METABOLOMIC PROFILING OF HUMAN EMBRYO DURING PRE-IMPLANTATION IN VITRO FERTILIZATION NON-INVASIVE APPROACH

A Thesis

by

MARYAM AL SHAIKH

BS, Texas A&M University-Corpus Christi, 2016

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

CHEMISTRY

Texas A&M University-Corpus Christi
Corpus Christi, Texas

May 2021

© Maryam Eissa Al Shaikh

All Rights Reserved

May 2021

METABOLOMIC PROFILING OF HUMAN EMBRYO DURING PRE-IMPLANTATION IN VITRO FERTILIZATION NON-INVASIVE APPROACH

A Thesis

by

MARYAM AL SHAIKH

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Hussain Abdulla, PhD
Chair

Fereshteh Billiot, PhD
Co-Chair/Committee Member

Xavier Gonzales, PhD
Committee Member

May 2021

ABSTRACT

In vitro fertilization (IVF) is a standard protocol used to treat infertility. However, the probability of successful embryo implantation during IVF is very low. Most of the IVF clinics depend on morphological scoring by embryologists to select high-quality embryos capable of implantation. But morphological scoring has only around 30% successful pregnancy rate. In this study, I investigated the potential of a new embryo scoring method based on measuring the change in culture media's metabolomic profiles. I analyzed 71 culture media samples with known pregnancy outcomes from two different culture media by ultraperformance liquid chromatography (UPLC) coupled with ultrahigh-resolution and accuracy mass spectrometer. I used a newly developed on-the-fly dynamic data acquisition technique to increase the percentage of metabolite compounds with MS^2 fragmentation spectra. To identify potential metabolomic pregnancy biomarkers, we used a combination of statistical analysis techniques like principal component analysis (PCA), differential analysis (volcano plots), and trend charts. We used Molecular Formula Calculator software, ChemSpider, and mzCloud databases to assign the molecule formula and chemical structure for the detected significant biomarkers. Also, we applied in-silico fragmentation and FISh scoring to validate the chemical structures of the identified biomarkers. Using PCA, we did not find any apparent clustering for pregnant or non-pregnant samples, but we could locate a few outliers' spectra. However, with volcano plots, we were able to identify a set of up-regulated biomarkers that are associated with non-pregnancy and down-regulated biomarkers that are associated with pregnancy in both media. Utilizing the KEGG and Metabolika databases, we recognized two possible metabolomics pathways. This study can improve selecting viable embryos, which will lead to an increase in the success rate of IVF. It will also provide a better

understanding of human embryos' metabolomic biochemical pathways during the pre-implantation stage.

DEDICATION

I dedicate this thesis study to my supporting family. Starting with my father, Eissa Alshaikh, and my mother, Ganiah Alsalh, who encouraged me at the age of 19 years old to study abroad and who supported me all the way. It is my life's dream to make them proud of me.

I also dedicate this work to my beloved husband, who stood by me and struggled for my success. I am very grateful to him.

Finally, I dedicate this work to my siblings. I could not have done it without their help and support.

ACKNOWLEDGEMENTS

First, I would like to acknowledge and thank Dr. Abdulla, my committee chair. Dr. Abdulla is an inspiration for me and for any student who gets to know him. He works hard to teach us and monitor our research. He always encourages me to read more research papers and learn about up-to-date methods and techniques related to untargeted metabolomic biomarkers discovery. Dr. Abdulla also encourages us to ask questions and get involved in the class and research discussions. I also would like to acknowledge and thank Dr. Billiot, one of my committee members. Dr. Billiot helped me through my graduate experience, especially with the most harrowing moments. She was an outstanding advisor and always motivated me to keep going regardless of the circumstances. Also, I would like to thank Dr. Gonzales, who is another committee member. I am very grateful for Dr. Gonzales's immediate help and response, especially when I was out of the country due to the COVID-19 pandemic.

In addition, I would like to thank Overture Life for providing me the samples for my project and for the opportunity to work with them. I would also like to thank my colleague Daniela Bergmann for helping me stay connected with the laboratory computer when I returned to my country. I appreciate her help and time, and I wish her the best in her academic future. Finally, I would like to thank all colleagues and the physical and engineering science department faculty and staff at Texas A&M University-Corpus Christi. This research was financially supported by Welch Departmental grants and by Overture life Inc. company.

TABLE OF CONTENTS

CONTENTS PAGE

ABSTRACT iv

DEDICATION vi

ACKNOWLEDGEMENTS vii

TABLE OF CONTENTS viii

LIST OF FIGURES xii

CHAPTER I. INTRODUCTION 1

1.1 BACKGROUND 1

1.2 RELATED WORK 4

CHAPTER II: METHODS & PROCEDURES 7

2.1 SAMPLE PREPARATION 7

2.2 METABOLOMIC ANALYSIS BY UPLC-ORBITRAP 7

2.3 STATICAL ANALYSIS 9

2.4 MOLECULAR FORMULA CALCULATOR 9

CHAPTER III: RESULTS & DISCUSSION 10

3.1 PRINCIPLE COMPONENTS ANALYSIS 11

3.2 DIFFERENTIAL ANALYSIS (VOLCANO PLOTS) 14

3.2.1 MOLECULAR FORMULA &PYTHON CODE 18

3.3 BOX-WHISKER CHART 19

3.4 STRUCTURE 20

3.2.3 METABOLOMIC PATHWAY.....22

CHAPTER IV: CONCLUSION 27

REFERENCES 29

LIST OF APPENDICES.....33

APPENDICES 35

Appendix 1: M2 M2 samples pregnant compounds that we can identify using MF calculator and python code.....35

