurement increases S/N by only the factor of 10. The peak position misalignments were noticed even after spectral calibration, one such misalignment is shown in Figure 10. This is due to change in pH and it is mostly unavoidable especially in this type of samples (serum). The whole range (δ 0-8.35) spectrum is shown in Figure 11. The list of metabolites identified is presented in Table 5.

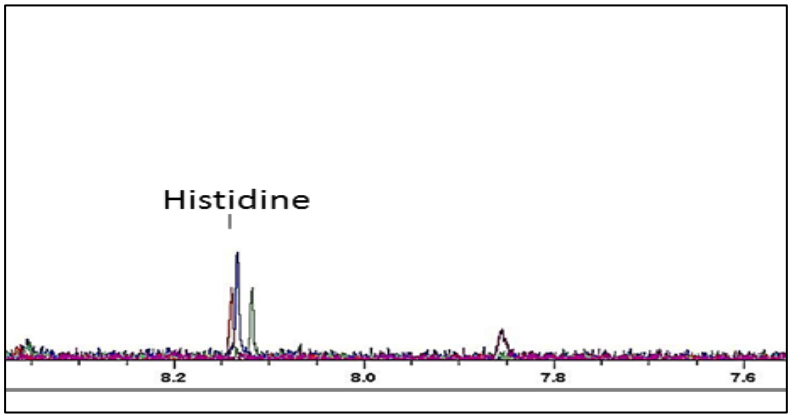


Figure 10: ^{1}H NMR spectra of one sample in each categories overlaid (δ 7.5- 8.35). The protons peaks of histidine can be seen at different chemical shifts for different samples. Amino acid histidine which is known to have nitrogen bound labile protons are readily influenced by slight changes in their chemical environment as manifested in figure 10.

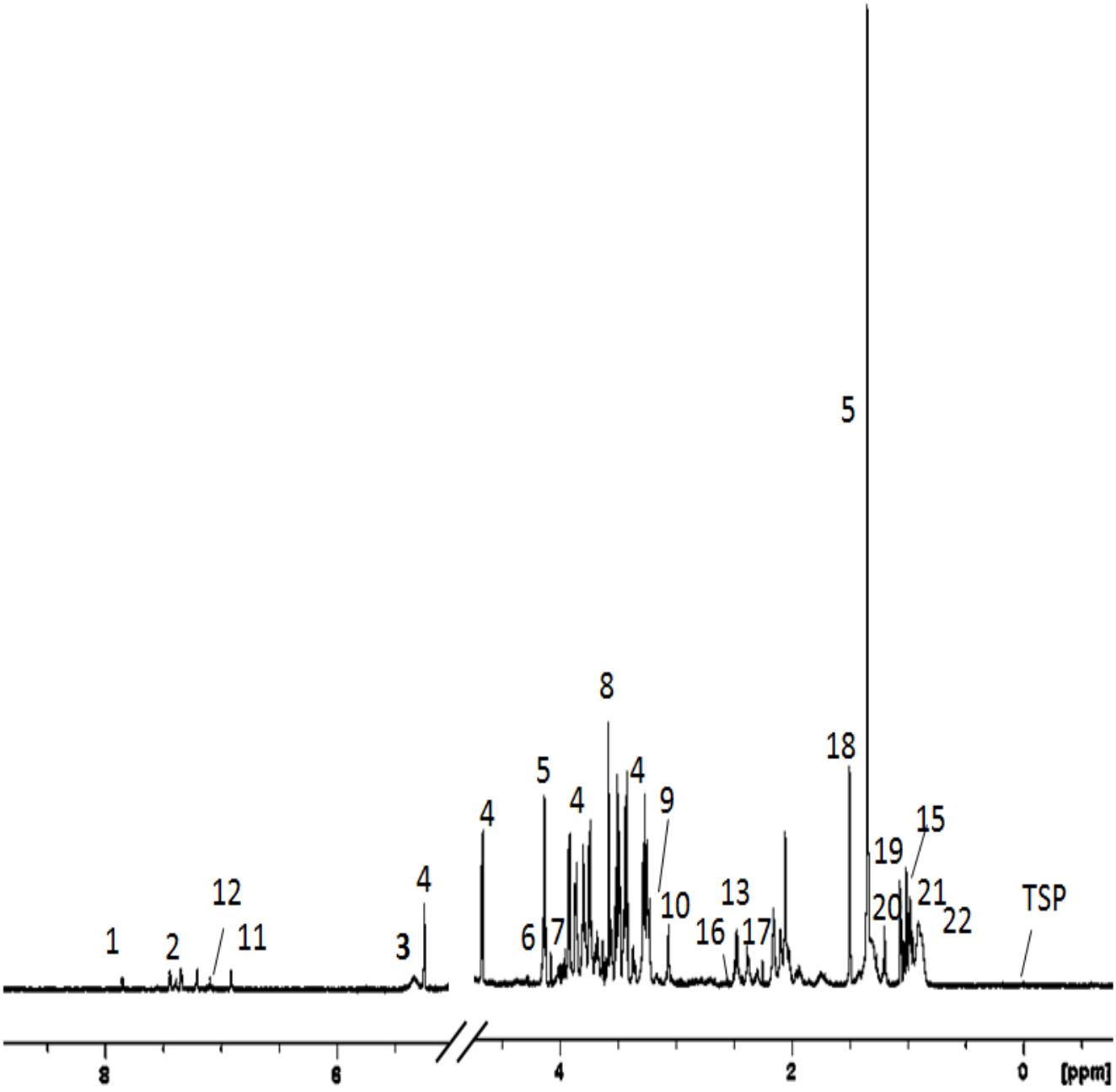


Figure 11: 1D ^{1}H NMR spectra of healthy donor. Region of water signal (δ 4.7-4.9) removed manually. The region above δ 8.2 is not shown because no prominent peaks were noted in that region. Table 5 shows 24 metabolites and their corresponding chemical shifts.

Table 5. Metabolites and the groups responsible for corresponding peak positions.

Metabolite	Group	^{1}H Chemical shift δ (ppm)	No.
α and β Glucose	CH	5.25, 4.65, 3.9, 3.87,3.72, 3.2	4
Unsaturated lipids	CH*=CHCH$_2$	5.33	3

Lactate	CH_3 CH	4.12, 1.3	5
Phenylalanine	Aromatic ring CH	7.34, 7.39	2
Tyrosine	Aromatic ring CH	6.9, 7.2	11
Histidine	Aromatic ring CH	7.09	12
Alanine	β CH_3	1.50	18
Valine	γ CH_3	1.05	19
Choline	$N^+(CH_3)3$	3.22	9
Creatine	N (CH_3)	3.05	10
Glutamate	γ CH_2	2.36	13
Glutamine	γ CH_2	2.46	13
Leucine and Isoleucine	δ CH_3	0.85-0.9	15
Glycine	CH_2	3.573	8
Threonine	CH_2	4.26	6
Serine	CH_2	3.97, 3.99	7
3-Hydroxy-butyrate	CH_3	1.19	20
VLDL	CH_3	0.9, 1.3	21
LDL	CH_3	0.8, 1.2	22
Pyruvic acid	CH_3	2.38	17
Formic acid	OH	8.4	1
Citric acid	CH_2	2.55, 2.66	16

4.1 Metabolic differences

The samples were randomly chosen from each category and overlaid to identify the differences. The standalone spectra and overlaid spectra are shown in Figure 12 a and 12 b. The level of lactate (δ 4.12, 1.3) is found to be high in acute LD and late LD patient sample compared to healthy donor and negative Lyme as can be appreciated from Figure 12 a, simultaneous decrease in the level of glucose (δ 4.65) suggested that there can be a correlation between the levels of these metabolites.

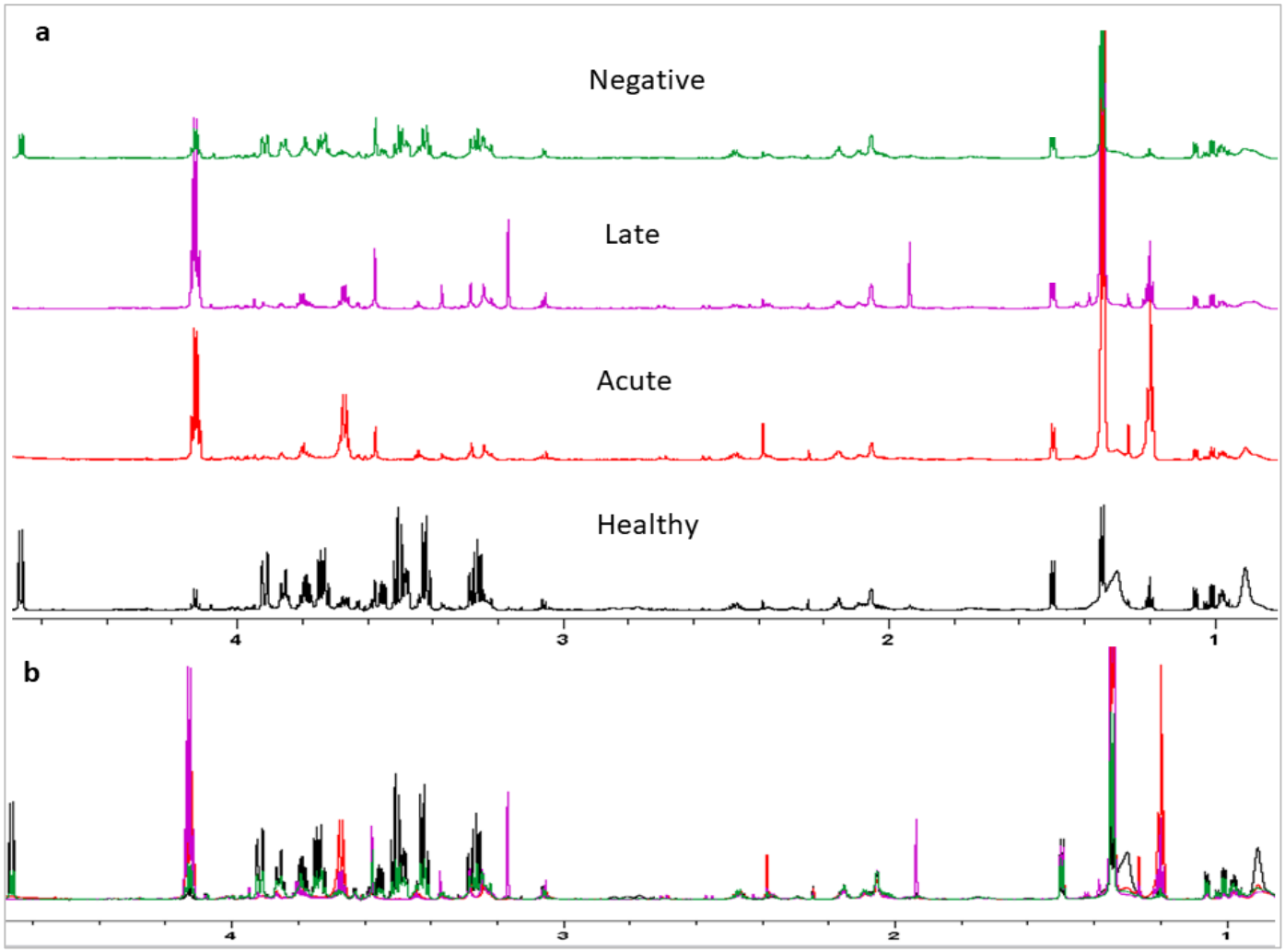


Figure 12: ^{1}H NMR spectra of one sample from each category in different layouts (δ 0.5-4.5 ppm) a) Individual spectra of one random sample in each of the categories b) Overlay spectra of the same samples from each category.The differences among the spectra can be clearly seen in overlay spectra.

However, the difference in the level of metabolites such as glycine, creatine etc., were not consistent, even among the samples of same category. For this reason sum spectra were prepared. The individual spectra of all samples in each category except one sample in negative category were taken into account for sum spectra. All individual spectra in each of the categories were summed up to prepare the sum spectra shown in Figure 13 a and 13 b. The sum spectra showed clear differences in the levels of α glucose, β glucose, lactate, citric acid and choline which eventually prove their importance in biomarker analysis. No remarkable changes were noticed in metabolites such as formic acid, phenylalanine and tyrosine and it can be seen in Figure 13 b. For further investigation, the integral area of each metabolite that represents, the metabolite concentration was compared with the mean integral value of the respective metabolite in healthy donors. The mean integral values represent the concentration of the metabolites. The percentage of acute and late LD sample population with either increased or decreased level of metabolites compared to healthy donors is presented in Table 6.

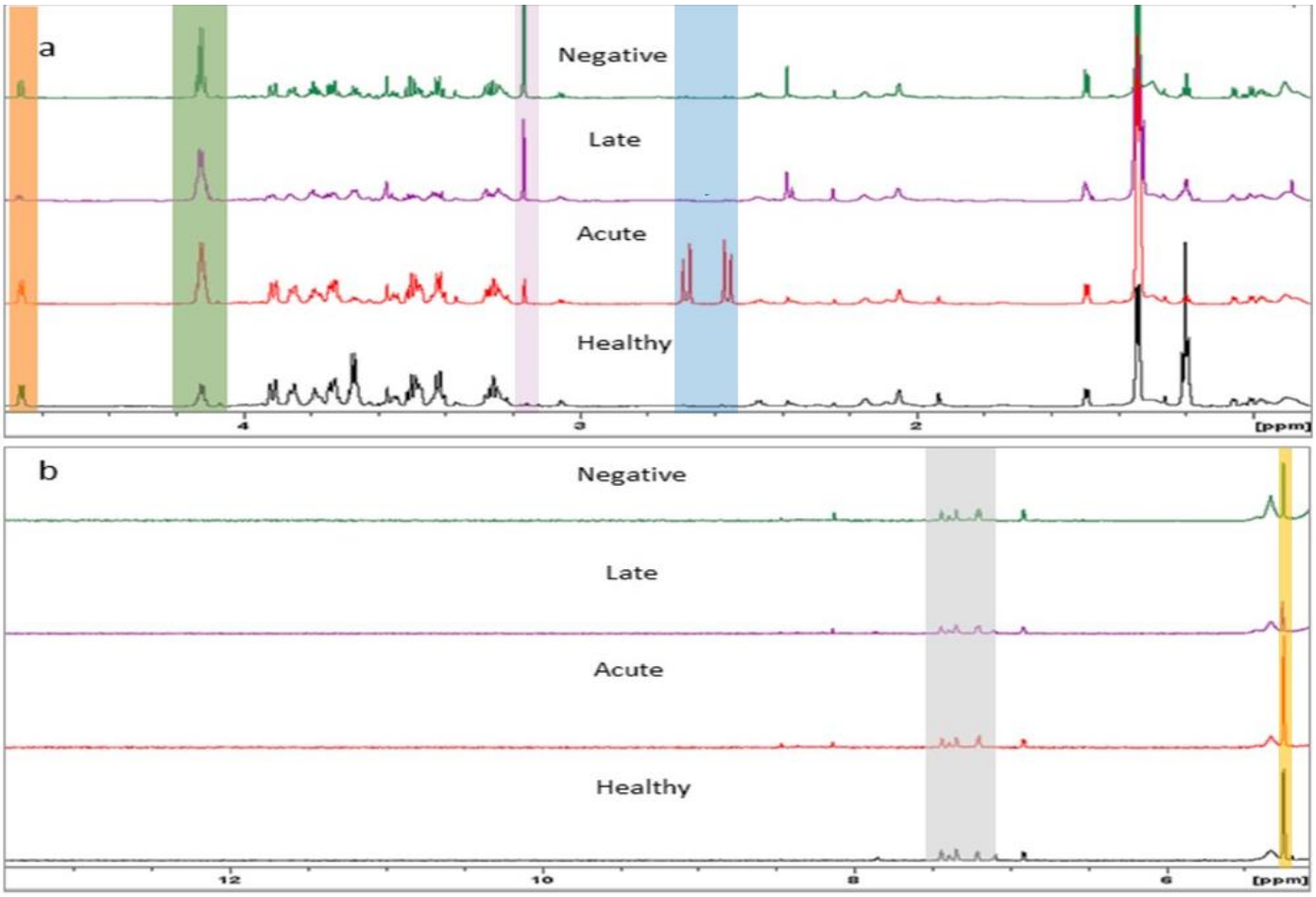


Figure 13: Sum spectra of all sample in each category a) Spectra range δ 0-4.65 ppm- α glucose, β glucose peaks are shaded orange, lactate peaks are shaded green, citric acid peak shaded blue and choline peak shaded pink b) Spectra range δ 5-12 ppm- β glucose peaks are shaded orange, the amino acids phenylalanine and tyrosine peaks are shaded grey. The spectral region δ 4.7-5 ppm that belongs to water signal is not shown.

In 70 % of acute LD patient sera lactate level was higher and in 80 % of acute LD patient sera α glucose, β glucose level was lower compared to respective metabolite levels in healthy donor samples, as presented in Table 6. All acute LD patients with higher level of lactate shown lower level of glucose. Choline and alanine level was lower in 70 % and 80 % acute LD patient sera respectively compared to respective metabolite levels in healthy donor samples, as presented in Table 6. The level of the metabolites such as creatine, glutamine, glycine, leucine, VLDL and unsaturated fatty acids were lower in 60 % to 70 % of the acute LD samples compared to respective metabolite levels in healthy donor samples, as presented in Table 6. Other metabolites namely valine, glutamate and LDL were present in higher level compared to respective metabolite levels in healthy donor samples, in half of the acute LD sample population and lower in the rest of the acute LD sample population, without showing any trend. So these metabolites were considered to be least important metabolites for biomarker analysis. Similarly, in late LD samples α glucose, β glucose, lactate, choline, and alanine seem to follow the same pattern as in the case of acute LD samples i.e. 80-90 % of the acute and late LD patient have decreased level of α glucose, β glucose, choline and alanine; 70-90 % of the acute and late LD patient have increased levels of lactate suggesting that few of these metabolic changes could be the biosignatures of LD patients. The level of the metabolites namely glutamine, leucine, valine, LDL and VLDL were lower

compared to respective metabolite levels in healthy donor samples, in 70-80 % of the late LD sample population, as presented in Table 6. Creatine and unsaturated fatty acids levels where higher compared to respective metabolite levels in healthy donor samples, in 50 % of the late LD samples. Level of glycine was higher compared to glycine level in healthy donors, in 60 % late LD, in contrast with acute LD patients sera, as presented in Table 6. Similarly, level of citric acid and glutamate was higher in 80 % and 70 % of late LD patient sera respectively compared to respective metabolite levels in healthy donor samples, but this was not the case in acute LD samples, as presented in Table 6. Only 60 % of the acute LD sample shown lower level of citric acid compared to citric acid level in healthy donor samples. So glycine, glutamate and citric acid could be the signature feature of late LD samples and might be useful to distinguish late LD from acute LD samples. Notably in negative Lyme samples, compared to healthy donors the level of α glucose, β glucose, glutamate, glutamine, glycine, leucine, valine VLDL were lower in 67 % of the samples compared to respective metabolite levels in healthy donor samples, level of lactate was higher in 67 % of the sample and alanine level was higher in 22 % of the samples, choline was lower in 89 % as presented in Table 6. Creatine level was higher in 78% of the negative Lyme samples compared to respective metabolite levels in healthy donor samples, as presented in Table 6. We cannot draw any conclusions about the utility of the identified metabolites as biomarkers at this point. Yet, these information about the metabolic difference in samples were useful in statistical biomarker analysis.