Appendix 2: M2M2 samples nonpregnant compounds that we're able to identify using MF calculator and python code, compounds that are highlighted are matching Compound Discoverer results.....36

Appendix 3: M11 samples pregnant compounds before removing the outlier that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....37

Appendix 4: M11 samples nonpregnant compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....38

Appendix 5: M11 samples pregnant compounds after removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....39

Appendix 6: : M11 samples nonpregnant compounds After removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....40
Appendix 7: All samples Down-regulated compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....41
Appendix 8: All samples UP regulated compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....42
Appendix 9: All samples Down-regulated compounds after removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....43
Appendix 10: All samples UP-regulated compounds after removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....44
Appendix 11: Structures for up-regulated compounds M2M2 samples.....45
Appendix 12: Structures for down-regulated compounds M2M2 samples.....46
Appendix 13: Structures for up-regulated compounds M11 samples.....47
Appendix 14: Structures for down-regulated compounds M12 samples.....48
Appendix 15: Structures for up-regulated compounds all samples.....50
Appendix 16: Structures for down-regulated compounds all samples.....51
Appendix 17: FISh spectra for two identified nonpregnant compounds in M2M2 samples.....52
Appendix 18: FISh spectra for two identified pregnant compounds in M2M2 samples.....53

Appendix 19: FISh spectra for two identified nonpregnant compounds in M11 samples.....54

Appendix 20: FISh spectra for two identified pregnant compounds in M11 samples.....55

Appendix 21: : FISh spectra for two identified nonpregnant compounds in all samples.....56

LIST OF FIGURES

FIGURES PAGE

Figure 1: Multiple birth rate (delivery of two or more children) in IVF/ ICSI over 14 years in Europe.....3

Figure 2: Diagram of Omics Cascade. Adopted from Dettmer (2007).....4

Figure 3: a) Principal component analysis (PCA) score plot of the M2M2 samples (n= 36) of 4300 metabolomic compounds detected by UPLC-OT-FTMS b) PCA Loading plots of M2M2's 4300 metabolomic compounds.....12

Figure 4: a) Principal component analysis (PCA) score plot of the M11 samples (n= 29) of 2800 metabolomic compounds detected by UPLC-OT-FTMS B) PCA Loading plots of M1's 2800 metabolomic compounds.....13

Figure 5: a) Principal component analysis (PCA) score plot of the M11 samples (n= 29; removing three outlier samples) of 2800 metabolomic compounds detected by UPLC-OT-FTMS b) PCA Loading plots of M1's 2800 metabolomic compounds.....13

Figure 6: A) Principal component analysis (PCA) score plot of both M1M1 and M22 samples (n= 68) of 5089 metabolomic compounds detected by UPLC-OT-FTMS B) PCA Loading plots of 5089 metabolomic compounds detected in all samples.....14

Figure 7: a) Principal component analysis (PCA) score plot of both M1M1 and M22 samples (n= 65; after removing three outlier samples) of 5089 metabolomic compounds detected by UPLC-OT-FTMS b) PCA Loading plots of 5089 metabolomic compounds detected in all samples.....14

Figure 8: Volcan plots for M2M2 samples, there were 26 nonpregnant compounds and 72 pregnant compounds.....16

Figure 9: Volcano plots for M11 samples before removing the outlier; there were eight pregnant compounds and ten pregnant compounds.....16
Figure 10: Volcano plots for M11 samples after removing the outlier, there were 15 nonpregnant compounds and 34 pregnant compounds.....17
Figure 11: Volcano plots for All samples; before removing the outlier, there were 14 nonpregnant compounds and 23 pregnant compounds.....17
Figure 12: Volcano plots for All samples After removing the outlier, there were 48 nonpregnant compounds and 29 pregnant compounds.....18
Figure 13: Box Whisker of some nonpregnant compounds that are found in all samples run.....19
Figure 14: Box Whisker of some pregnant compounds that are found in all samples run.....19
Figure 15: The tryptophan amino acid metabolomic pathway that is associated with non-pregnant samples.....24
Figure 21: Super pathway of branched chain amino acid biosynthesis (identified pathway compounds 16).....25
Figure 22: Vitamin and cofactor metabolism detected by KEGG.....26

LIST OF TABLES

TABLES PAGE

Table 1: Treatment result after ART (assisted reproductive technology) in European countries in 2010.....2