Histograms were prepared for fifteen metabolites with mean of the integral values of metabolities to compare the categories and identify the general differences. Histograms reveal the overall metabolic changes in each of the categories, similar to sum spectra. Histograms shown that the levels of metabolites glycine, alanine, LDL, VLDL, unsaturated fatty acids, α glucose, β glucose, lactate and citric acid were higher in acute LD patient samples compared to respective metabolites in the healthy donor samples and choline was lower compared to choline in the healthy donor samples [Refer appendix I]. In late LD patient samples, there was no change in the level of glycine and LDL, unlike the acute LD patients but an noticeable increase in the level of lactate, citric acid, VLDL, unsaturated fatty acids, creatine and decrease in the levels of alanine, choline, α glucose and β glucose compared to the healthy donors, were also observed [Refer appendix I]. In negative Lyme sample population choline, glycine and alanine level was low and citric acid, LDL, unsaturated fatty acids, α glucose, β glucose levels were higher and no significant changes can be noticed in VLDL and creatine [Refer appendix I].

Table 6. Percentage of the acute LD, late LD and negative Lyme samples with elevated (high) or decreased (low) level of metabolites compared to healthy donors. One sample in negative Lyme category had exactly the same level of α glucose compared to healthy donors, but it was counted as high (*).

Metabolites	Acute LD		Late LD		Negative Lyme	
	High	Low	High	Low	High	Low
α glucose	20	80	10	90	33 *	67

β glucose	20	80	10	90	33	67
Lactate	70	30	90	10	67	33
Choline	30	70	20	80	11	89
Creatine	40	60	50	50	22	78
Glutamine	40	60	40	60	33	67
Glutamate	50	50	30	70	33	67
Glycine	40	60	60	40	33	67
Leucine	40	60	30	70	33	67
Valine	50	50	40	60	33	67
Alanine	20	80	20	80	22	78
LDL	50	50	20	80	55	45
VLDL	40	60	20	80	33	67
Citric acid	60	40	80	20	100	-
Unsaturated fatty acids	40	60	50	50	67	33

4.2 Biomarker analysis

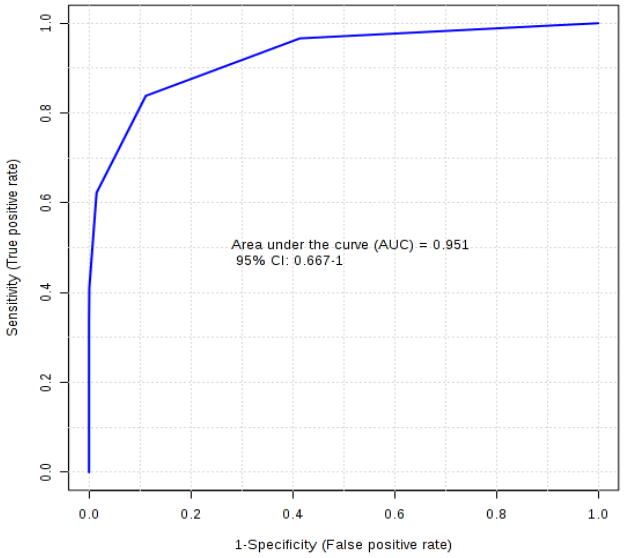
PLS-DA algorithm was tested with a data set (containing integrals values of metabolities in two samples the belong to two different category) to check the predictive accuracy of PLS-DA and the corresponding result obtained from the MetaboAnalyst tool is translated in Table 7. The predictive accuracy test result of PLS-DA algorithm obtained from MetaboAnalyst demonstrated that 34 feature model has the highest predictive accuracy of 87 % [Refer appendix III]. From now on, metabolite and ratios of metabolites will be mentioned as features for convenience.

Table 7. Predictive accuracy with different features.

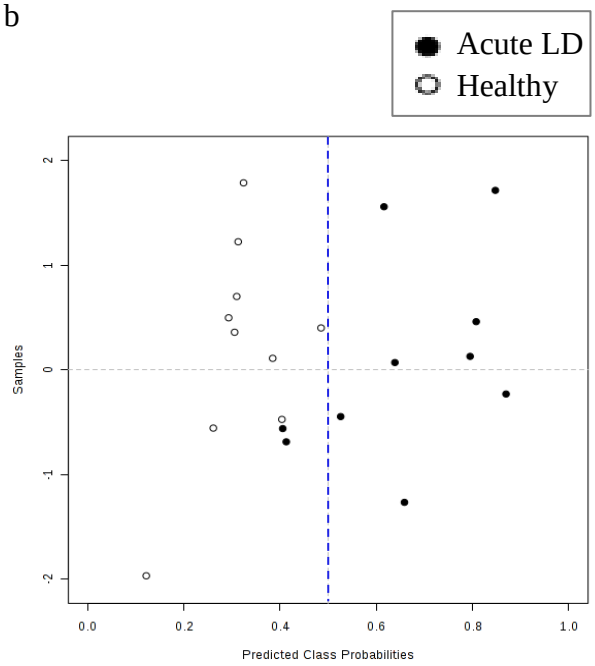
Numbers of features	Predictive accuracy
	PLS-DA (%)
5	84
10	81
34	87

The features of this 34 feature model includes 15 individual metabolites and 19 metabolite ratios. These 19 metabolites ratios were suggested automatically by the biomarker analysis tool based on their individual AUC values. The software also allows the user to select features for class prediction based their own theories.

a



b



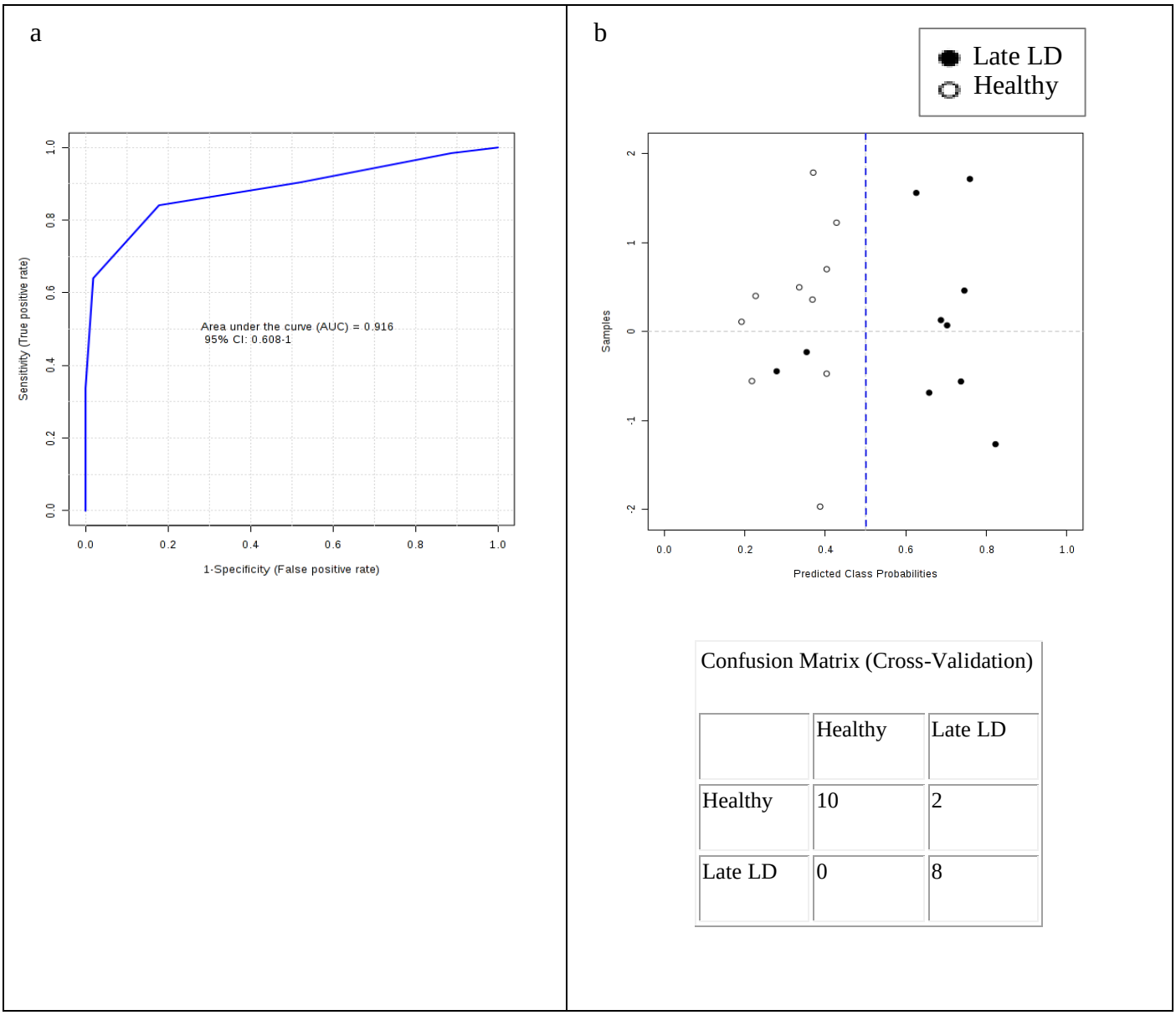
Confusion Matrix (Cross-Validation)

	Healthy	Acute LD
Healthy	10	2
Acute LD	0	8

Figure 14: Biomarker evaluation results of healthy donors and acute LD data set a) ROC curve of 34 feature model b) Class prediction plot using 34 features- the classification boundary is the dashed blue line at 0.5 in x-axis. For list of all 34 features refer appendix IV.

The biomarker analysis tool available in MetaboAnalyst software allows the user to compare only two sets of data at a time. So each category was compared against each other as a pair. ROC curve of 34 feature model of a healthy donor vs acute LD data sets is shown in Figure 14 a. The AUC value of ROC curve is 0.951, which suggests that 34 features model is an excellent model [Refer Table 1] for acute LD and healthy donor classification. Average of the predicated class probabilities of each sample across 100 cross-validations using 34 feature model is shown in Figure 14 b. The matrix presented along, shows that out of 10 healthy donors samples all 10 healthy donor samples were correctly classified as healthy donor or true negative based on the 34 features and out of 10 acute LD samples 8 were correctly classified as patient sample and only 2 samples

were wrongly classified or grouped with healthy donor samples. From equation 1.1 and 1.2, the sensitivity of the 34 feature model for the acute LD and healthy donor dataset is 80 % and the specificity is 100 %.



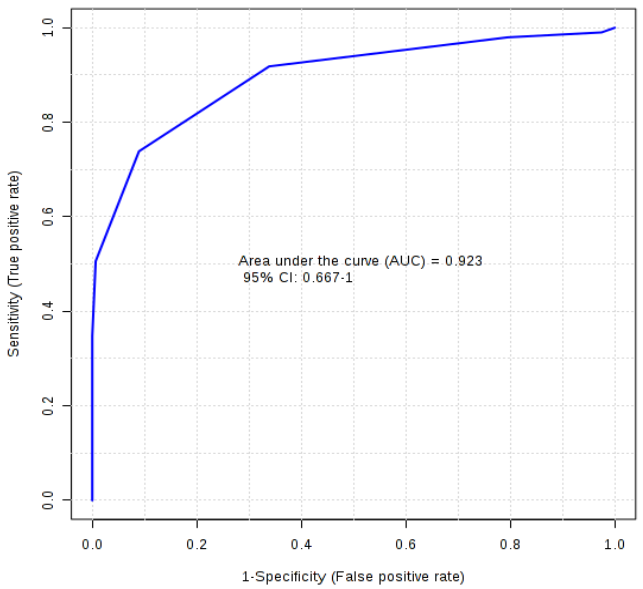
Confusion Matrix (Cross-Validation)

	Healthy	Late LD
Healthy	10	2
Late LD	0	8

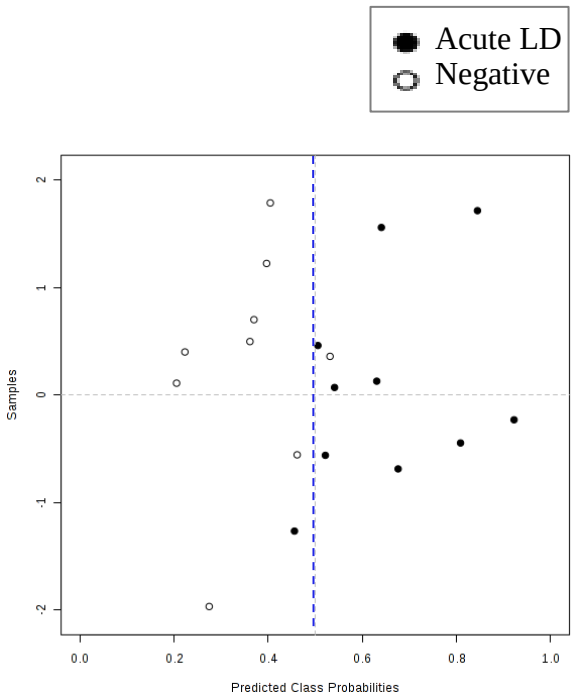
Figure 15: Biomarker evaluation results of healthy donor and late LD data set a) ROC curve of 34 feature model b) Class prediction plot using 34 feature model-the classification boundary is the dashed blue line at 0.5 in x-axis. For the list of all 34 features refer appendix IV.

ROC curve and class prediction plot of 34 feature model of healthy donors and late LD data sets is shown in Figure 15 a and b. The AUC value of ROC curve is 0.916, which suggests that 34 features model is an excellent model [Refer Table 1] for the healthy donor and late LD classification. The class prediction plot in Figure 15 b and the matrix presented along, shows that out of 10 healthy donors samples all 10 samples were correctly classified as healthy donor and out of 10 acute LD samples 8 were correctly classified and two samples were wrongly grouped with healthy donor samples. The sensitivity of the 34 feature model for the late LD and healthy donor dataset is 80 % and the specificity is 100 %.

a



b

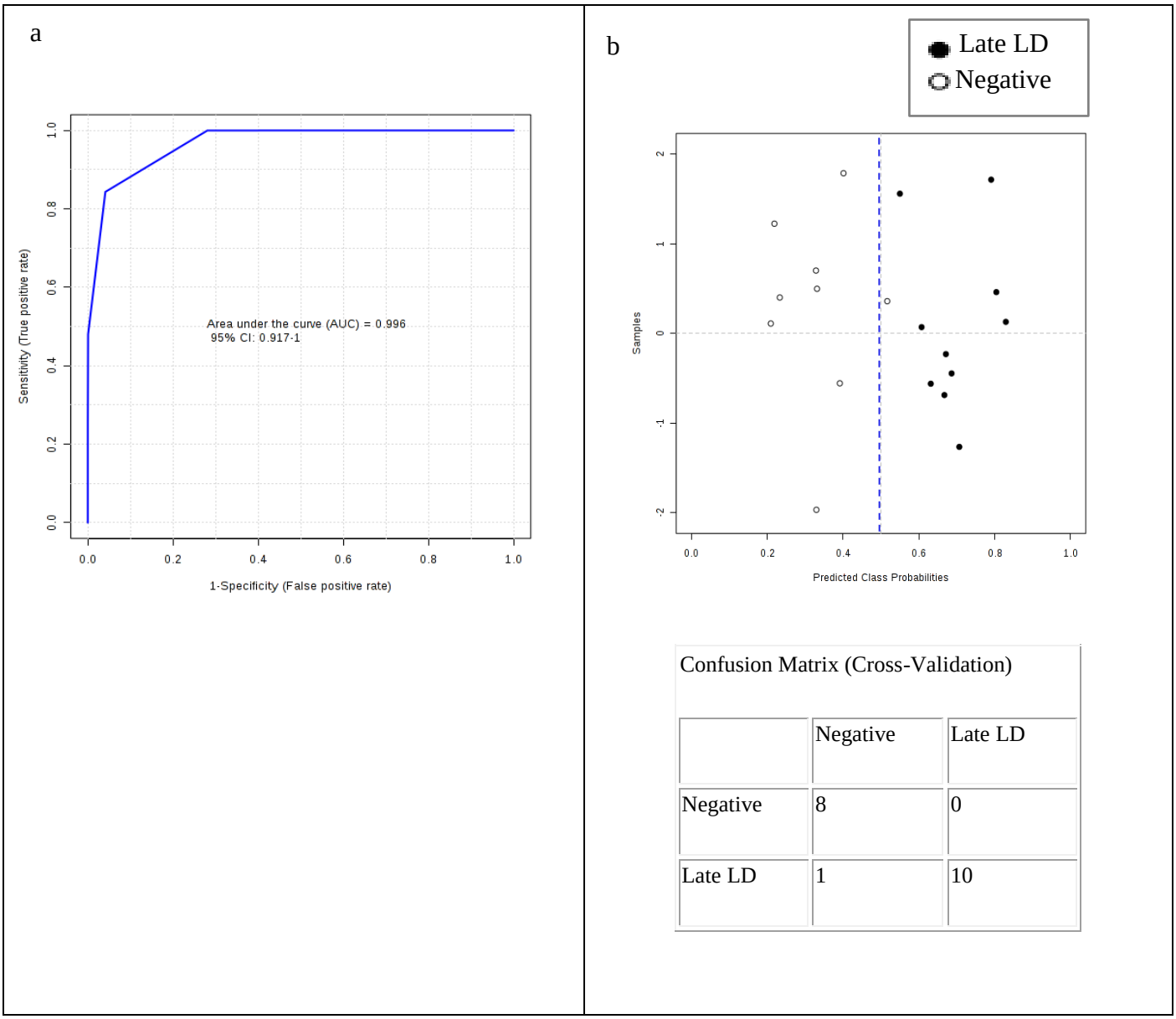


Confusion Matrix (Cross-Validation)

	Negative	Acute LD
Negative	8	1
Acute LD	1	9

Figure 16: Biomarker evaluation results of negative Lyme and acute LD data set a) ROC curve of 34 feature model b) Class prediction plot using 34 feature model-the classification boundary is the dashed blue line at 0.5 in x-axis. For the list of all 34 features refer appendix IV.

The 34 feature model was tested for acute LD and negative Lyme data set. This was done to understand whether the identified metabolic changes are unique features of LD patients. The biomarker analysis results of acute LD and negative Lyme data set are presented in Figure 16. ROC curve analysis results are shown Figure 16 a, the AUC value of the ROC curve is 0.923. The class prediction plot in Figure 16 b and the matrix presented along, shows that out of 9 negative Lyme samples, 8 samples were correctly classified as negative Lyme samples and out of 10 acute LD samples 9 were correctly classified as acute LD samples and only 1 sample was wrongly grouped with negative Lyme samples. The sensitivity of the 34 feature model for the acute LD and negative Lyme dataset is 90 % and the specificity is 89 %.



Confusion Matrix (Cross-Validation)

	Negative	Late LD
Negative	8	0
Late LD	1	10

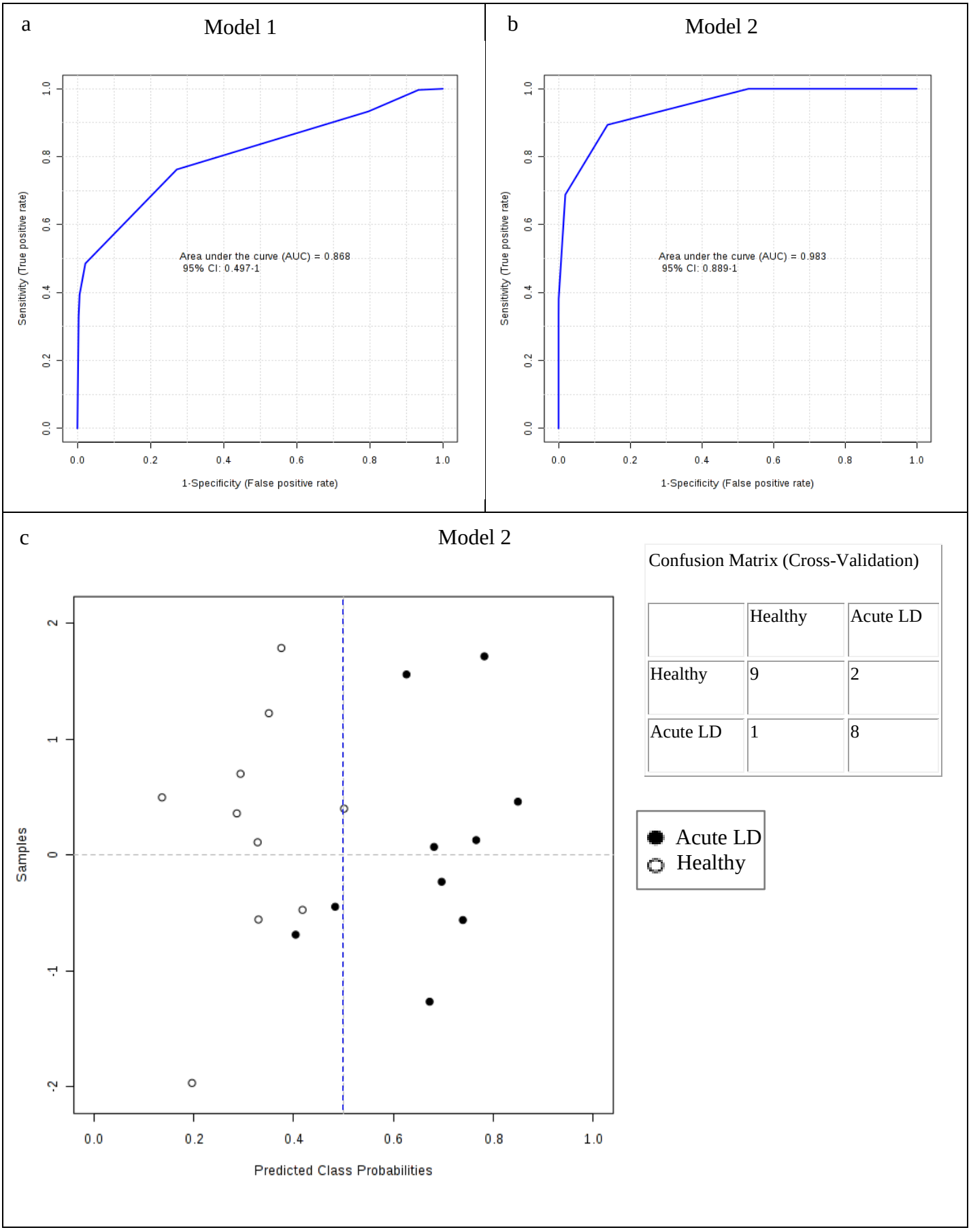
Figure 17: Biomarker evaluation results of negative Lyme vs late LD a) ROC curve of 34 feature model b) Class prediction plot of 34 feature model-the classification boundary is the dashed blue line at 0.5 in x-axis. For list of all 34 features refer appendix IV.

In a similar way, the 34 feature model was tested for late LD and negative Lyme data set. The biomarker analysis results of late LD and negative Lyme data set are presented in Figure 17. ROC curve analysis results is shown Figure 17 a, the AUC value of the ROC curve is 0.996. The class prediction plot in Figure 17 b and the matrix presented along, shows that out of 9 negative Lyme samples, 8 samples were correctly classified as negative Lyme samples based on the 34 features and out of 10 late LD samples all 10 were correctly classified as late LD samples. The sensitivity of the 34 feature model for the late LD and negative Lyme dataset is 100 % and the specificity is 89 %. From a practical point of view, detecting the levels of more than 10 different metabolites in patients for clinical diagnosis of LD will be a difficult task and cost effectiveness of such as diagnostic method is a question mark. So it is appropriate to aim for less number of features, which can be utilized as biomarkers. For this purpose five feature models in which different choices and combinations of features were chosen, is tested. The predictive accuracy of five feature model is 84 %, as

presented in Tabel 7. The AUC values and fold change of the individual features of healthy donor and acute LD data set are presented in Table 8. Based on the four criterions mentioned in section 3.3, lactate, β glucose, citric acid, choline and alanine were selected and named as model 1 for convenience. Model 1 was evaluated. The ROC curve was generated and presented in Figure 18 a. AUC value of ROC curve analysis of model 1 for the healthy donor and acute LD data set is 0.868. Similarly, model 2 with features such as citric acid, lactate/alanine, α glucose/lactate, β glucose/lactate, β glucose/citric acid was tested and the results are presented in Figure 18 b. The ROC curve analysis of the model 2 (AUC 0.983), suggests that lactate/alanine, α glucose/lactate, β glucose/lactate, β glucose/citric acid and citric acid can be the interesting features that we are hunting for.

Table 8. AUC and fold change corresponding to individual features in healthy donors and acute LD [Refer appendix IV for the AUC of all 34 features including the top 20 metabolite ratios]

Features	AUC	Fold Change
Lactate/Alanine	0.99	1.5436
α glucose/Citric acid	0.9	-0.33907
β glucose/Citric acid	0.9	-0.20773
α glucose/Lactate	0.84	1.5436
β glucose/Lactate	0.82	-0.07999
α glucose	0.86	1.5436
β glucose	0.83	1.5104
Lactate	0.8	-1.6579
Citric acid	0.77	-1.3877
Choline	0.645	0.17834
Alanine	0.62	0.13759
Glutamine	0.6	-0.07999
Glycine	0.59	-0.20773
Unsaturated	0.58	0.022535
VLDL	0.58	0.071083
Creatine	0.56	-0.33907
Leucine	0.54	-0.14752
LDL	0.53	-0.21441
Glutamate	0.5	-0.32629
Valine	0.5	-0.11413



Confusion Matrix (Cross-Validation)

	Healthy	Acute LD
Healthy	9	2
Acute LD	1	8

Figure 18: Biomarker evaluation results of healthy donor and acute LD data sets using five features a) ROC curve of five features model-Model 1 b) ROC curve of five features model-Model 2 c) Class prediction plot using model 2

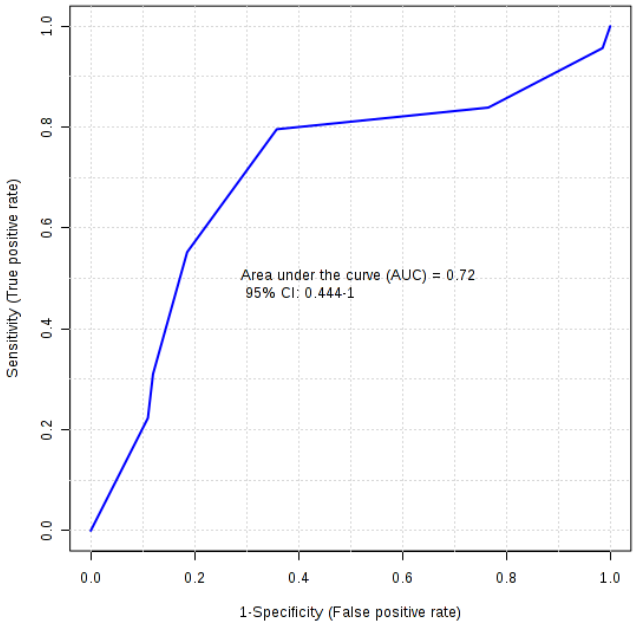
The AUC values and fold change of the individual features of healthy donors and late LD data set are presented in Table 9. Model 1 was tested and evaluated for the healthy donor and late LD dataset. The ROC curve analysis of model 1 for this data set is shown in Figure 19 a. The ROC curve (AUC 0.72) obtained indicates that the features used in model 1 may not the best choice for this classification and stresses us to move for different choice of features. Model 3 which includes features such as citric acid, α glucose / lactate, β glucose lactate, glucose/citric acid and choline/glycine was tested. The ROC curve analysis results of model 3 (AUC 0.979) shown in Figure 19 b, indicates that based on this model 3, late LD patients can be distinguished from healthy donors with excellent sensitivity 80 % and specificity 100 %.

Table 9. AUC and fold change corresponding to individual features in healthy donors and late LD samples. [Refer appendix IV for the AUC of all 34 features including the top 19 ratios]

Features	AUC	Fold Change
Choline/Glycine	0.94	-1.2798
β glucose/Citric acid	0.89	2.9602
α glucose/Citric acid	0.88	-0.20454
α glucose/Lactate	0.86	-0.448
β glucose/Lactate	0.84	0.23507
α glucose	0.86	2.583
β glucose	0.86	2.9602
Lactate	0.86	-1.2798
Citric acid	0.695	-0.27711
Choline	0.69	0.45362
Unsaturated	0.68	-0.448
VLDL	0.63	0.14051
Leucine	0.62	0.23507
Creatine	0.6	-0.16996
Glutamate	0.57	-0.21181
Valine	0.57	-0.00147
Glutamine	0.56	0.076048
Alanine	0.55	0.10184

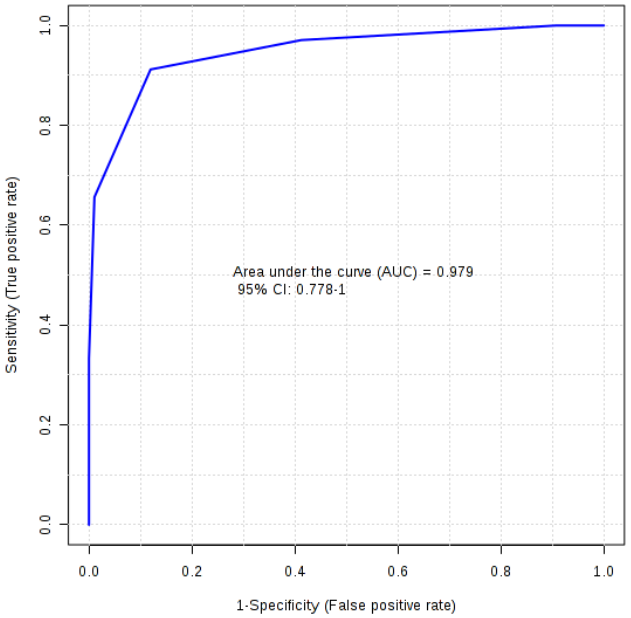
a

Model 1



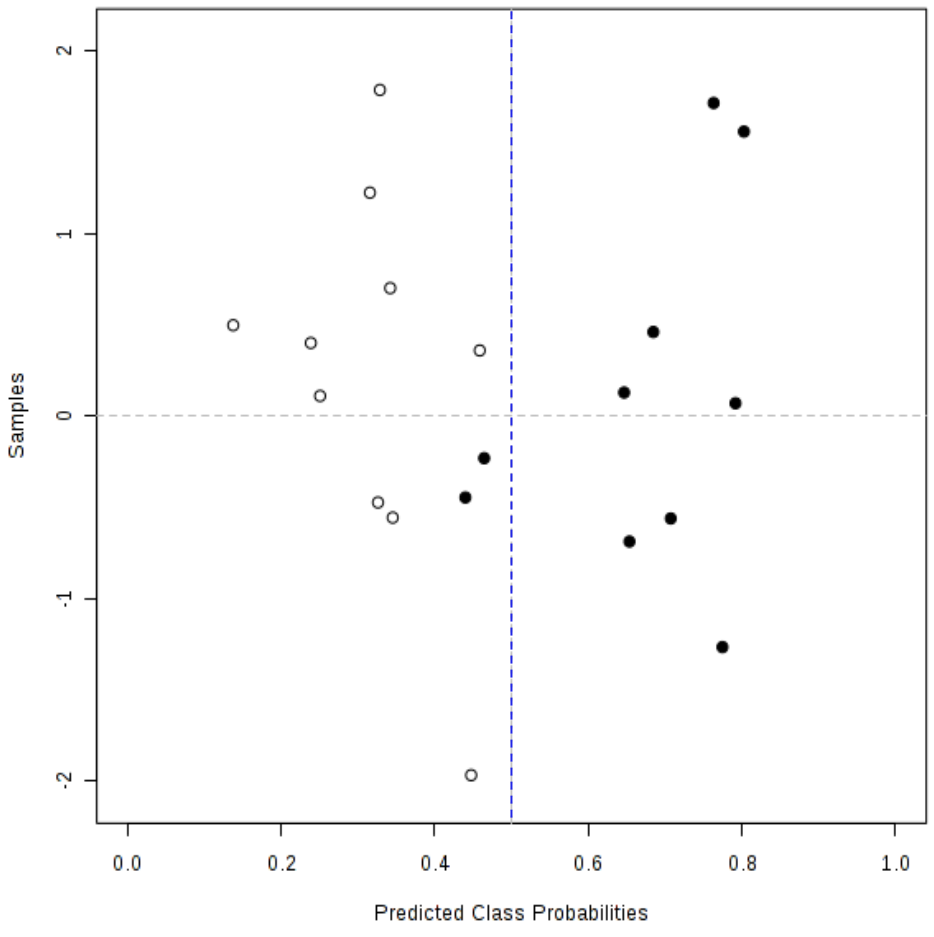
b

Model 3



c

Model 3



Confusion Matrix (Cross-Validation)

	Healthy	Late LD
Healthy	10	2
Late LD	0	8

Late LD
Healthy

Figure 19: Biomarker evaluation results of healthy donor and late LD data sets using five features a) ROC curve of five features model-Model 1 b) ROC curve of five feature model-Model 3 c) Class prediction plot using model 3

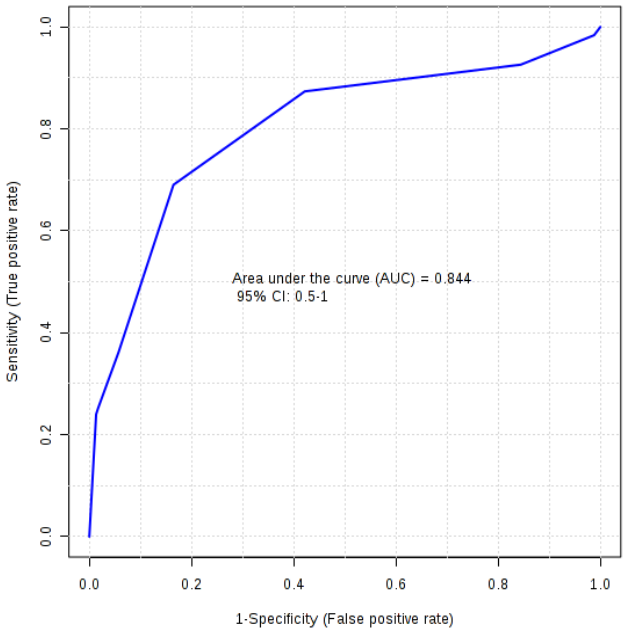
The AUC values and fold change of the features of acute LD and negative Lyme data set are presented in Table 10. Model 1 and 2 were tested for acute LD and negative Lyme samples data set. The ROC curve analysis and class prediction results are presented in Figure 20. The ROC curve of model 1 (AUC 0.844) is shown in Figure 20 a. The ROC curve of model 2 (AUC 0.869) is shown in Figure 20 b. Another model, model 4 was built with features such as choline, valine, alanine, LDL and unsaturated fatty acids and the test, the ROC curve analysis result (AUC 0.836) was similar to model 1 and 2. A new model, model 5 with features such as unsaturated fatty acids, choline/valaine, choline/alanine, choline/LDL and unsaturated fatty acids/choline was tested. The ROC curve is presented in Figure 21 b. The ROC curve of model 5 (AUC 0.789) built shows that based on these features acute LD samples can be classified from negative Lyme.