Table 2: Studies that used culture media analysis to evaluate embryo.....5

CHAPTER I. INTRODUCTION

1.1 BACKGROUND

Parenthood is a universally desired goal for couples; however, infertility can prevent or delay this goal's achievement. Infertility is the inability to conceive after one year of regular unprotected intercourse (Kovac et al., 2013). The World Health Organization (WHO) has recognized infertility as a public health issue worldwide (Boivin et al., 2009). It affects about 15% of couples and equally affects both males and females (Kovac et al., 2013). One standard protocol to treat infertility is in vitro fertilization (IVF) techniques as an assisted reproductive technology (ART). IVF aims to provide high-quality embryos capable of implantation and development, which will result in pregnancy and delivery of healthy infants (Ménézo et al., 2013). The first successful IVF was achieved in 1978 by Patrick Steptoe and Robert Edwards, which resulted in the birth of Louise Brown. From 1978 to 2013, five million IVF inf ants were born worldwide, and each year over 200,000 IVF babies are born (Qin et al., 2014). However, this does not mean that IVF always results in pregnancy. For example, the average pregnancy rate of IVF in different European countries was 29.2 per aspiration (Table 1). Furthermore, pregnancies lost to follow-up starting from the clinical pregnancy stage were 7.3% for IVF, which implies that the mean delivery rate per aspiration for IVF is 22.4% (Kupka et al., 2014). Because of the low rate, most clinics transfer more than one embryo at each attempt to increase IVF success. As a result of this practice, the current rate of multiple gestational pregnancies in IVF pregnancies higher than the rate in spontaneously conceived (SC) pregnancies (Qin et al., 2014). Figure 1 shows the rate of multiple births (delivery of two or more children) in IVF/ICSI over 14 years in Europe. Multiple gestational pregnancies cause an economic burden and affect the health of both mother and infants. For instance, multiple

gestational pregnancies can cause premature delivery and cause ovarian hyperstimulation

Country	Initiated cycles IVF + ICSI	IVF			ICSI			FER			ART infants[a]	ART infants per national births (%)
		Aspirations	Pregnancies per aspiration (%)	Deliveries per aspiration (%)	Aspirations	Pregnancies per aspiration (%)	Deliveries per aspiration (%)	Thawings FER	Pregnancies per thawing (%)	Deliveries per thawing (%)		
Austria	6161	1014	32.5		4768	34.9		620	25.6		1559	2.0
Belgium	20 572	4526	26.8	19.6	13 131	25.8	17.9	8815	18.3	12.4	5199	4.0
Bulgaria	4673	666	29.9	24.5	3960	28.8	22.7	272	22.4	17.3	1595	2.1
Czech Republic	12 864							4303				
Denmark	12 234	6304	25.1	22.4	5417	25.6	23.0	3371	17.0	14.6	3724	5.9
Finland	4861	2516	31.0	25.0	2147	27.8	22.1				1859	3.0
France		21 783	24.1	18.6	34 709	27.5	21.6	21 376	16.1	11.6	16 500	2.0
Germany	51 720	9545	27.9	19.1	35 150	27.6	19.0	17 876	19.2	12.0	14 123	2.1
Greece	2754	710	33.7	27.7	1703	31.1	23.8	461	28.0	21.3	1102	
Hungary	5091	1211	32.5		3863	30.3						
Iceland		279	26.2	21.5	223	32.7	27.8	205	29.3	23.9	216	4.4
Ireland	3176	1483	31.2	25.6	1173	32.7	26.1	882	17.5	13.9	971	1.2
Italy	52 661	7606	24.1	16.6	39 843	23.0	15.7	3758	17.2	11.5	9794	1.7
Kazakhstan	1630	1282	34.9	25.8	348	32.5	22.4	289	25.6	17.6	701	
Lithuania	112	57	26.3		53	34.0		19	36.8	0.0		
Macedonia	1409	283	44.2	32.2	1009	41.3	33.9	49	24.5	16.3	611	
Moldova	624	272	38.2	31.6	328	39.6	33.5	0			242	0.6
Montenegro	446	27	48.1	48.1	417	28.3	21.3	6	33.3	33.3	136	1.8
Norway	6557	2942	29.3	24.4	3314	27.9	23.7	2443	19.9	15.7	2098	4.1
Poland	8968	335	37.3	16.7	8501	34.5	26.3	3733	23.2	14.7	3500	0.8
Portugal	5875	1571	35.1	26.6	3856	31.0	24.1	921	20.3	15.2	1962	1.9
Romania	923	506	41.1		352	48.9		208	24.5			
Russia	27 310	13 817	34.2	25.2	12 508	33.8	23.8	3760	24.7	15.0	9500	
Serbia	1484	410	31.5	23.9	1050	35.4	27.5				484	
Slovenia	3604	1205	34.4	28.0	2310	28.1	22.3	760	21.3	17.1	1131	5.1
Spain	32 503	2880	33.8	17.5	25 994	32.7	19.7	8760	27.9	15.4	13 385	2.8
Sweden	11 592	5348	31.8	24.5	5499	31.3	24.3	5520	25.0	19.0	4025	3.5
Switzerland	5482	741	21.5	16.7	4452	23.9	18.4	4058	19.1	13.4	1733	2.2
The Netherlands	16 898	7895	28.5	21.3	7639	31.8	24.3				5015	2.7
Ukraine	5264	2328	38.2	27.7	2794	36.5	30.4	1240	26.9	20.2	2455	
The UK	44 642	18 738	30.9	27.0	23 160	31.2	27.6	10 476	21.9	19.3	17 014	2.2
All[b]	352 090	118 280	29.2	22.4	249 671	28.8	21.1	104 181	20.3	14.1	120 634	

For IVF and ICSI there were for France, Greece, Ireland, Kazakhstan, Russia and Spain, respectively 177, 46, 1, 8, 543 and27 deliveries with unknown outcome. These were accepted as singletons to calculate the ART infants. For FER there were for France, Greece, Kazakhstan, Russia and Spain, respectively 41, 4, 2, 8 and 4 deliveries with unknown outcome. These were accepted as singletons to calculate the ART infants. For the Netherlands no data on the number of thawings were available. For ED there were for France, Greece, Kazakhstan, Poland, Russia, Spain and Ukraine, respectively 1, 2, 1, 1, 23, 8 and 9 deliveries with unknown outcome. These were accepted as singletons to calculate the ART infants. For PGD there was for Russia 1 delivery with unknown outcome. This one was accepted as singleton to calculate the ART infants. In the Czech Republic, IVF and ICSI were reported together, no details on pregnancies and deliveries.