Table 10. AUC and fold change corresponding to individual features in acute LD and negative Lyme samples. [Refer appendix IV for the AUC of all 34 features including the top 19 ratios]

Feature	AUC	Fold Change
Choline/Valine	0.94	-0.12519
Choline/Alanine	0.94	1.8015
Choline/LDL	0.91	-0.23055
Choline/Glycine	0.9	1.3482
Unsaturated/Choline	0.88	-0.096599
Unsaturated	0.77	1.8015
LDL	0.72	1.3482
Choline	0.71	-0.68828
Unsaturated/Lactate	0.7	-0.61597
Lactate/Alanine	0.66	0.65634
Glutamine	0.65	0.70928
Creatine	0.63	0.74165
Glycine	0.63	0.65634
Valine	0.63	0.74228
VLDL	0.63	0.76789

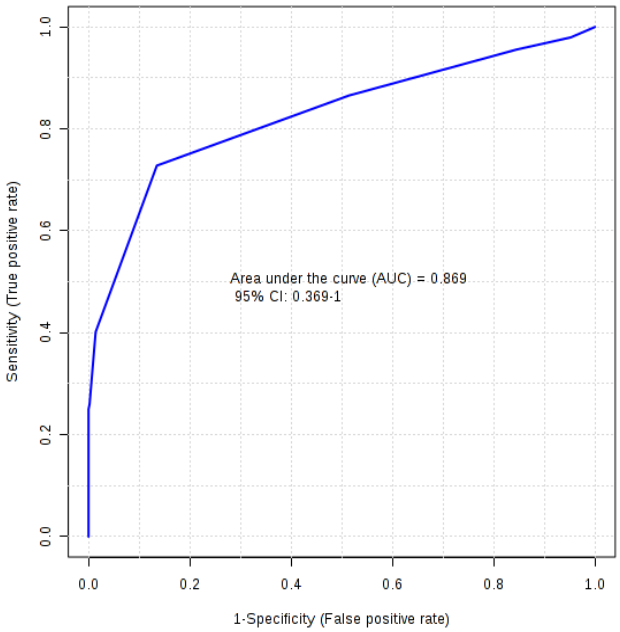
a

Model 1



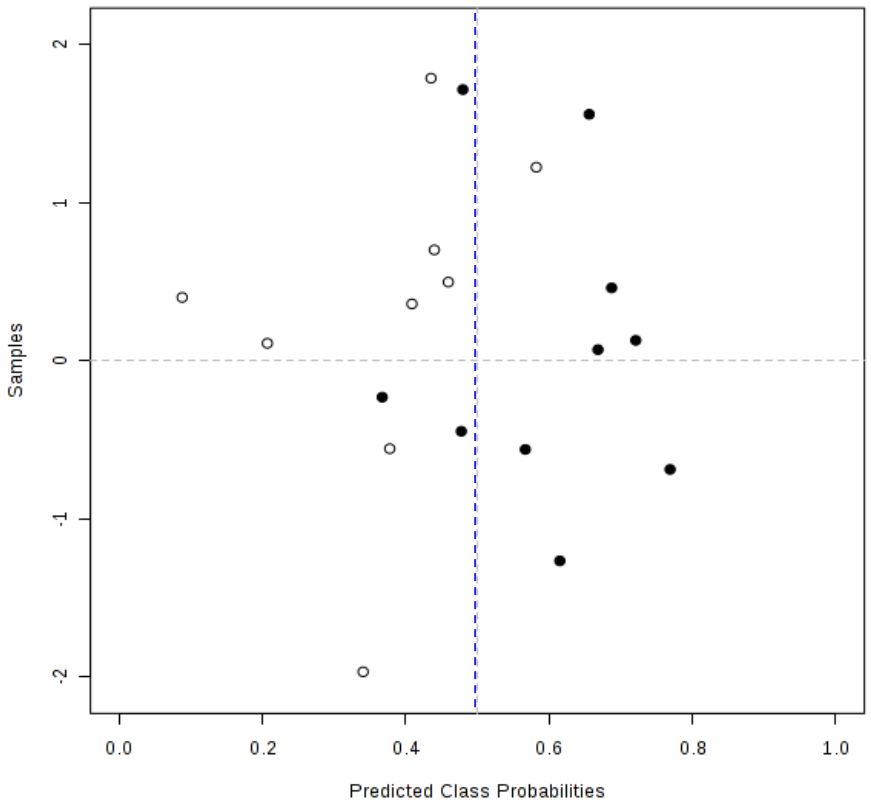
b

Model 2



c

Model 2

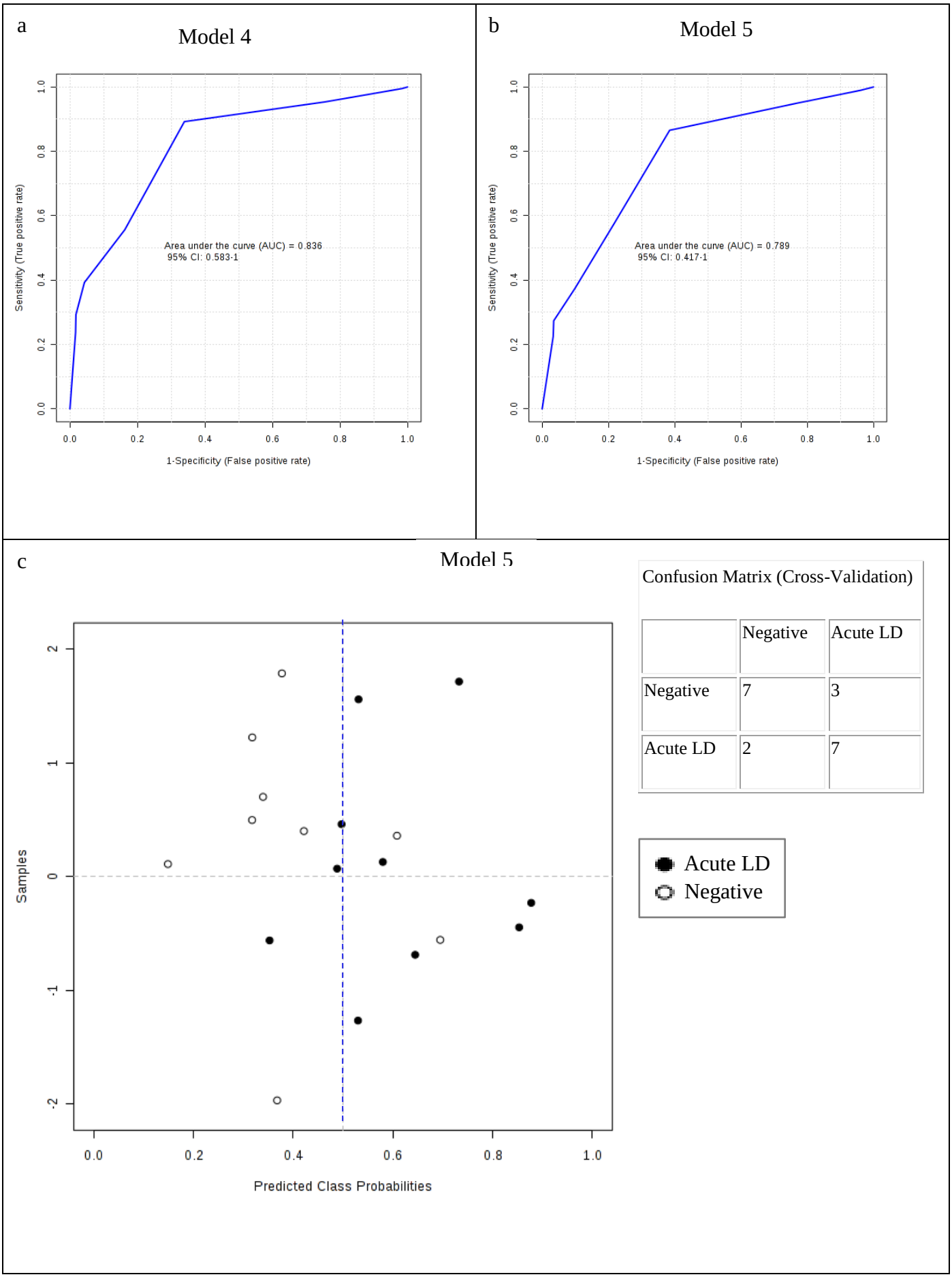


Confusion Matrix (Cross-Validation)

	Negative	Acute LD
Negative	8	3
Acute LD	1	7

Acute LD
Negative

Figure 20: Biomarker evaluation results of negative Lyme and acute LD data set a) ROC curve of five feature model-Model 1 b) ROC curve of five feature model- Model 2 c) Class prediction plot using model 2



Confusion Matrix (Cross-Validation)

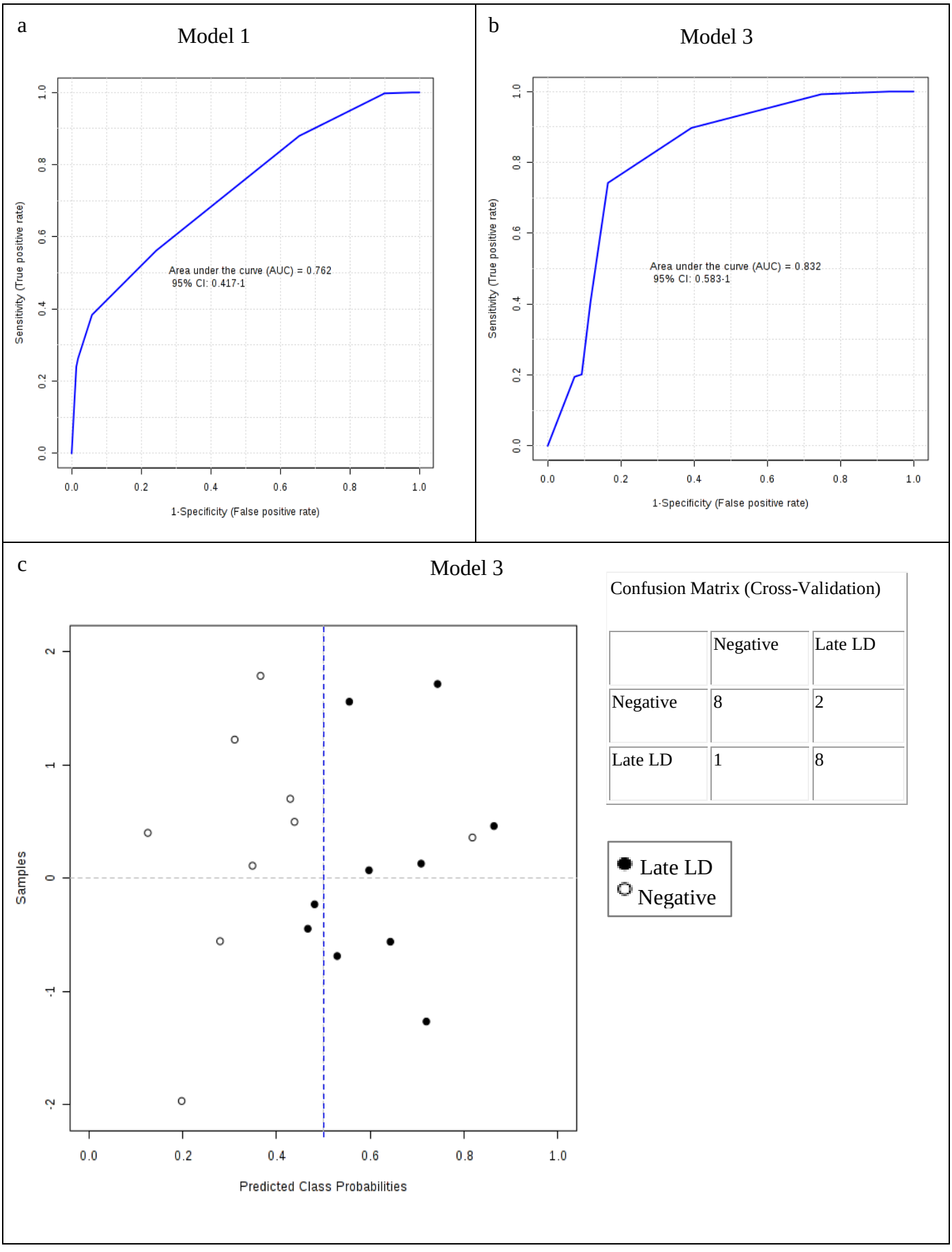
	Negative	Acute LD
Negative	7	3
Acute LD	2	7

Figure 21: Biomarker evaluation results of negative Lyme and acute LD data sets using five features a) ROC curve of five features model-Model 4 b) ROC curve of five features model-Model 5 c) Class prediction plot using model 5

The class prediction plot of negative Lyme and acute LD data set using model 2, presented in Figure 20 c and the matrix presented along, shows that out of 9 negative Lyme samples, 8 samples were correctly classified as negative Lyme samples and out of 10 acute LD samples, 7 were correctly classified as acute LD samples. The class prediction plot in Figure 21 c using model 5 and the matrix presented along, shows that out of 9 negative Lyme samples, 7 samples were correctly classified as negative Lyme samples and out of 10 acute LD samples, 7 were correctly classified as acute LD samples. Late Lyme and negative Lyme data set was tested with model 1, 3, 4 and 5. The AUC values and fold change of the individual features of late LD and negative Lyme data set are presented in Table 11. ROC curve analysis results of model 1 (AUC 0.762), model 3 (AUC 0.832), model 4 (AUC 0.782) and model 3 (AUC 0.93) are presented in Figure 22 a, Figure 22 b, Figure 23 a and Figure 23 b.

Table 11. AUC and fold change corresponding to individual features in late LD and negative Lyme samples. [Refer appendix IV for the AUC of all 34 features including the top 19 ratios]

Features	AUC	Fold Change
Choline/Leucine	1.0	-0.26352
Choline/LDL	1.0	-0.41624
Choline/Valine	0.97778	0.015272
Unsaturated/Choline	0.95556	0.52059
Choline/Glycine	0.87778	-1.2221
Choline/Alanine	0.86667	-0.4475
Choline	0.83333	-1.2221
Glycine	0.72222	-0.41624
Lactate	0.7	-0.88652
LDL	0.68889	0.76736
Glutamate	0.68333	-0.4475
Unsaturated	0.61111	0.52059
Creatine	0.61111	-0.26352
Leucine	0.58889	0.30063



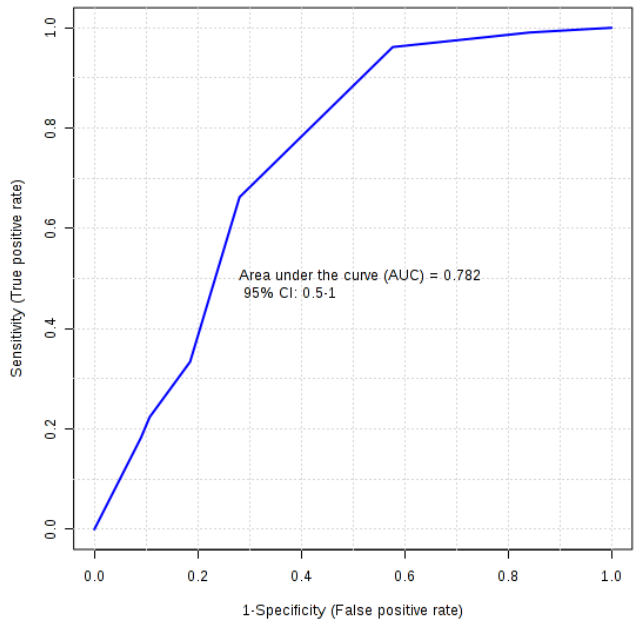
Confusion Matrix (Cross-Validation)

	Negative	Late LD
Negative	8	2
Late LD	1	8

Figure 22: Biomarker evaluation results of negative Lyme and late LD data sets using five features a) ROC curve of five features model-Model 1 b) ROC curve of five features model-Model 3 c) Class prediction plot using model 3

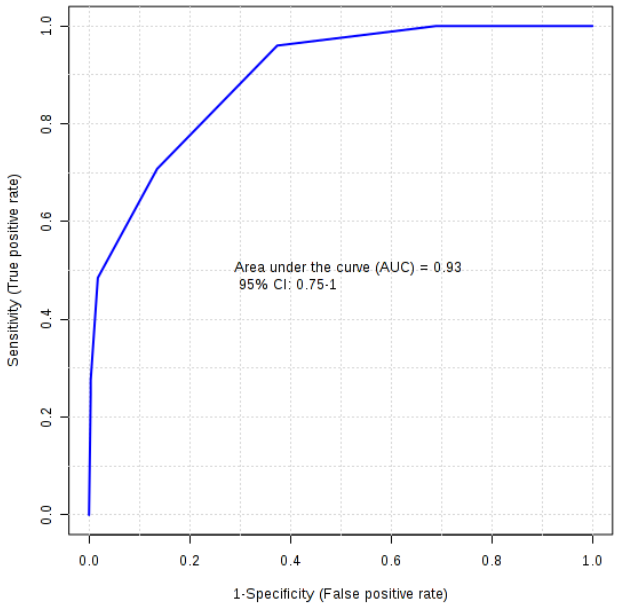
a

Model 4



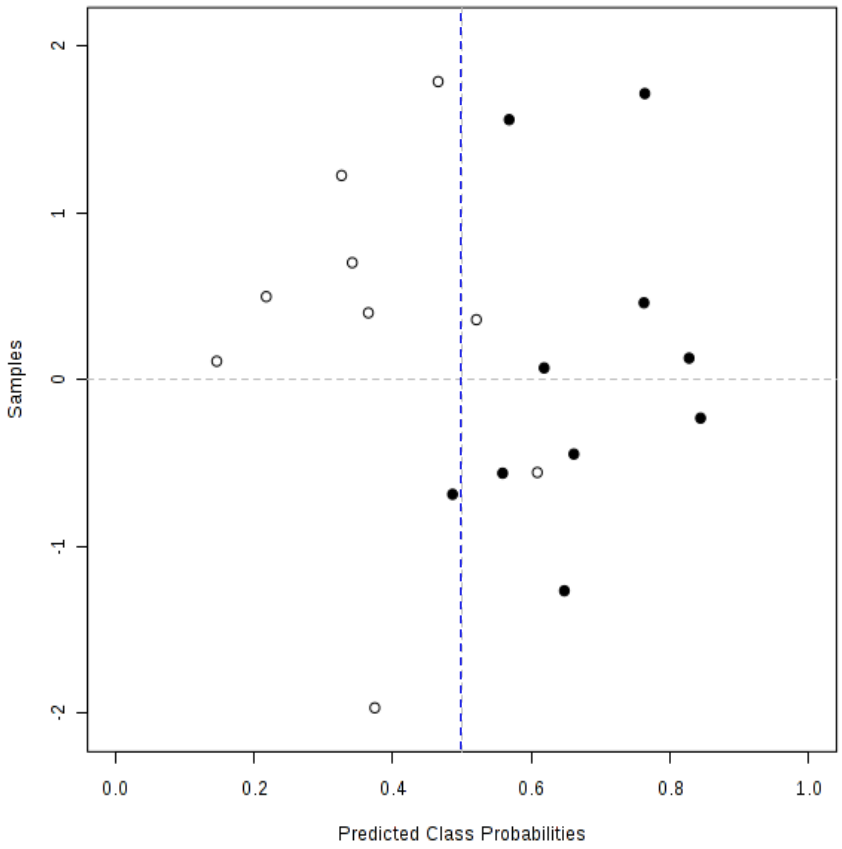
b

Model 5



c

Model 5



Confusion Matrix (Cross-Validation)

	Negative	Late LD
Negative	7	1
Late LD	2	9

Late LD
Negative

Figure 23: Biomarker Evaluation results of negative Lyme and late LD dataset using five features a) ROC curve of five features model-Model 4 b) ROC curve of five features model-Model 5 c) Class prediction plot using model 5

The class prediction plot of negative Lyme and late LD dataset using model 3 is shown in Figure 22 c and the matrix presented along, shows that out of 9 negative Lyme samples, 8 samples were correctly classified as negative Lyme samples and out of 10 late LD samples, 8 were correctly classified as late LD samples. The class prediction plot in Figure 23 c using model 5 and the matrix presented along, shows that out of 9 negative Lyme samples, 7 samples were correctly classified as negative Lyme samples and out of 10 late LD samples, 9 were correctly classified as late LD samples.