[a]ART infants also include ED.

Table 1: Treatment Results after ART (assisted reproductive technology) in European countries in 2010

syndrome (OHSS) (Kupka et al., 2014).

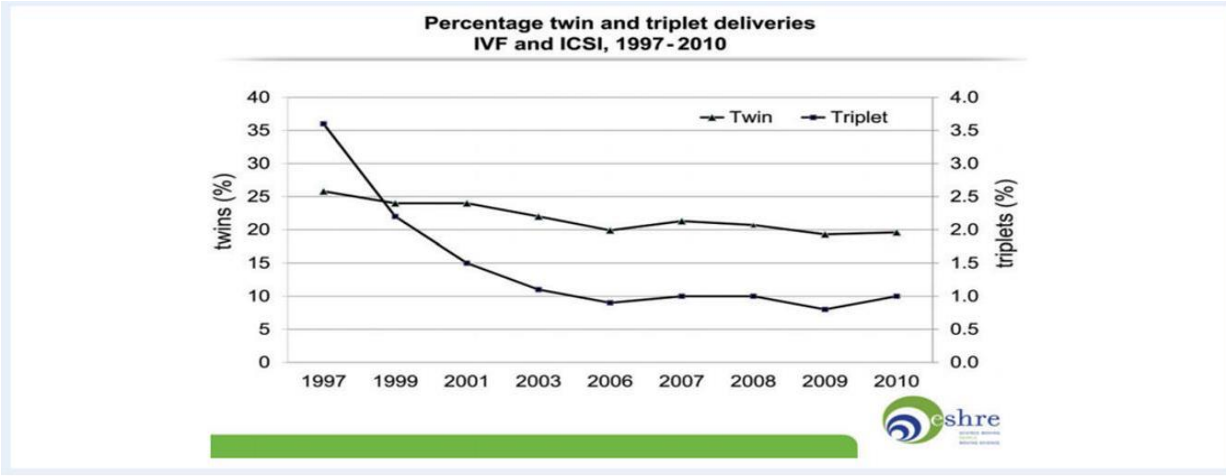


Figure 1: Multiple birth rate (delivery of two or more children) in IVF/ICSI over 14 years in Europe.

To increase the pregnancy rate and embryo development, we need to develop a better clinical approach for selecting a viable embryo to transfer via IVF. There are two ways of selecting a viable embryo: invasive methods and non-invasive methods. An example of invasive techniques is pre-implantation genetic screening (PGS). PGS utilizes the recent advances in molecular-based screening technology, such as a generation sequencing platform and analyzing a fertilized embryo's chromosomal normalcy. However, this method has many disadvantages since it requires micromanipulation with an embryo of one or more blastomeres in the early development stages. Second, not all euploid embryos are developmentally competent, so not all of them result in healthy babies' pregnancy. Also, embryo mosaicism is still a challenge for PGS (e.g., Spinella et al. 2018).

On the other hand, non-invasive methods do not involve interfering with the embryo. For example, morphological scoring has been used in clinics since the first successful IVF in 1978 until now (Rødgaard et al. 2015). However, the pregnancy rate prediction using this method is around 30%. There is an urgent need to develop a new non-invasive approach that has higher predictivity.

The new approach should reflect the embryo's function or physiological state without affecting the embryo's viability. Recently, a few studies tried to use the culture media of embryos during the pre-implantation stages to assist the embryo selection, which are indirect and non-invasive methods (Rødgaard et al. 2016). These studies investigate the changes in proteomics or the selected number of targeted metabolites. However, none of those methods have been applied as clinic routine yet, and none of them utilize the untargeted metabolite approach.

Metabolomes are the final downstream products of gene transcription (considered a molecular phenotype), which means any change in the embryo's biochemical pathways will be vastly amplified in the metabolomic profile relative to transcriptome and proteome profiles (see Figure 2).

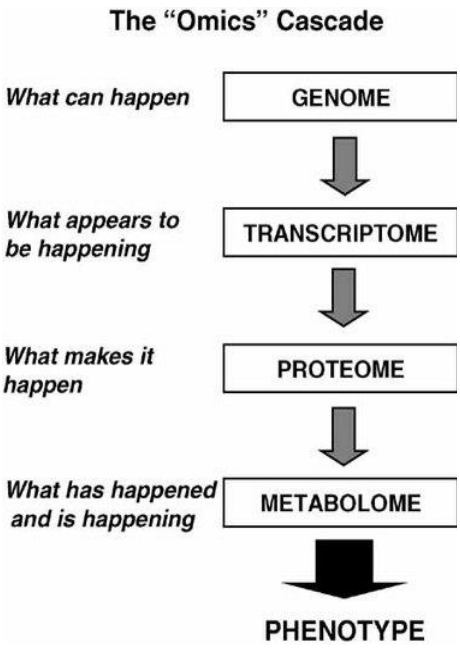


Figure 2. Diagram of Omics Cascade. Adopted from Dettmer (2007)

1.2 RELATED WORK

Analyzing culture media to determine the embryo's quality has not been applied as clinic methods yet, because they fail to validate the used analyzing technology. Most analytical techniques used showed poor reproductivity when involved in clinical trials (Sanchez, 2017). For example, some studies used Near-infrared (NIR) spectroscopic metabolomic profiling of spent embryo-culture media to assess the embryo potential for implementation (Ahlström et al., 2011). The study suggests that morphology and embryo metabolism are independent of one another. NIR is suitable for a clinic routine because of the minimal sample preparation. Still, there are two

significant NIR issues: First, NIR is limited to metabolomics because it only measures functional groups' vibrations and not the individual metabolite. Second, the NIR pregnancy rate did not show any advantage over morphology when applied in the clinical trial (Rødgaard et al., 2015). In another study, using the ^{1}H-NMR technique, Pudakalakatti et al. (2013) found higher pyruvate uptake and a lower ratio of pyruvate/alanine in day three culture media for embryos result in pregnancy relative to the one that did not. In contrast with Pudakalakatti et al. (2013), Seli et al. (2011) did not show any correlation between the high intake of pyruvate or low pyruvate/alanine ratio pregnancy rate. Instead, Seli et al.'s (2011) study showed that more elevated glutamate, lower alanine levels, and a decrease in the alanine to lactate ratio are associated with embryo implantation. Nonetheless, NMR is not ideal for studying metabolites due to its low sensitivity

Table2: Shows some studies that used culture media as non-invasive approach to evaluate embryos' capability of implantation.

Table 1 Overview of recent metabolomic findings.

Study	*Method*	*Species*	*Day*	*Sample size*	*Findings*
Wiener-Megnazi et al., 2011	TLC	Human	2 and 3	133 patients–284 embryos	Oxidative status can be used as a transfer marker at day 3
Vergouw et al., 2011	NIR	Human	5	115 patients–127 embryos	Correlation between increased viability score and live-birth outcome
Seli et al., 2011	NIR	Human	5	198 embryos	Viability score better indicator for transfer outcome than morphologic grading
Gardner et al., 2011	Micro-fluorimetry	Human	4 and 5	50 patients–50 embryos	Increased glucose uptake indicates higher pregnancy rate and female embryos have higher uptake than male
Combelles et al., 2012	ELISA	Human	3	91 patients–122 embryos	SOD1 (superoxide dismutase-1) cannot be used as a biomarker for transfer potential
Pudakalakatti et al., 2013	^{1}H NMR	Human	2 and 3	48 patients–127 embryos	Pyruvate/alanine ratio can predict transfer outcome at day 3
Fu et al., 2013	NIR	Human	3	172 patients–239 embryos	Viability index predicts transfer outcomes
Wallace et al., 2014	^{1}H NMR	Human	2	37 patients–58 embryos	The ratios between different metabolites are correlated with transfer potential
Kirkegaard et al., 2014	^{1}H NMR	Human	3 and 5	161 patients–148 embryos	No correlation between NMR variations and pregnancy outcome
Vergouw et al., 2014	NIR	Human		924 patients	Meta-analysis of four studies: NIR profiling does not improve birth rates
Muñoz et al., 2014a	FTIR	Bovine	6 and 7	69 embryos	In a subgroup of blastocysts, using FTIR is correlated with higher birth rates
Muñoz et al., 2014b	FTIR	Bovine	7	49 embryos	FTIR profiling correlates with viability of in vivo embryos however is affected by embryo management procedures

ELISA = enzyme-linked immunosorbent assay; FTIR = Fourier transform infrared spectroscopy; ^{1}H NMR = proton nuclear magnetic resonance spectroscopy; NIR = near infrared spectroscopy; TLC = thermochemiluminescence.

and difficulty identifying individual metabolites within a complex mixture. A summary of recent IVF studies and the used analytical techniques. Table 2 Shows

This study aims to discover metabolomic biomarkers secreted by the embryo and the culture medium that could predict the implantation potential. We investigated the changes in the culture's metabolomic profile using ultraperformance liquid chromatography (UPLC) coupled with high resolution and high accuracy mass spectrometer (Fusion Orbitrap mass spectrometer). This study could increase the success rate of IVF, which will be beneficial in many ways: 1). It promotes a better chance of pregnancy for couples who suffer from infertility, 2). It minimizes the need to transfer more than one embryo at one process, alleviating the health complications arising from multiple pregnancies for both infants and mothers. 3). Economic wise, if the first IVF attempt is successful, no more trials would be needed, lowering the financial cost (Thompson, 2016).

CHAPTER II: METHODS & PROCEDURES

2.1 SAMPLE PREPARATION

We analyzed samples from two clinics (M1M1) and (M22) with different culture media collected from day five-stage (blastocyst stage, the embryo about 20 cells). Thirty-nine of those culture media samples lead to pregnancy (P), and Thirty-two culture media samples from an embryo did not result in pregnancy (NP). Overture Life Inc. provided these samples as a part of an analysis contract.

We diluted 20µL of the sample with 480µL UPLC water to be a total of 500µL in Amicon® Ultra 0.5mL 3kDa centrifugal filters. Amicon® Ultra was spin at 14,000 x g at 4 ^{o}C for 10 min to separate the metabolites from the proteins. The protein was recovered by flipping the membrane filter and spun again in a new tube at 4,000 x g at 4 ^{o}C for 3 min. We froze the protein extract at -20^{o}C for future analysis. While we store the metabolomics fraction (~480 µL) frozen at -20^{o}C until the analysis by UPLC-Fusion Orbitrap mass spectrometer.