An attempt was made to distinguish acute LD from late LD samples. The AUC values and fold change of the individual features of acute LD and late LD data set are presented in Table 12. All models, model 1, model 2, model 3, model 4 and model 5 were tested for acute LD and late LD dataset. The ROC analysis of all these models are presented in Figure 24 and 25. The results indicates that none of these model as capable of differentiating acute LD from late LD. Even the model 5 which worked quite well for acute LD-negative Lyme classification and extremely well for late LD-negative Lyme did not work well acute LD and late LD classification, as it can be seen in Figure 25 b. Model 5 is poor model for acute LD and late LD classification because the AUC value is 0.37 the least score so far. Model 6 with features leucine, valine, alanine, glycine and citric acid was tested. The ROC curve of model 6 shown in Figure 26 a, (AUC 0.43) indicates that this is a poor model. Model 7 with features leucine/valine, leucine/alanine, glycine/citric acid, glycine and citric acid was tested and the ROC curve shown in Figure 26 b (AUC 0.76) seem to be a fair model.

Table 12. Acute LD and late LD comparison and the ROC values of individual metabolites. [Refer appendix IV for the AUC of all 34 features including the top 19 ratios]

Features	AUC	Fold Change
Leucine/Valine	0.83	1.566
Leucine/Alanine	0.77	0.37716
Glycine/Citric acid	0.73	1.1805
β glucose	0.79	1.566
α glucose	0.69	1.1805
Glycine	0.65	0.001311
Unsaturated	0.6	-0.47573
Creatine	0.6	0.15917
Citric acid	0.58	1.0788
Alanine	0.57	-0.03581
Leucine	0.57	0.38529
Glutamate	0.53	0.10914

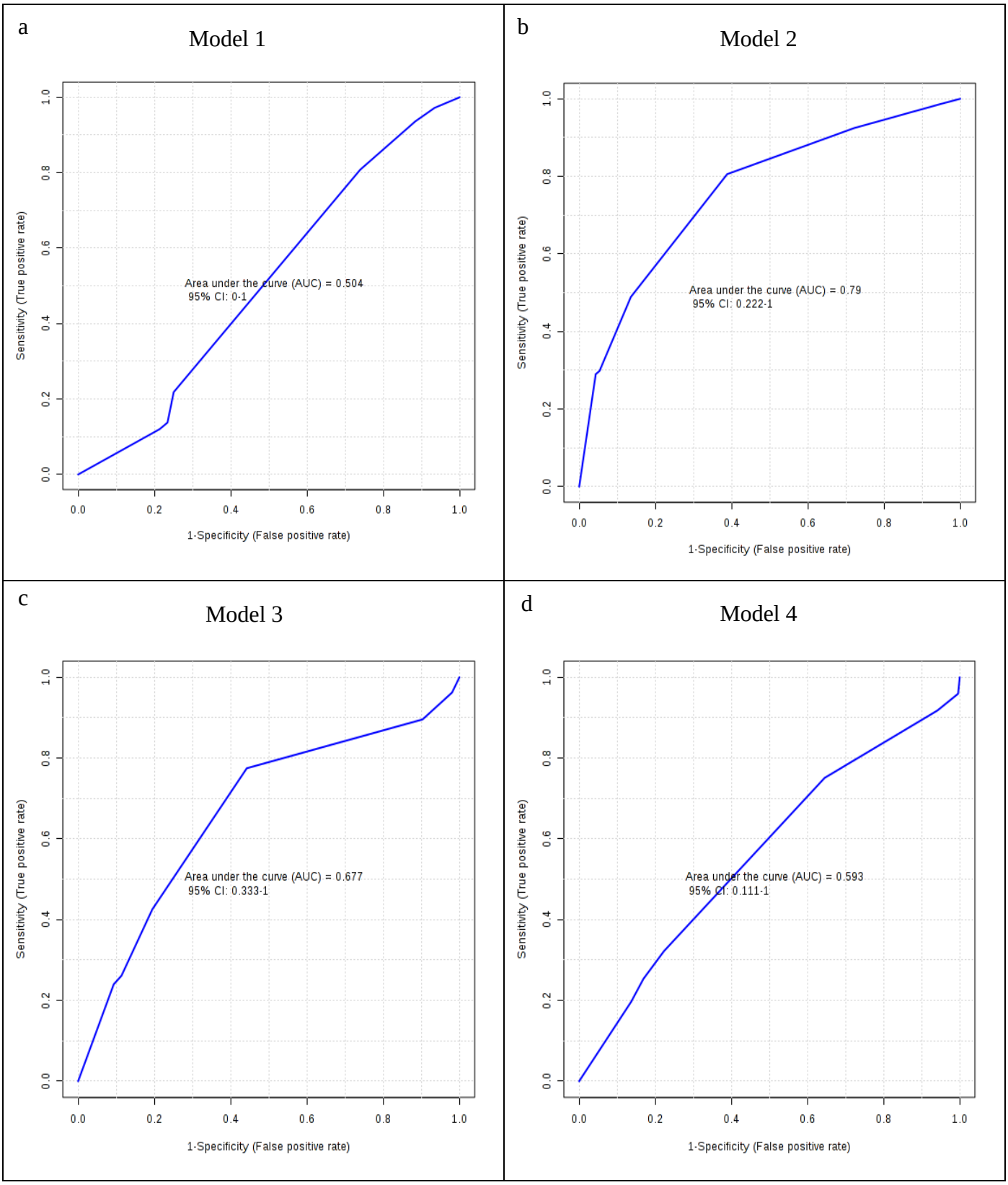
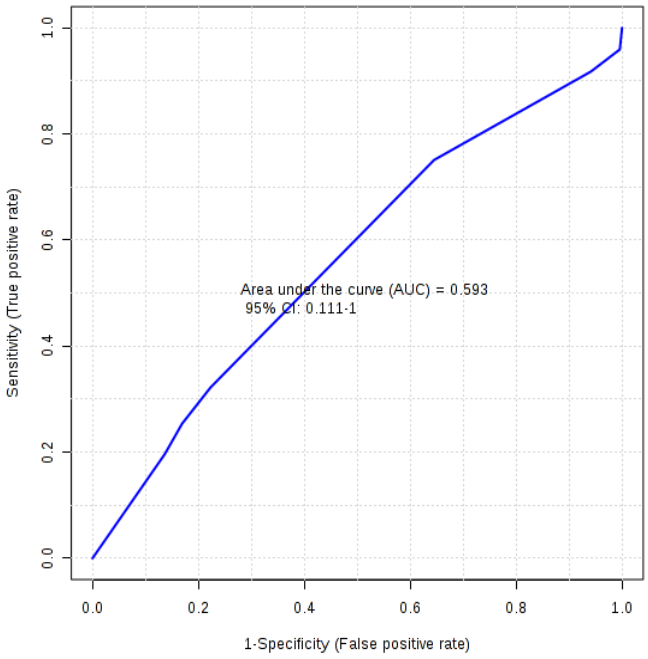


Figure 24: Biomarker evaluation results of acute LD vs Late LD a) ROC curve of model 1 b) ROC curve of model 2 c) ROC curve of model 3 d) ROC curve of model 4

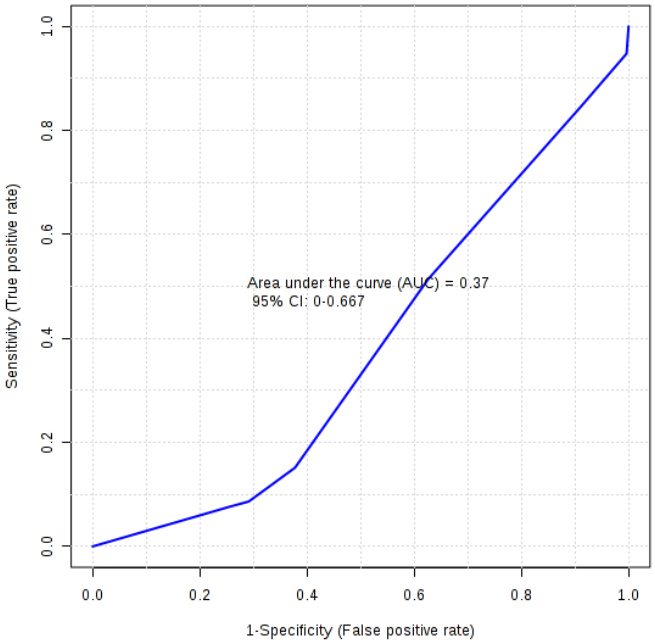
a

Model 4



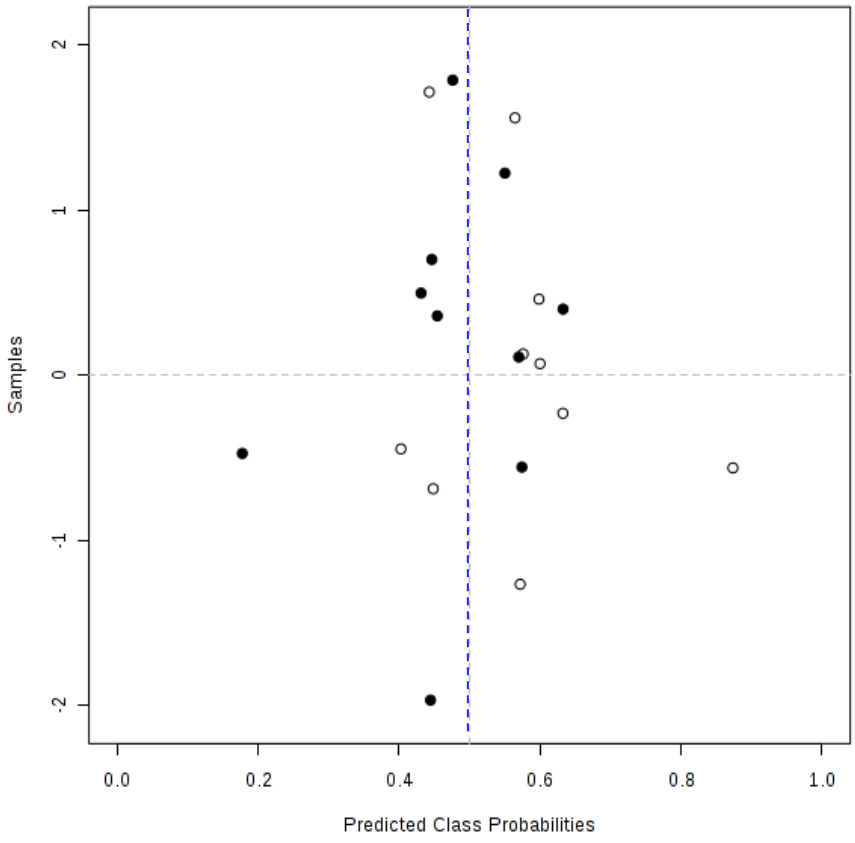
b

Model 5



c

Model 5



Confusion Matrix (Cross-Validation)

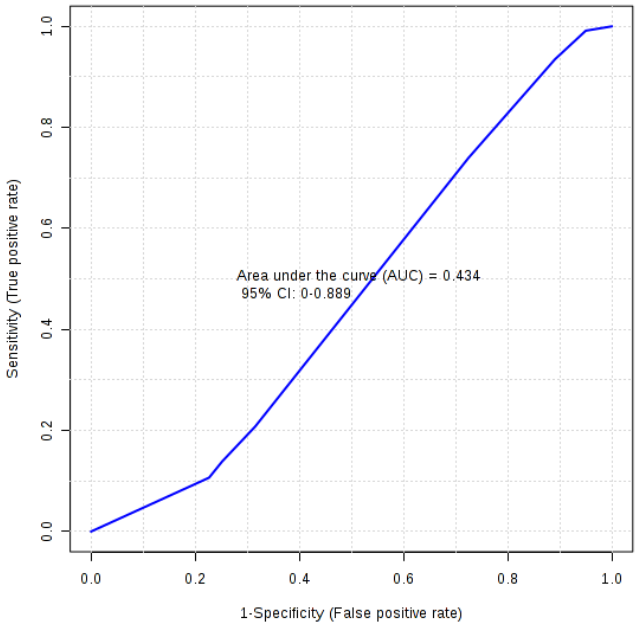
	Acute LD	Late LD
Acute LD	6	3
Late LD	4	7

Acute LD
Late LD

Figure 25: Biomarker Evaluation results of acute LD and late LD a) ROC curve of model 4 b) ROC curve of model 5 c) Class prediction plot using model 5

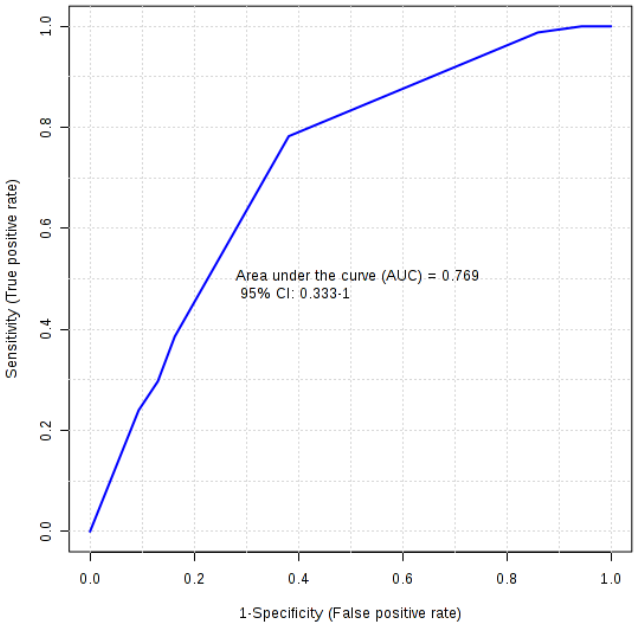
a

Model 6



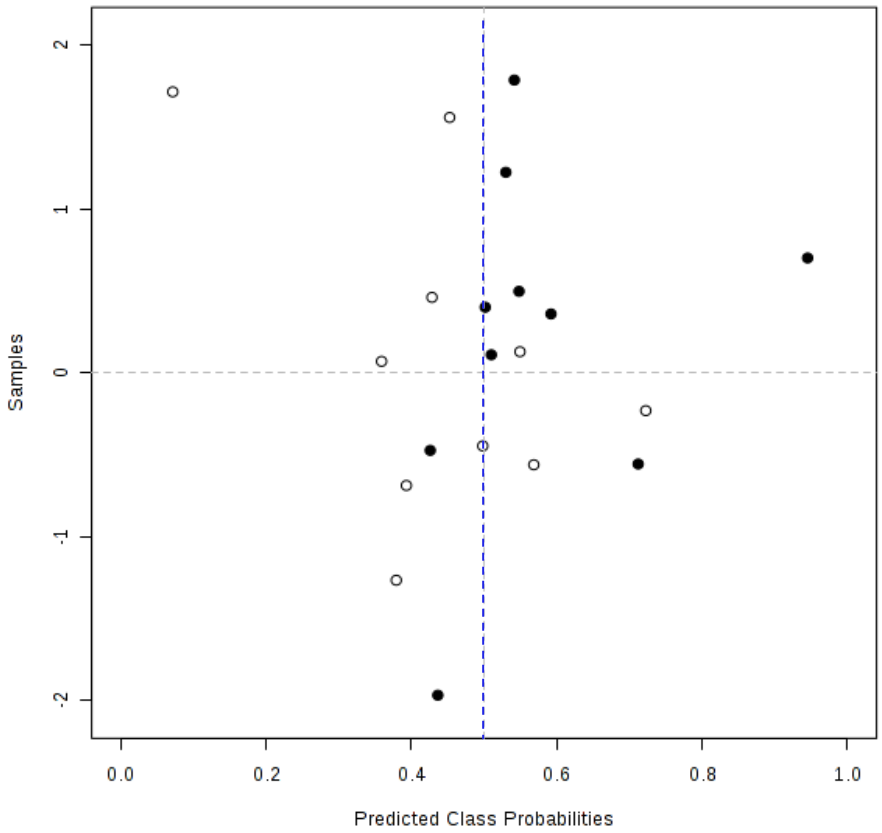
b

Model 7



c

Model 7



Confusion Matrix (Cross-Validation)

	Acute LD	Late LD
Acute LD	8	3
Late LD	2	7

Acute LD
Late LD

Figure 26: Biomarker Evaluation results on Acute and late comparison a) ROC curve of model 6 b) ROC curve of model 7 c) Class prediction plot using model 7

The ROC analysis results of model 2 (AUC 0.98) for acute LD and healthy donor classification and model 3 (AUC 0.98) for late LD and healthy donor classification shows that these two models are excellent models, refer Table 1 for AUC value and corresponding biomarker utility gradings. The sensitivity and specificity of model 2 for acute LD and healthy donor data set is 80 % and 90 % respectively. The sensitivity and specificity of model 3 for late LD and healthy donor data set is 80 % and 100 % respectively. Model 2 was not tested for late LD and healthy donor data set because of the practicalities in MetaboAnalyst software. The ROC analysis results of model 2 (AUC 0.86) with sensitivity 70% and specificity 89 % for acute LD and negative Lyme classification and model 3 (AUC 0.83) for late LD and negative Lyme classification shows that these two models are very good models with sensitivity 80% and specificity 89 %. This difference in the effectiveness of these models for different data sets indicates that one or few of the features used in these models may also be the common features of negative Lyme samples. Model 4 and model 5 were tested for acute LD- negative Lyme and late LD-negative Lyme data sets. Model 5 seem to be working well for all the data sets, particularly for late LD and negative Lyme data set. The sensitivity of the model 5 for late LD and negative Lyme data set in 90 % and specificity is 80 %. The reason why model 5 worked well for late LD and negative Lyme dataset and not equally well for acute LD and negative Lyme dataset is not clear. One convincing reason could be, the large differences in the level of metabolites/features used in model 5 in acute LD and late LD samples. In other words, some or all of the features namely unsaturated fatty acids, choline, valaine, alanine, LDL, unsaturated fatty acids were largely varying in late LD samples compared to negative Lyme samples. But in acute LD samples the difference is not large enough compared to negative Lyme and the class prediction was less effective. Again this strengthens the fact that there are significant metabolic difference between acute LD and late LD samples. Model 1, model 2, model 3, model 4 and model 5 were all tested on acute LD and late LD data sets. The ROC curve analysis results show that none of the models are suitable for acute LD and late LD classification except model 2 for which the AUC value was 0.7 fairly high compared to all other models. Model 6 and model 7 were tested to distinguish acute LD from late LD. The ROC analysis result of model 6 (AUC 0.434) was not impressive. But the ROC analysis result of model 7 (AUC 0.769) showed that it is fair model for distinguish acute LD from late LD samples.