2.2 METABOLOMIC ANALYSIS BY UPLC-ORBITRAP FUSION MASS SPECTROSCOPY

The metabolomics approach has many challenges to be overcome, such as structural diversity, background interference, sample limitation (volume limitation, density limitation, and dynamic range of polarity). The recent coupling of ultraperformance liquid chromatography (UPLC) with an ultrahigh-resolution and high-mass accuracy mass analyzer (i.e., Orbitrap mass spectrometer) has overcome many these challenges. UPLC reduced metabolites' chemical complexity by separating the metabolites mixture on a second orthogonal dimension (retention time) to the first dimension (m/z) based on their polarity, electrical charge, and molecular size. High resolving power provides the ability to measure small mass differences required to assign the organic molecular formulas of different isobaric ions accurately. In this study, I used the

Orbitrap Fusion mass spectrometer (OT-FTMS), the state-of-the-art mass analyzer with a resolving power up to 500,000 Full-Width at Half Maximum peak Hight (FWHM) at m/z 200, and scan time is one second. On the other hand, the high mass accuracy analyzer allows determining the molecule's elemental composition by eliminating most other possibilities. Which make it an ideal mass analyzer for metabolomics and lipidomic that required at least two (2) ppm accuracy (Makarov and Scigelova, 2010),

I analyzed the metabolomic extract's aliquots on Vanquish UPLC – coupled with heated ESI (H-ESI) source Orbitrap Fusion Tribrid Mass Spectrometer (UPLC-OT-FTMS) and operate in positive mode. I used a 1.7 μm ACQUITY UPLC BEH C18 reversed-phase column (Waters, 30Å, 1.7 μm, 2.1 mm X 100 mm). Eluent A was Milli-Q water with 0.1% (v/v) formic acid, and eluent B was acetonitrile with 0.1% (v/v) formic acid. I used the following gradient: 5% of B for 2 min; ramp to 65% B for 16 min; ramp to 100% B for 7 min and hold for 8 min. An 8 min column re-equilibration with the starting ratio of eluents was carried out between sample analyses. The flow rate was 0.2 $ml \cdot min^{-1}$ with an injection volume of 20 μL. The H-ESI setting was 3200 volts, 30 Sheath gas, 10 Aux gas, 325^{o}C ion transfer tube temp, and 200^{o}C vaporizer temp. The Orbitrap full scan was run at 500,000 (FWHM at m/z 200) resolutions with a scan range of 100-1000 m/z and RF Lens at 40%. For MS^2, the isolation window was set at 0.7 m/z with performing both collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD) using an ion trap mass spectrometer as the detector. The AGC was set at 1.0e4. To increase the percentage of compounds that get MS/MS fragmentations, I applied a novel intelligent data-dependent technique called the "dynamic exclusion technique." The technique is based on adding the compounds m/z that have already been fragmented into a temporary exclusion list for 30-90 s, so they will not be fragment again, giving other molecular ions to get fragmented within that time.

2.3 STATICAL ANALYSIS

We used Compound Discoverer software 3.1 to identify the metabolomics compounds and perform multivariate analyses (Volcano plot, Principal Components Analysis (PCA), and trend analysis). The retention times (RT) of all chromatography spectra were aligned using an adaptive curve with a maximum shift of 0.2 min and five ppm mass tolerance. To identify a compound, we required it to meet the following conservative criteria: 1) a signal-to-noise (S/N) above 3, 2) a minimum of 5 mass scans per chromatographic peak, 3) a minimum peak intensity of 50,000, and 4) at least one isotope peak (M+1) was detected. We have used the ratio of the M+1 to parent peaks to confirm the number of carbon atoms. We also used the M+2 peak ratio to confirm the presence of the S atom in the compound. The workflow considered the possibility of the presence of multiple positive adducts ($[M+H]^{+1}$; $[M+K]^{+1}$; $[M+Na]^{+1}$; $[M+NH_4]^{+1}$; $[2M+ACN+H]^{+1}$; $[2M+ACN+Na]^{+1}$; $[2M+H]^{+1}$; $[2M+K]^{+1}$; $[2M+Na]^{+1}$; $[2M+NH_4]^{+1}$; $[M+2H]^{+2}$; $[M+3H]^{+3}$; $[M+ACN+2H]^{+2}$; $[M+ACN+H]^{+1}$; $[M+ACN+Na]^{+1}$; $[M+H+K]^{+2}$; $[M+H+MeOH]^{+1}$; $[M+H+Na]^{+2}$; $[M+H+NH_4]^{+2}$; $[M+H-H_2O]^{+1}$; $[M+H-NH_3]^{+1}$). In case multiple adducts were detected, all adducts of the same compound were grouped with a tolerance of 0.2 min (in retention time).