5. DISCUSSIONS

Early Lyme diagnosis is challenging even with the serological based two tier testing method and the assay sensitivity of this method is less than 69 % for early LD [16]. Meta-analytic data summarizing the sensitivity and specificity of various two tier methods and a brief summary on the studies conducted since 1995 in North America on the LD diagnostic tests accuracy, can be acquired from the review article by Waddell and collegues [16]. As discussed in this article by Wadell and collegues, Immunetics the commercially available C6 B.burgdorferi ELISA™ is so far the effective diagnostic test but still the low assay sensitivity of the test in patients with early LD remain to the be limitation [16]. So there is an immense need for more sensitive early LD diagnosis technique. Lyme pathogens interacts with host cells and interfere in cellular functions

which results in fluctuations in the level of metabolites [56–58] and so we hypothesize that, by detecting the metabolic changes in patients, LD patients can be possibly identified and it is supported by our results, refer Figure 12 a, 12b, 13a, 13 b and Table 6. In other words LD patients can be clinical diagnosed on the basis of metabolites as biomarkers, similar to other diseases namely cancer, hepatocellular necrosis, etc.,[29–33] [refer Table 2 for some examples metabolite biomarkers], as demonstrated here.

Using 1D proton NMR spectroscopy technique, we identified 24 metabolites in human sera as presented in Figure 11 and Table 5. In another metabolomic study by Gao and collegues, eighteen metabolites were detected using the 1D proton NMR spectroscopy technique in human hepatocellular carcinoma and liver cirrhosis patients [59]. Higher levels of metabolites such as pyruvate, glutamine, tyrosine, 1-methylhistidine and phenylalanine, etc., in human hepatocellular carcinoma and liver cirrhosis patients compared to healthy donors, together with lower levels of low-density lipoprotein, isoleucine, valine, creatine, choline and unsaturated lipids compared to healthy donors were reported [59]. Although there can be over 4000 metabolites in human serum including lipids, it is nearly impossible to detect and identify the proton peaks corresponding to all of those metabolites. The major reasons for this incapability are, concentration of each metabolite in sera and concentration of the sera itself. The attainable S/N in 1D proton NMR spectroscopy is proportional to concentration of the serum. Another reason is spectral overlap, which stems from metabolites which have protons that resonate at nearby frequency range. Owing to finite linewidth of proton resonances and different amount of metabolites, this often results in broadened resonance clumps that are difficult to deconvolute and quantify. If we decide to overcome the sensitivity problem by increasing the concentration of the sera, then other essential factors such as magnetic field homogeneity, ionic concentration, etc., that largely influence the spectra might be disturbed.

Though 24 metabolites were identified using online databases (BMRB, HMDB) and literatures, only 15 were chosen for further analysis. The metabolites were chosen mainly based on the lineshape of the respective peaks and spectral overlaps with neighboring peaks to avoid errors and misconceptions. The result of one sample in negative Lyme category [sample number D3, Refer appendix I] was excluded from the further analysis because the spectrum obtained was abnormal (data not shown). The prime metabolic change compared to the metabolic profile of the healthy donor samples are high level of lactate in both acute LD and late LD, as shown Figure 12, 13 and Table 6. It is known that, in critically ill patients such as cancer patients, patients with severe respiratory illnesses, etc., higher lactate level in blood is usually due to hypoxemia condition (low level of oxygen in blood), inadequate perfusion and deficiency of enzymes responsible for lactate metabolism such as pyruvate dehydrogenase deficiency, glucose-6-phosphate dehydrogenase deficiency, etc., [60]. One or few of these mentioned reasons could be the reson for increased level of lactate in LD patients sera samples compared to healthy donors sera samples. In sepsis patients, raised blood lactate level can be observed even in the absence of tissue hypoxia conditions [60]. Sepsis is the condition in which extreme inflammatory response to infection causing inflammation throughout the body and inappropriate blood clots in vital organs resulting in

organ failure [61,62]. Notably, the severity of the late LD also has a strong association with host inflammatory responses [63]. All acute and late LD samples with increased level of lactate shown lower level of glucose, suggesting an inverse proportionality of these two features. An elevated lactate level, together with lower or negligible glucose level, as presented in Figure 12 a could be important characteristics of LD patients as it one of the important difference among the observed. Metabolites such as phenylalanine, tyrosine did not show any consistent increase or decrease in LD sample population, refer Figure 13 b. Various metabolic pathways have to be studied to figure out the most convincing reason for the metabolic changes that are presented in Figure 12, 13 and Table 6, such as elevated level of citric acids in 60 % and 80 % of the acute and late LD samples, decreased or lower level of amino acids such as alanine, glycine, leucine, creatine, choline, glutamine. Molins and collegues have reported the lipophilic structures as biosignatures of LD patients [18]. The characterization techniques such as LC-MS can be used to detect the lipidic structures and associated biomarkers in the studied LD samples. We aimed for other small molecule metabolities in this study.

MetaboAnalyst software used for biomarker analysis is a user friendly software. To use this software one does not have to have a strong statistical background [47]. The biomarker analysis results of 34 feature models and some of the 5 feature models demonstrated that acute and late LD samples can be distinguished from healthy donor and negative Lyme samples, based on the detected metabolic features namely glucose, lactate, citric acid, alanine, valine, leucine, choline, glutamate, glutamine, glycine, creatine, LDL, VLDL and unsaturated fatty acids, refer section 4.2. The 34 feature model which employs the ratio of individual metabolites seem to be an excellent model [Refer Table 1 for AUC values and utility of biomarkers based on AUC] with AUC values between 0.9 to 1.0 as presented in Figures 14 a, 15 a, 16 a and 17 a . The sensitivity of the 34 feature model is greater than or equal to 80 % for both acute LD and late LD samples, from the Figures 14 a, 14 b, 15 a, 15 b, 16 a, 16 b, 17 a and 17 b. This is comparatively greater than the sensitivity of the serological based methods used for acute LD diagnosis [16,64]. The specificity of the 34 feature model is 89 - 100 % for both acute LD and late LD as presented in Figures 14 a, 14 b, 15 a , 15 b, 16 a, 16 b, 17 a and 17 b. From the biomarker analysis results presented in section 4.2, identification of acute LD samples and late LD samples based on the metabolites is evident and it is also supported by the class-seperation plots and cross-validation matrices in Figures 14 b, 15 b, 16 b and 17 b. For practical reasons, simple five feature models were tested. The number of combinations of 34 features as a set of 5 is 278,256 and testing all of these models in the MetaboAnalyst will take a long time. So features were selected for testing were based on assumptions and different criterions, as mentioned in section 3.3. The ROC curve analysis of five feature model results, indicated that model 1 is a fair model, model 2, 3 and 5 are also excellent models similar to 34 feature model, as presened in Figure 18 b and c, 20 b and c. Thus, the possibility of using some or all of the metabolic features reported in this study as potential biomarkers for early LD and late LD diagnosis is undeniable.

6. CONCLUSION

Lyme disease is an endemic disease in various parts of the world and an emerging threat in some of the south Asian countries too. LD diagnosis is challenging because of its complexity and low assay sensitivity of the two tier current diagnostic methods (TTTAs). The reported work is a step towards simple and potential diagnosis of LD based on metabolites. 1D proton NMR spectroscopy is highly suitable sophisticated technique for semi-quantitative metabolomic study. With this technique whole metabolic profile of a person can be acquired from the body fluids. Though identification of features is challenging especially when we are looking for a molecular compound that have not been reported, still NMR proves to be a powerful technique. From the results of this metabolomic study, it is evident that acute and late LD patients can be identified based on their metabolic profile. The biomarker analysis using 34 features and 5 features models further suggests that acute LD and late LD samples can be distinguished from healthy donors and negative Lyme samples with sensitivity 80 % and specificity 90 – 100 %. Our future goal is to detect and identify more metabolites present in LD patient sera using the NMR spectroscopy and LC-MS. The studies about the use of metabolites as biomarkers could be a game changer in the Lyme disease diagnostics associated challenges and could draw an end line to the complexity and limitations in LD diagnosis.

REFERENCES

[1] K. Tilly, P.A. Rosa, P.E. Stewart, Biology of Infection with Borrelia burgdorferi, 22 (2008) 217–234.

[2] L.C. Kingry, D. Batra, A. Replogle, L.A. Rowe, B.S. Pritt, J.M. Petersen, Whole genome sequence and comparative genomics of the novel lyme borreliosis causing pathogen, Borrelia mayonii, PLoS One. 11 (2016) 1–21.

[3] H.J. Christen, F. Hanefeld, H. Eiffert, R. Thomssen, Epidemiology and clinical manifestations of Lyme borreliosis in childhood. A prospective multicentre study with special regard to neuroborreliosis, Acta Paediatr.Suppl. 386 (1993) 1–75.

[4] M. Leena, Characterization and Immunological Aspects of Borrelia Burgdorferi Pleomorphic Round Bodies Leena Meriläinen Characterization and Immunological Aspects of Borrelia Burgdorferi Pleomorphic Round Bodies, n.d.

[5] E.S. Theel, LYME DISEASE, 54 (2016) 1191–1196. doi:10.1128/JCM.03394-15.Editor.

[6] S. Nayak, A. Sridhara, R. Melo, L. Richer, N.H. Chee, J. Kim, V. Linder, D. Steinmiller, S.K. Sia, M. Gomes-Solecki, Microfluidics-based point-of-care test for serodiagnosis of Lyme Disease, Sci. Rep. 6 (2016) 1–9.

[7] E.L. Maloney, The Need for Clinical Judgment in the Diagnosis and Treatment of Lyme Disease, J. Am. Physicians Surg. 14 (2009) 82–89.

[8] E.D. Shapiro, Lyme Disease, N. Engl. J. Med. 370 (2014) 1724–1731. doi:10.1056/NEJMcp1314325.

[9] A.E. Fish, Y.B. Pride, D.S. Pinto, Lyme Carditis, Infect. Dis. Clin. North Am. 22 (2008) 275–288.

[10] J.M. Siwula, G. Mathieu, Acute onset of facial nerve palsy associated with Lyme disease in a 6 year-old child., Pediatr. Dent. 24 (2002) 572–574.

[11] a P. van Dam, H. Kuiper, K. Vos, a Widjojokusumo, B.M. de Jongh, L. Spanjaard, a C. Ramselaar, M.D. Kramer, J. Dankert, Different genospecies of Borrelia burgdorferi are associated with distinct clinical manifestations of Lyme borreliosis., Clin. Infect. Dis. 17 (1993) 708–717.

[12] R. Rosenberg, N.P. Lindsey, M. Fischer, C.J. Gregory, A.F. Hinckley, P.S. Mead, G. Paz-Bailey, S.H. Waterman, N.A. Drexler, G.J. Kersh, H. Hooks, S.K. Partridge, S.N. Visser, C.B. Beard, L.R. Petersen, Vital Signs: Trends in Reported Vectorborne Disease Cases - United States and Territories, 2004-2016., MMWR. Morb. Mortal. Wkly. Rep. 67 (2018) 496–501.

[13] A.T. Borchers, C.L. Keen, A.C. Huntley, M.E. Gershwin, Lyme disease : A rigorous review of

diagnostic criteria and treatment, J. Autoimmun. 57 (2015) 82–115.

[14] N. Rudenko, M. Golovchenko, L. Grubhoffer, J.H. Oliver, Updates on borrelia burgdorferi sensu lato complex with respect to public health, Ticks Tick. Borne. Dis. 2 (2011) 123–128.

[15] D.J. Cameron, L.B. Johnson, E.L. Maloney, Evidence assessments and guideline recommendations in Lyme disease: The clinical management of known tick bites, erythema migrans rashes and persistent disease, Expert Rev. Anti. Infect. Ther. 12 (2014) 1103–1135.

[16] L.A. Waddell, J. Greig, M. Mascarenhas, S. Harding, R. Lindsay, N. Ogden, The accuracy of diagnostic tests for lyme disease in humans, a systematic review and meta-analysis of north American research, PLoS One. 11 (2016) 1–23.

[17] L.J. Lahey, M.W. Panas, R. Mao, M. Delanoy, J.J. Flanagan, S.R. Binder, A.W. Rebman, J.G. Montoya, M.J. Soloski, A.C. Steere, R.J. Dattwyler, P.M. Arnaboldi, J.N. Aucott, W.H. Robinson, Development of a Multiantigen Panel for Improved Detection of Borrelia burgdorferi Infection in Early Lyme Disease, J Clin Microbiol. 53 (2015) 3834–3841.

[18] C.R. Molins, L. V. Ashton, G.P. Wormser, A.M. Hess, M.J. Delorey, S. Mahapatra, M.E. Schriefer, J.T. Belisle, Development of a metabolic biosignature for detection of early lyme disease, Clin. Infect. Dis. 60 (2015) 1767–1775.

[19] O. Beckonert, H.C. Keun, T.M.D. Ebbels, J. Bundy, E. Holmes, J.C. Lindon, J.K. Nicholson, Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts., Nat. Protoc. 2 (2007) 2692–2703.

[20] A.C. Dona, M. Kyriakides, F. Scott, E.A. Shephard, D. Varshavi, K. Veselkov, J.R. Everett, A guide to the identi fi cation of metabolites in NMR-based metabonomics / metabolomics experiments, CSBJ. 14 (2016) 135–153.

[21] S.K. Bharti, R. Roy, Quantitative1H NMR spectroscopy, TrAC - Trends Anal. Chem. 35 (2012) 5–26.

[22] A.C. Dona, B. Jime, M.R. Lewis, J.T.M. Pearce, E. Holmes, J.C. Lindon, J.K. Nicholson, Precision High-Throughput Proton NMR Spectroscopy of Human Urine, Serum, and Plasma for Large-Scale Metabolic Phenotyping, (2014).

[23] A.H. Emwas, R. Roy, R.T. McKay, D. Ryan, L. Brennan, L. Tenori, C. Luchinat, X. Gao, A.C. Zeri, G.A.N. Gowda, D. Raftery, C. Steinbeck, R.M. Salek, D.S. Wishart, Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis, J. Proteome Res. 15 (2016) 360–373.

[24] G.A. Nagana Gowda, D. Raftery, Recent advances in NMR-based metabolomics, Anal. Chem. 89 (2017) 490–510.

[25] L. Poretsky, Principles of diabetes mellitus, Princ. Diabetes Mellit. (2010) 1–887.

[26] A.J. Atkinson, W.A. Colburn, V.G. DeGruttola, D.L. DeMets, G.J. Downing, D.F. Hoth, J.A. Oates, C.C. Peck, R.T. Schooley, B.A. Spilker, J. Woodcock, S.L. Zeger, Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, Clin. Pharmacol. Ther. 69 (2001) 89–95.

[27] J. Xia, D.I. Broadhurst, M. Wilson, D.S. Wishart, Translational biomarker discovery in clinical metabolomics : an introductory tutorial, (2013) 280–299.

[28] E. Simin, Amino acid profiling as a method of discovering biomarkers for early diagnosis of cancer, (2016) 1339–1345.

[29] F. Valenza, G. Aletti, T. Fossali, G. Chevallard, F. Sacconi, M. Irace, L. Gattinoni, Lactate as a marker of energy failure in critically ill patients: Hypothesis, Crit. Care. 9 (2005) 588–593.

[30] O. Danne, M. Möckel, Choline in acute coronary syndrome: An emerging biomarker with implications for the integrated assessment of plaque vulnerability, Expert Rev. Mol. Diagn. 10 (2010) 159–171.

[31] M.C. Mimmi, N. Finato, G. Pizzolato, C.A. Beltrami, F. Fogolari, A. Corazza, G. Esposito, Absolute quantification of choline-related biomarkers in breast cancer biopsies by liquid chromatography electrospray ionization mass spectrometry, Anal. Cell. Pathol. 36 (2013) 71–83.

[32] A. Ribes, S. Pajares, Creatine as Biomarker 14 ´, (2015) 333–361.

[33] S. Pajares, A. Arias, J. García-Villoria, P. Briones, A. Ribes, Role of creatine as biomarker of mitochondrial diseases, Mol. Genet. Metab. 108 (2013) 119–124.