2.4 MOLECULAR FORMULA CALCULATOR

I calculated the molecular formula for each peak using a molecular formula calculator software (Molecular Formula Calc version 1.0 NHMFL, 1998) with the following parameters: $C_{1-72}H_{2-200}O_{0-50}N_{0-10}S_{0-3}P_{0-3}$. I removed the molecular formulas that are unlikely to occur in nature (or that are not chemically possible), as described in detail in (Abdulla et al., 2013). In summary, I applied a modified version of the rules set in Kind and Fiehn (2007), which requires that formulas satisfy the following inequalities: $H/C < 2.50$, $O/C \leq 1.20$, $O/P \geq 3.00$, and $N/C < 0.50$. All assigned formulas were further tested for the physical existence of chemical structures using LEWIS and SENIOR chemical

rules, again according to Kind and Fiehn (2007). The molecular ^{13}C isotope and ^{34}S isotope peaks (when they were detected above the S/N thresh-old) were also validated with the chemical building block approach (e.g., CH_2 homologies series) described by Koch et al. (2007). As I did not use internal standards when analyzing these samples, the calculated masses of the assigned formulas are within 2.0 ppm of the masses detected by OT-FTMS. I used a Python code (designed by Breeana Cross) to filter the incorrect molecular formula results using the above parameters. *De novo* compound structure elucidation was identified based on the tandem mass spectrum and compared with fragmentation database software mzCloud and in silico fragmentation by Mass Frontier software.

CHAPTER III: RESULTS & DISCUSSION

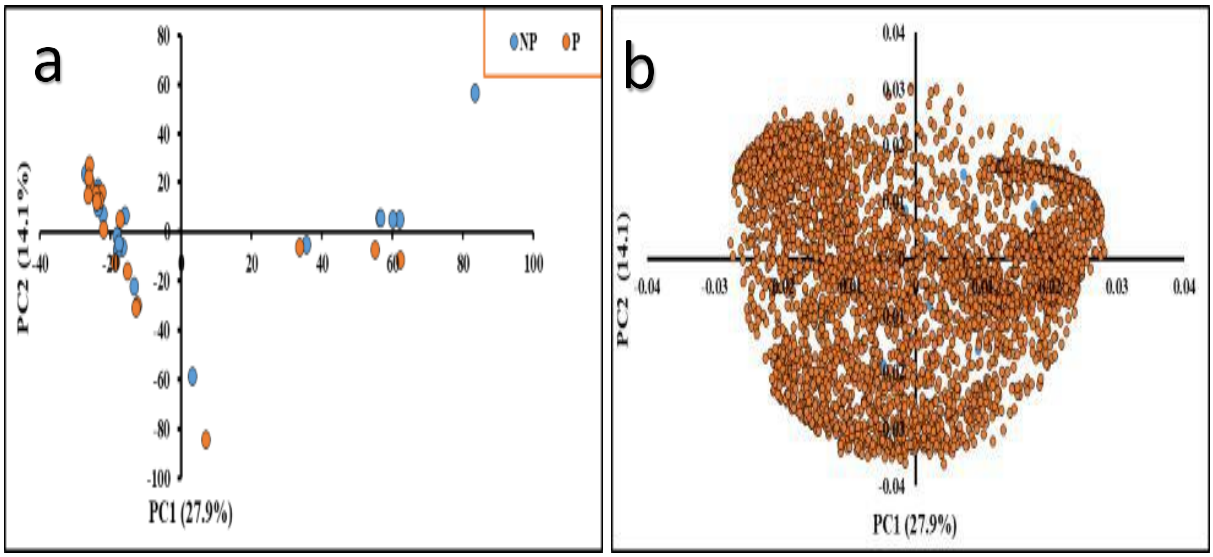
Over 4200 unique metabolomics compounds were detected in 36 M2M2 samples by UPLC-OT-FTMS; out of that, we got MS/MS (MS^2) fragmentation for 3100 metabolites. With our dynamic exclusion technique, I increased MS^2 fragmentation for about 74% of detected metabolites. For our 32 M samples, I detected over 2800 metabolomic compounds, and out of these compounds, 2100 compounds were fragmented MS^2 which accounts for 72% of the total compound detected. Analyzing all the samples together (both M1M1 and M22 samples), I was able to identify 5089 compounds, with 78 % of these detected compounds have MS^2 fragmentation. These results illustrate the percentage of compounds that get MS^2 fragmentation to over 70% of the compound detected. These percentages are significantly higher than the typical data-dependent acquisition (DDA) technique that can only fragment around 20% of the compound detected in complex metabolomic samples, which increases our ability for structural elucidation higher number of metabolomics compounds in each sample.

3.1 Principal Components Analysis

I applied principal components analysis (PCA) to investigate the differences in different samples' metabolomics profiles. PCA aims to reduce our dataset's dimensionality and identify new meaningful underlying variables between pregnant and non-pregnant. Applying PCA on M2M2 samples shows the first two principal components explained 42% of the conditioned media differences (Figure 5). We did not notice any clusters in pregnant (P) or non-pregnant (NP) samples, and we did not identify any sample outliers. For M1 samples, the first two principal components explained 44.7% of the different culture media samples (Figure 6). Even though I did not see any clustering for P and N samples, I identified three outlier samples. Taking a closer look at the raw spectrum of these three samples, I noticed an irregularity in their chromatographs relative to other samples. These three spectra have high intensity of a handful of compounds not observed in the other samples, indicating potential contamination of these three samples during the transfer for the analysis. I removed those outliers and conducted a new PCA again. In the new PCA of the M1 samples (n=29), I did not observe any P vs. NP samples' clustering, but the samples don't show any outliers (Figure 7). Combining M1M1 and M22 samples, I didn't also observe any clustering (Figure 8).