[34] P. Newsholme, M.M.R. Lima, J. Procopio, T.C. Pithon-Curi, S.Q. Doi, R.B. Bazotte, R. Curi, Glutamine and glutamate as vital metabolites, Brazilian J. Med. Biol. Res. 36 (2003) 153–163.

[35] Y. Miyagi, M. Higashiyama, A. Gochi, M. Akaike, T. Ishikawa, T. Miura, N. Saruki, E. Bando, H. Kimura, F. Imamura, M. Moriyama, I. Ikeda, A. Chiba, F. Oshita, A. Imaizumi, H. Yamamoto, H. Miyano, K. Horimoto, O. Tochikubo, T. Mitsushima, M. Yamakado, N. Okamoto, Plasma free amino acid profiling of five types of cancer patients and its application for early detection, PLoS One. 6 (2011) e24143.

[36] Y. Gu, T. Chen, S. Fu, X. Sun, L. Wang, J. Wang, Y. Lu, S. Ding, G. Ruan, L. Teng, M. Wang, Perioperative dynamics and significance of amino acid profiles in patients with cancer, J. Transl. Med. 13 (2015) 1–14.

[37] H. Joung, S. Hun, J. Ryu, J. Eun, Y. Chul, M. Ki, T. Won, S. Lee, H. Nakamura, N. Nishikata, M. Mori, Y. Noguchi, H. Miyano, K. Young, Lung Cancer The performance of a novel amino acid multivariate index for detecting lung cancer : A case control study in Korea, 90 (2015) 522–527.

[38] Y. Ma, W. Liu, J. Peng, P. Zhang, H. Chen, H. Qin, [Study on specific metabonomic profiling of serum from colorectal cancer patients by gas chromatography-mass spectrometry]., Zhonghua Wei Chang Wai Ke Za Zhi. 12 (2009) 386–390.

[39] J. Maeda, M. Higashiyama, A. Imaizumi, T. Nakayama, H. Yamamoto, T. Daimon, M. Yamakado, F. Imamura, K. Kodama, Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study, BMC Cancer. 10 (2010) 690.

[40] N. Okamoto, Y. Miyagi, A. Chiba, M. Akaike, M. Shiozawa, A. Imaizumi, H. Yamamoto, T. Ando, M. Yamakado, O. Tochikubo, Diagnostic modeling with differences in plasma amino acid profiles between non-cachectic colorectal/breast cancer patients and healthy individuals, Int. J. Med. Sci. 1 (2009) 1–8.

[41] B.C. Batch, K. Hyland, L.P. Svetkey, Branch chain amino acids : biomarkers of health and disease, 17 (2014) 86–89.

[42] Y. Qiu, G. Cai, M. Su, T. Chen, X. Zheng, Y. Xu, Y. Ni, A. Zhao, L.X. Xu, S. Cai, W. Jia, Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS, J Proteome Res. 8 (2009) 4844–4850.

[43] A.B. Leichtle, J.M. Nuoffer, U. Ceglarek, J. Kase, T. Conrad, H. Witzigmann, J. Thiery, G.M. Fiedler, Serum amino acid profiles and their alterations in colorectal cancer, Metabolomics. 8 (2012) 643–653.

[44] A. Trpkovic, I. Resanovic, J. Stanimirovic, D. Radak, S.A. Mousa, D. Cenic-Milosevic, D. Jevremovic, E.R. Isenovic, Oxidized low-density lipoprotein as a biomarker of cardiovascular diseases, Crit. Rev. Clin. Lab. Sci. 52 (2015) 70–85.

[45] K.A. Rye, Biomarkers associated with high-density lipoproteins in atherosclerotic kidney disease, Clin. Exp. Nephrol. 18 (2014) 247–250.

[46] K.E. Corey, M. Lai, L.G. Gelrud, J. Misdraji, L.L. Barlow, H. Zheng, K.L. Andersson, M. Thiim, D.S. Pratt, R.T. Chung, Non-High-Density Lipoprotein Cholesterol as a Biomarker for Nonalcoholic Steatohepatitis, Clin. Gastroenterol. Hepatol. 10 (2012) 651–656.

[47] J. Xia, D.S. Wishart, Web-based inference of biological patterns , functions and pathways from metabolomic data using MetaboAnalyst, Nat. Protoc. 6 (2011) 743–760.

[48] T.L. James, Chapter 1 Fundamentals of NMR, Control. 27 (1998) 1–31.

[49] U. Holzgrabe, B.W.K. Diehl, I. Wawer, NMR spectroscopy in pharmacy, J. Pharm. Biomed. Anal. 17 (1998) 557–616.

[50] J.C. Lindon, J.K. Nicholson, Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics, Annu. Rev. Anal. Chem. 1 (2008) 45–69.

[51] D. The, J. Keeler, Fourier transformation and data processing, (2004).

[52] A. Ross, G. Schlotterbeck, F. Dieterle, H. Senn, NMR Spectroscopy Techniques for Application to Metabonomics, Handb. Metabonomics Metabolomics. (2007) 55–112.

[53] J. Van Brakel, Ceteris paribus laws, Behav. Brain Sci. 15 (1992) 584–585.

[54] J. Trygg, E. Holmes, T. Lundstedt, Chemometrics in metabonomics, J. Proteome Res. 6 (2007) 469–479.

[55] A. Wang, E.A. Gehan, Gene selection for microarray data analysis using principal component analysis, Stat. Med. 24 (2005) 2069–2087.

[56] A. Szczepanski, M.B. Furie, J.L. Benach, B.P. Lane, H.B. Fleit, Interaction between Borrelia burgdorferi and endothelium in vitro, J. Clin. Invest. 85 (1990) 1637–1647.

[57] S.K. Williams, Z.P. Weiner, R.D. Gilmore, Human neuroglial cells internalize Borrelia burgdorferi by coiling phagocytosis mediated by Daam1, PLoS One. 13 (2018) 1–18.

[58] R. Ebady, A.F. Niddam, A.E. Boczula, Y.R. Kim, N. Gupta, T.T. Tang, T. Odisho, H. Zhi, C.A. Simmons, J.T. Skare, T.J. Moriarty, Biomechanics of Borrelia burgdorferi Vascular Interactions, Cell Rep. 16 (2016) 2593–2604.

[59] H. Gao, Q. Lu, X. Liu, H. Cong, L. Zhao, H. Wang, D. Lin, Application of 1H NMR-based metabonomics in the study of metabolic profiling of human hepatocellular carcinoma and liver cirrhosis, Cancer Sci. 100 (2009) 782–785.

[60] C. Higgins, Lactate and lactic acidosis, Radiom. Medicall ApS. (2007) 8.

[61] N. Evangelatos, P. Bauer, M. Reumann, K. Satyamoorthy, H. Lehrach, A. Brand, Metabolomics in sepsis and its impact on public health, Public Health Genomics. 20 (2018) 274–285.

[62] M.C. Scott, Defining and Diagnosing Sepsis, Emerg. Med. Clin. North Am. 35 (2017) 1–9.

[63] A. Szczepanski, J.L. Benach, Lyme borreliosis: host responses to Borrelia burgdorferi, Microbiol Rev.

55 (1991) 21–34.

[64] A.R. Marques, Laboratory Diagnosis of Lyme Disease: Advances and Challenges, Infect. Dis. Clin. North Am. 29 (2015) 295–307.

APPENDIX

I. APPENDIX

The NMR data processing and analysis methods followed in this work is presented in this table.

Steps	**Mode**	**Functions**	**Softwares**
Phase correction	Manual	-	TopSpin 3.2
Spectra calibration	Manual	-	TopSpin 3.2
Baseline correction	Automatic	Polynomial of Degree ABSG	TopSpin 3.2
Line broadening	Automatic	Exponential weighting function	TopSpin 3.2
Peak picking and integration	Manual	-	TopSpin 3.2
Procedures		**Algorithm**	**Softwares**
Data formating	Manual	-	MS-Excel
Statistical analysis	Semi-automatic	PLSDA	MetaboAnalyst

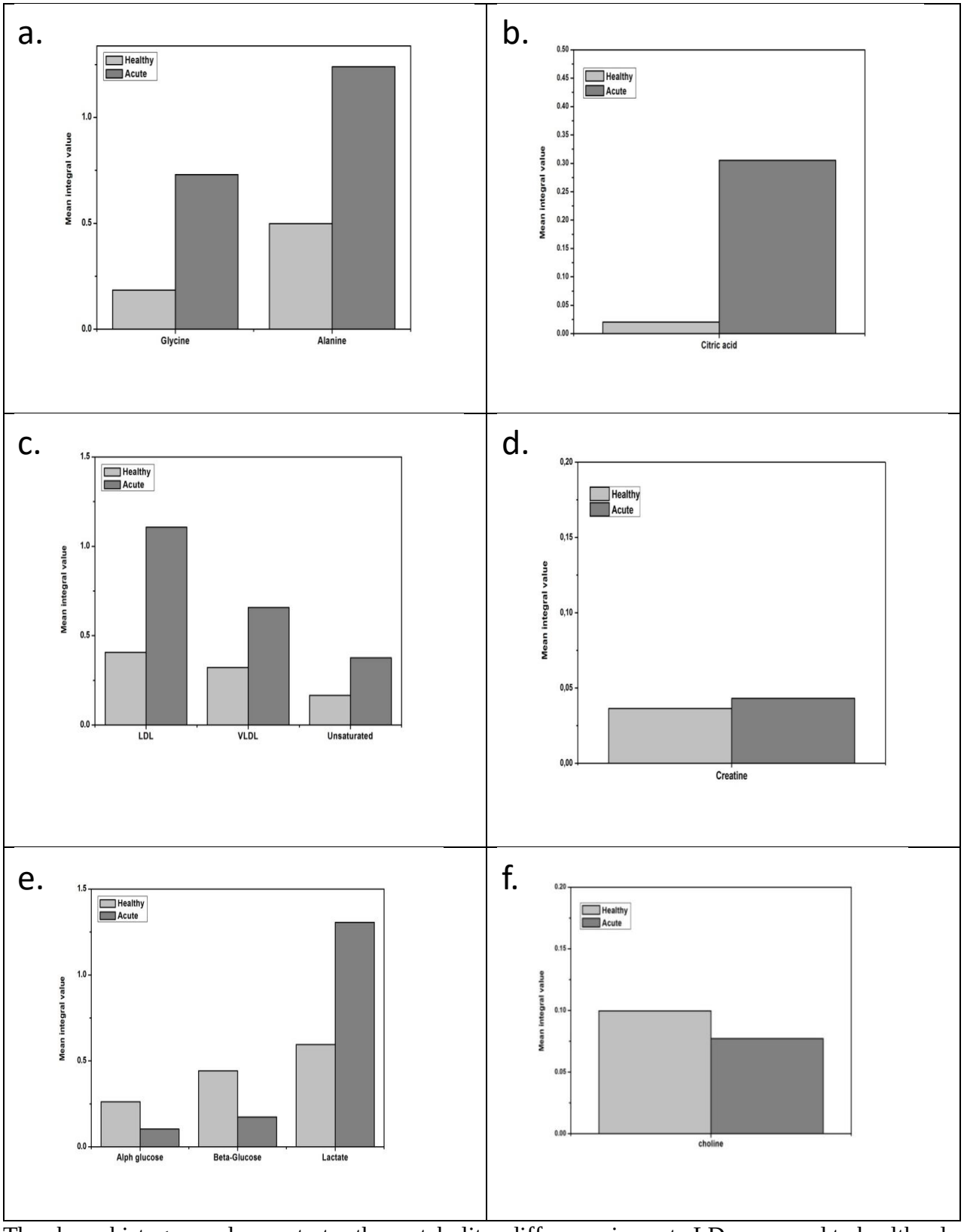
I. APPENDIX

The below table shows the list of samples and the buffer batch used.

Serial number	Sample name	Category	Comment
1	A1	Acute LD	- NB1
2	A2		- NB1
3	A3		- NB1
4	A4		- NB1
5	A5		- NB1
6	A6		- NB1
7	A7		- NB1
8	A8		- NB1
9	A9		- NB1
10	A10		- NB1
11	B1	Late LD	- NB1
12	B2		- NB1
13	B3		- NB1
14	B4		- NB2
15	B5		- NB1
16	B6		- NB2
17	B7		- NB2
18	B8		- NB2
19	B9		- NB2
20	B10		- NB2
21	F1	Healthy donors	- NB2
22	F2		- NB2
23	F3		- NB2
24	F4		- NB2
25	F5		- NB2
26	F6		- NB2
27	F7		- NB2
28	F8		- NB2
29	F9		- NB2
30	F10		- NB2
31	D1	Negative Lyme	- NB3
32	D2		- NB3
33	D3		-NB3(not included in data analysis
34	D4		- NB3
35	D5		- NB3
36	D6		- NB3
37	D7		- NB3
38	D8		- NB3
39	D9		- NB3
40	D10		- NB3

NB1, NB2, NB3- buffer batch numbers.

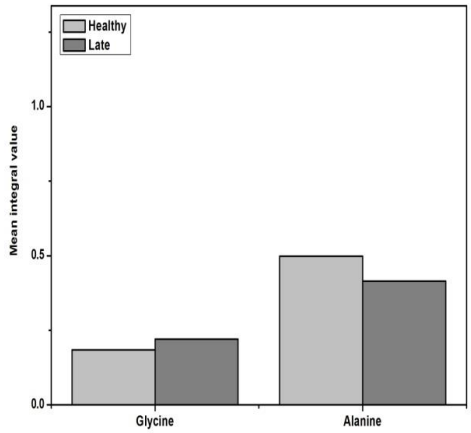
II. APPENDIX



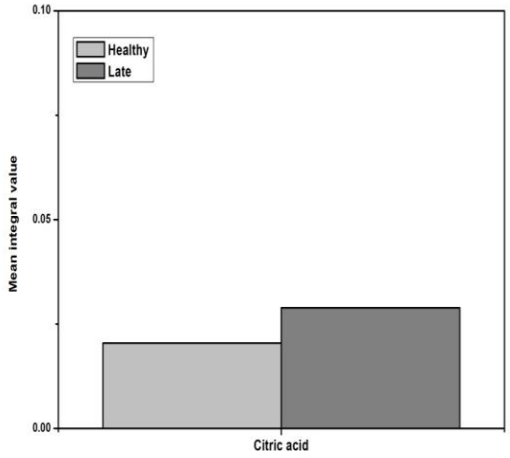
The above histograms demonstrates the metabolites difference in acute LD compared to healthy donors.

II. APPENDIX

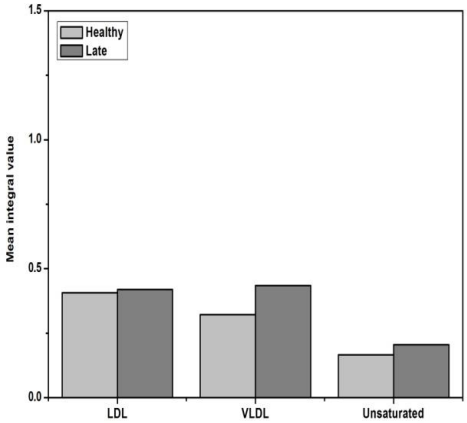
a.



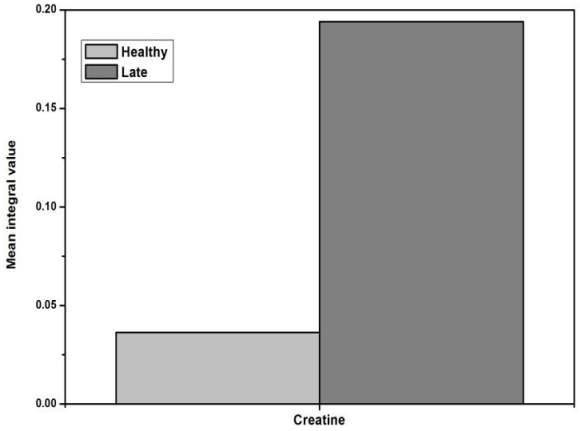
b.



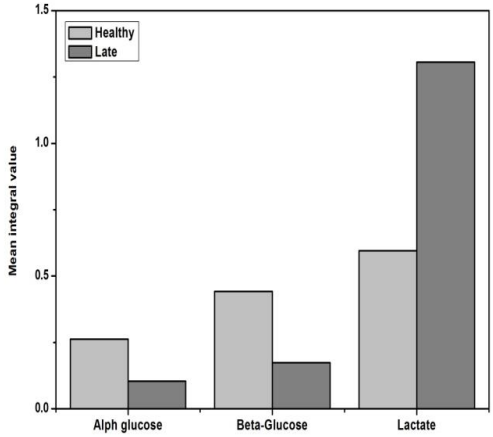
c.



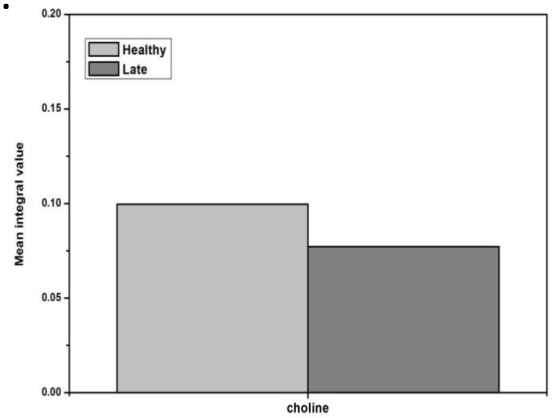
d.



e.



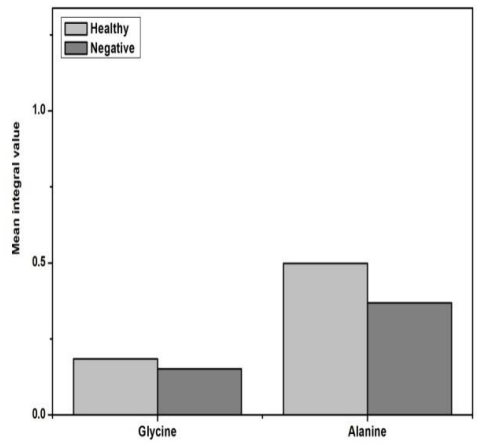
f.



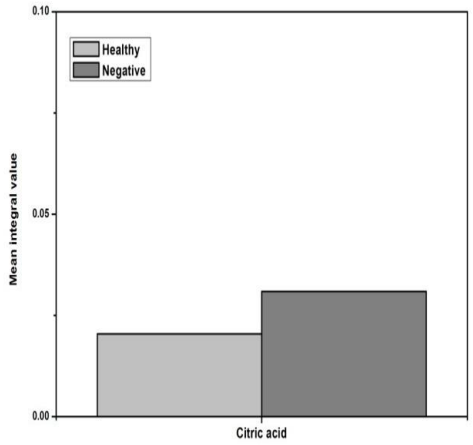
The above histograms demonstrates the metabolic difference in late LD patients compared to healthy donors.

II. APPENDIX

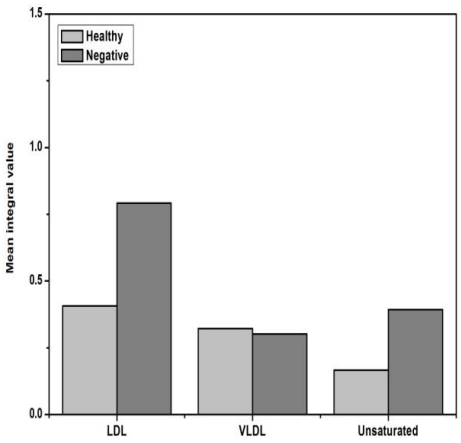
a.



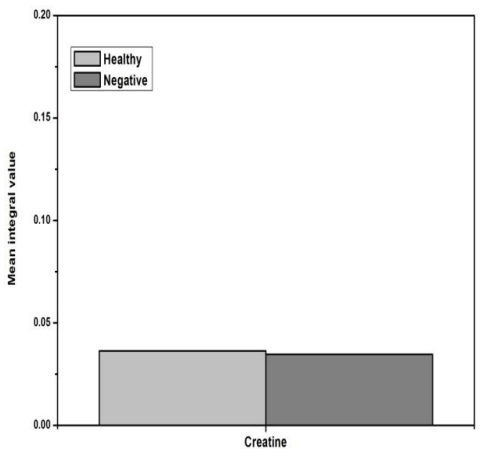
b.



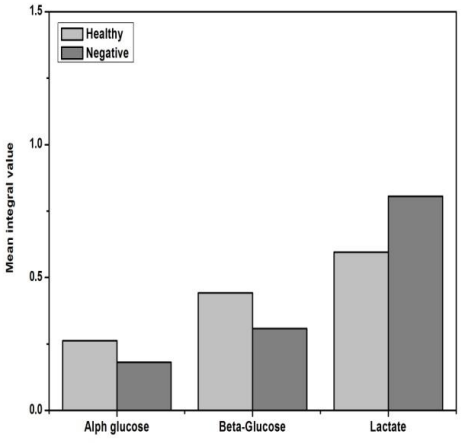
c.



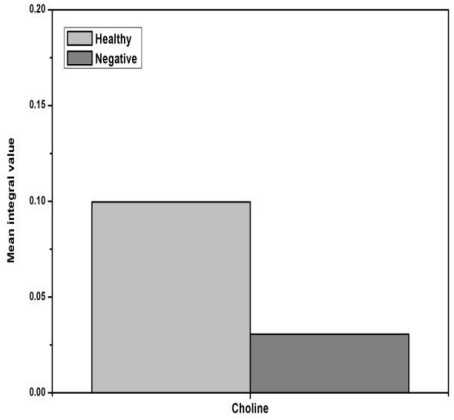
d.



e.

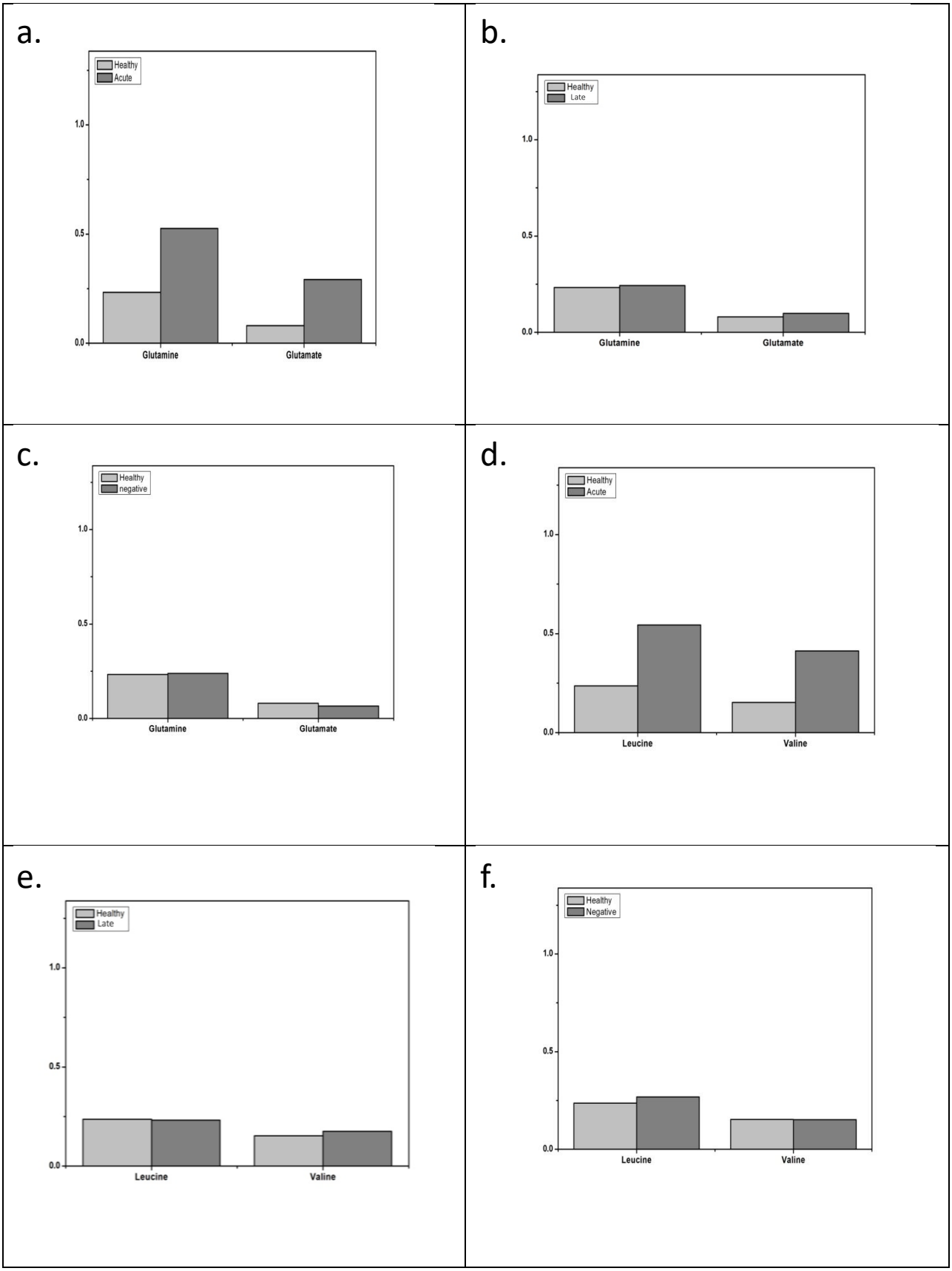


f.



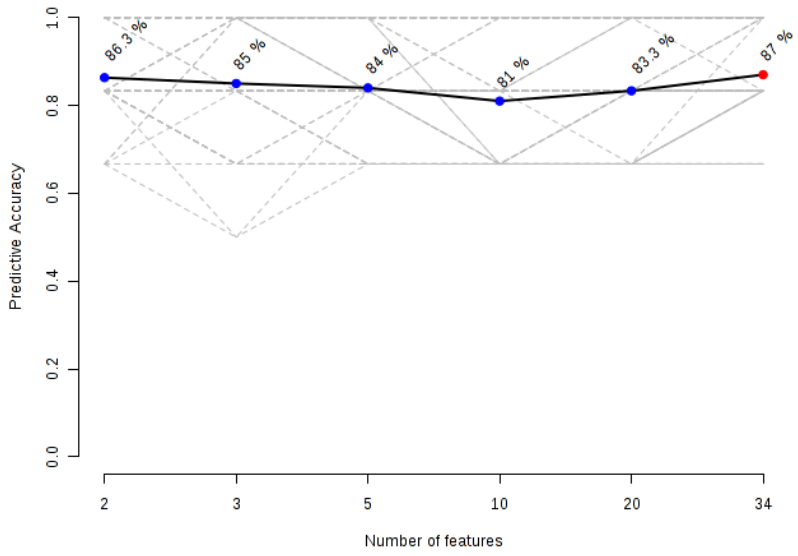
The above histograms demonstrates the metabolic difference in negative lyme and healthy donors.

II. APPENDIX

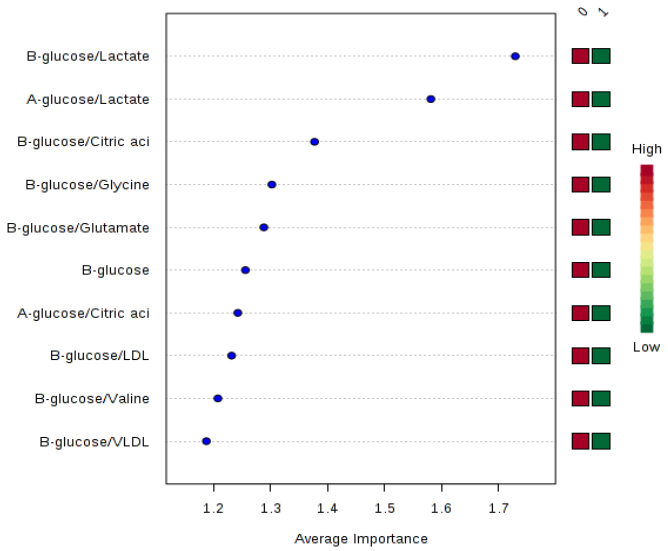


The above histograms demonstrates the difference in the levels of glutamate, glutamine, leucine and valine in acute LD, late LD, negative lyme compared to healthy donors.

III. APPENDIX



Predictive accuracy of different features using- PLS-DA. Number features and their respective predictive accuracy is presented in the above figure. This result will vary according to algorithm used and the latent variables (in case of PLS-DA).



Multivariate Exploratory ROC Analysis. Result of Multivariate Exploratory ROC Analysis tool available on MetaboAnalyst for acute LD and healthy donors is presented above. The image shows the top 10 features ordered based on their average importance. The list of features are specific to the dataset.

IV. APPENDIX

The below table shows the ROC values of all 34 features for healthy donor and acute LD data set.

Features	AUC	Fold Change
Lactate/Alanine	0.99	1.5436
Lactate/Glycine	0.97	0.071083
Lactate/VLDL	0.97	-1.6579
Lactate/Creatine	0.96	-0.11413
Lactate/Glutamine	0.95	0.13759
Lactate/Valine	0.95	0.022535
Lactate/Choline	0.94	-0.14752
Lactate/Leucine	0.94	-1.3877
Lactate/Glutamate	0.93	-0.21441
α glucose/Citric acid	0.9	-0.33907
β glucose/Citric acid	0.9	-0.20773
Lactate/LDL	0.9	1.5104
Unsaturated/Lactate	0.87	0.022535
α glucose	0.86	1.5436
α glucose/Lactate	0.84	1.5436
β glucose	0.83	1.5104
β glucose/Lactate	0.82	-0.07999
Lactate	0.8	-1.6579
α glucose/Glutamine	0.79	-1.6579
β glucose/Glutamine	0.79	-0.32629
α glucose/Creatine	0.78	1.5104
α glucose/Glutamate	0.78	0.17834
Citric acid	0.77	-1.3877
Choline	0.645	0.17834
Alanine	0.62	0.13759
Glutamine	0.6	-0.07999
Glycine	0.59	-0.20773
Unsaturated	0.58	0.022535
VLDL	0.58	0.071083
Creatine	0.56	-0.33907
Leucine	0.54	-0.14752
LDL	0.53	-0.21441
Glutamate	0.5	-0.32629
Valine	0.5	-0.11413

IV. APPENDIX

The below table shows the ROC values of all 34 features for healthy donor and late LD data set.

Features	AUC	Fold Change
α glucose/Glutamine	0.96	2.583
α glucose/Glycine	0.96	-1.2798
β glucose/Glutamine	0.96	-0.00147
β glucose/Glycine	0.96	0.018814
α glucose/VLDL	0.94	-0.21181
Choline/Glycine	0.94	-1.2798
β glucose/VLDL	0.93	2.583
α glucose/Valine	0.92	-0.16996
β glucose/Valine	0.92	-0.27711
β glucose/Leucine	0.9	0.14051
α glucose/Leucine	0.89	0.45362
α glucose/LDL	0.89	0.076048
β glucose/LDL	0.89	-0.448
β glucose/Citric acid	0.89	2.9602
α glucose/Citric acid	0.88	-0.20454
β glucose/Glutamate	0.88	0.10184
α glucose/Glutamate	0.87	2.9602
α glucose	0.86	2.583
β glucose	0.86	2.9602
Lactate	0.86	-1.2798
α glucose/Lactate	0.86	-0.448
β glucose/Lactate	0.84	0.23507
Citric acid	0.695	-0.27711
Choline	0.69	0.45362
Unsaturated	0.68	-0.448
VLDL	0.63	0.14051
Leucine	0.62	0.23507
Creatine	0.6	-0.16996
Glutamate	0.57	-0.21181
Valine	0.57	-0.00147
Glutamine	0.56	0.076048
Alanine	0.55	0.10184
Glycine	0.54	-0.20454
LDL	0.5	0.018814

IV. APPENDIX

The below table shows the ROC values of all 34 features for negative lyme and acute LD data set.

Features	AUC	Fold Change
Choline/Leucine	0.94	-1.8877
Unsaturated/Choline	0.93	5.2627
Choline/Glutamine	0.93	-2.2552
Choline/LDL	0.93	-1.2435
Choline/Valine	0.92	-3.1293
Choline/Alanine	0.9	-0.7361
Choline/VLDL	0.88	-1.8115
Choline	0.81	-0.73959
Choline/Glycine	0.81	-1.0377
Unsaturated/Creatine	0.8	2.1376
Choline/Glutamate	0.8	-3.9849
Unsaturated/VLDL	0.78	4.0673
Unsaturated/Lactate	0.77	5.961
Unsaturated/Glutamine	0.77	4.307
Unsaturated/Glycine	0.77	5.168
Lactate/LDL	0.77	-6.4137
Unsaturated/Valine	0.76	5.0891
Choline/Creatine	0.76	-6.5209
Unsaturated/Glutamate	0.75	3.149
Lactate/Alanine	0.74	-4.4588
Unsaturated	0.67	0.52311
Lactate	0.66	-0.80095
LDL	0.62	0.33594
β glucose	0.6	1.0059
Glutamine	0.55	-1.2361
α glucose	0.54	0.1268
Glutamate	0.54	-1.5318
Creatine	0.53	-1.4068
Glycine	0.53	-1.587
Valine	0.53	-1.3582
VLDL	0.53	-1.0228
Alanine	0.52	-1.9519
Citric acid	0.52	-1.9583
Leucine	0.5	-1.1724

IV. APPENDIX

The below table shows the ROC values of all 34 features for negative lyme and late LD data set.

Features	AUC	Fold Change
Choline/Leucine	1.0	-1.0125
Choline/LDL	1.0	-0.90712
Choline/Valine	0.97778	-1.0399
Unsaturated/Choline	0.95556	1.8365
Choline/Glutamine	0.95556	-0.82473
Glycine/Leucine	0.91111	-4.0485
Choline/Citric acid	0.9	-6.759
Alpha-glucose/Choline	0.88	4.8882
Choline/Glycine	0.87	-0.52405
Choline/Alanine	0.86	-0.48452
Creatine/Leucine	0.84	-0.35369
Glutamate/Leucine	0.84	-0.8569
Alanine/LDL	0.84	-3.4339
Choline	0.83	-0.39873
Choline/VLDL	0.83	-0.76079
Glycine/LDL	0.83	-1.3487
Glycine/Valine	0.82	-2.1915
Leucine/Alanine	0.82	1.3313.
Glutamate/LDL	0.81	-0.66321
Choline/Creatine	0.8	-5.0862
Glycine	0.72	-0.29132
Lactate	0.7	-2.7058
LDL	0.68	1.9866.
Glutamate	0.68	-0.18336
Unsaturated	0.61	1.3164
Creatine	0.61	-0.072426
Leucine	0.58	0.24551
Alanine	0.57	-0.14919
β glucose	0.56	0.7037
VLDL	0.54	-0.56266
α glucose	0.52	0.41867
Valine	0.52	-0.064572
Citric acid	0.51	0.028861
Glutamine	0.5	-0.0020669

IV. APPENDIX

The below table shows the ROC values of all 34 features for acute LD and late LD data set.

Features	AUC	Fold Change
Leucine/Valine	0.83	1.566
β glucose	0.79	1.566
α glucose/ β glucose	0.78	0.001311
Leucine/Alanine	0.77	0.37716
Unsaturated/ β glucose	0.76	1.1805
Unsaturated/LDL	0.76	0.15699
Unsaturated/Leucine	0.75	0.15917
Glycine/Leucine	0.75	-0.47573
Glycine/Citric acid	0.73	1.1805
Unsaturated/Lactate	0.71	1.566
α glucose	0.69	1.1805
Unsaturated/ α glucose	0.69	-0.47573
β glucose/Valine	0.69	0.23334
β glucose/Alanine	0.69	0.070927
β glucose/Glutamine	0.68	0.38529
Unsaturated/Glutamine	0.68	0.29418
Unsaturated/Citric acid	0.68	0.10914
β glucose/Glycine	0.67	-0.03581
β glucose/Glutamate	0.66	0.11474
Unsaturated/Creatine	0.66	0.37716
Glycine	0.65	0.001311
β glucose/VLDL	0.65	1.0788
Unsaturated	0.6	-0.47573
Creatine	0.6	0.15917
Citric acid	0.58	1.0788
Alanine	0.57	-0.03581
Leucine	0.57	0.38529
Glutamate	0.53	0.10914
Glutamine	0.52	0.15699
Valine	0.52	0.11474
LDL	0.52	0.23334
VLDL	0.51	0.070927
Lactate	0.5	0.37716
Choline	0.5	0.29418