Despite the PCA technique is used widely to highlight the difference between samples by reducing the dimensionality of extensive metabolomic data and maximizing the variance between the samples, it's not surprising that it cannot cluster each pregnant from non-pregnant samples. As PCA is an unsupervised learning technique, it doesn't consider if the sample results label (pregnant vs. non-pregnant). In my case, the samples share a similar metabolomics background (the original culture media), which makes it challenging to identify any differences between pregnant vs. non-pregnant samples using the unsupervised technique. Which makes PCA a non-

ideal technique to detect these minor differences, especially for achieving a supervise objective. Nevertheless, PCA is a powerful technique to identify any outliers in the dataset or any irregularity.



i re a rincipal component analysis score plot of t e samples n of metabolomic compo n s etecte by b oa in plots of s metabolomic compo n s.

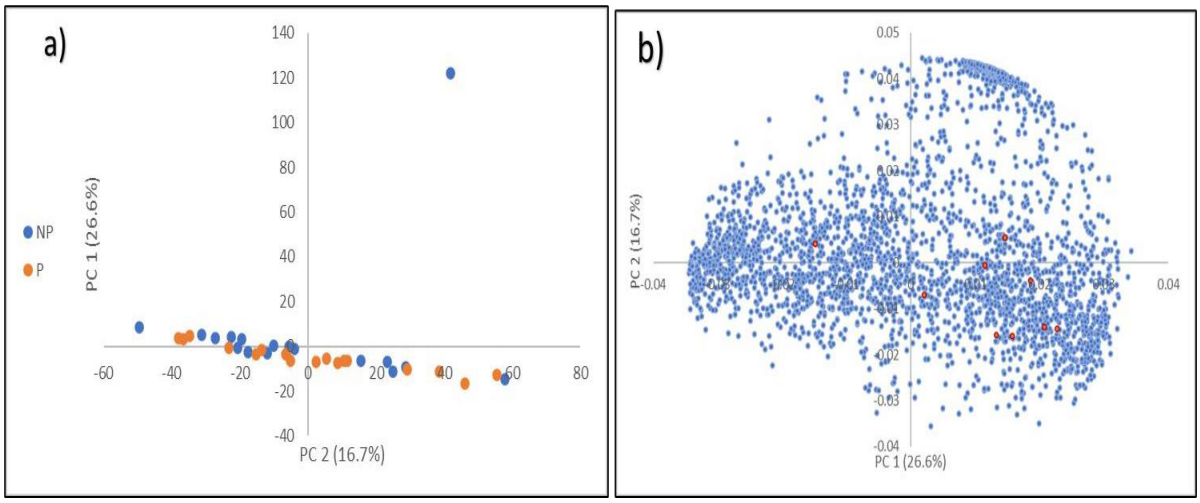


Figure 4: a) Principal component analysis (PCA) score plot of the M11 samples (n= 29) of 2800 metabolomic compounds detected by UPLC-OT-FTMS B) PCA Loading plots of M1's 2800 metabolomic compounds.

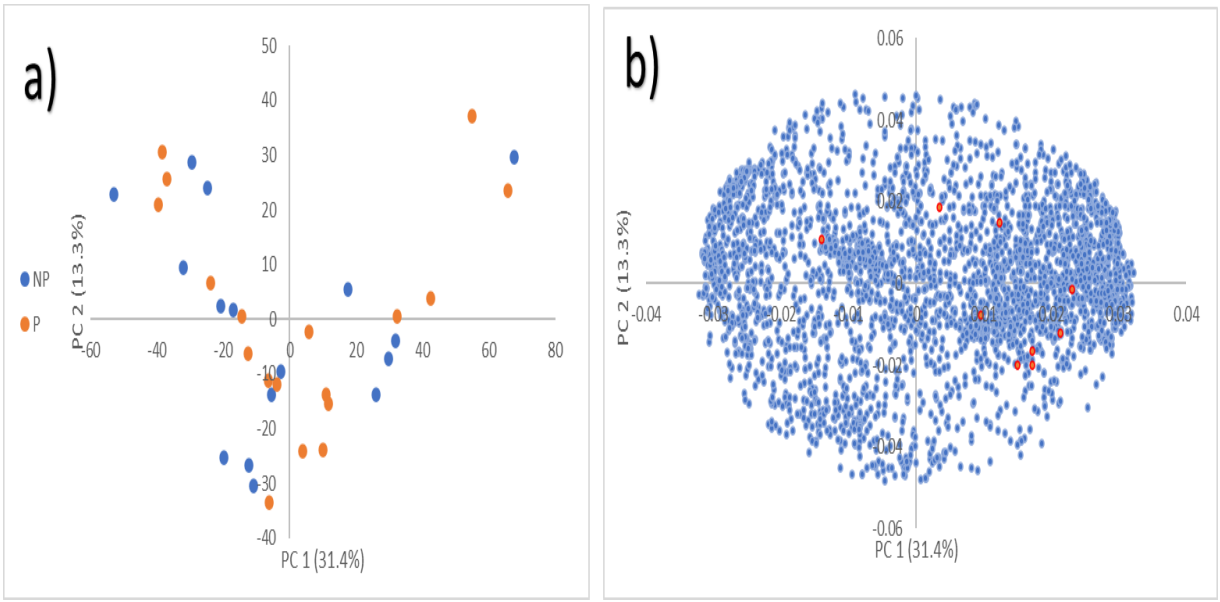


Figure 5: a) Principal component analysis (PCA) score plot of the M 11 samples (n= 29; removing three outlier samples) of 2800 metabolomic compounds detected by UPLC-OT-FTMS b) PCA Loading plots of M1's 2800 metabolomic compounds.

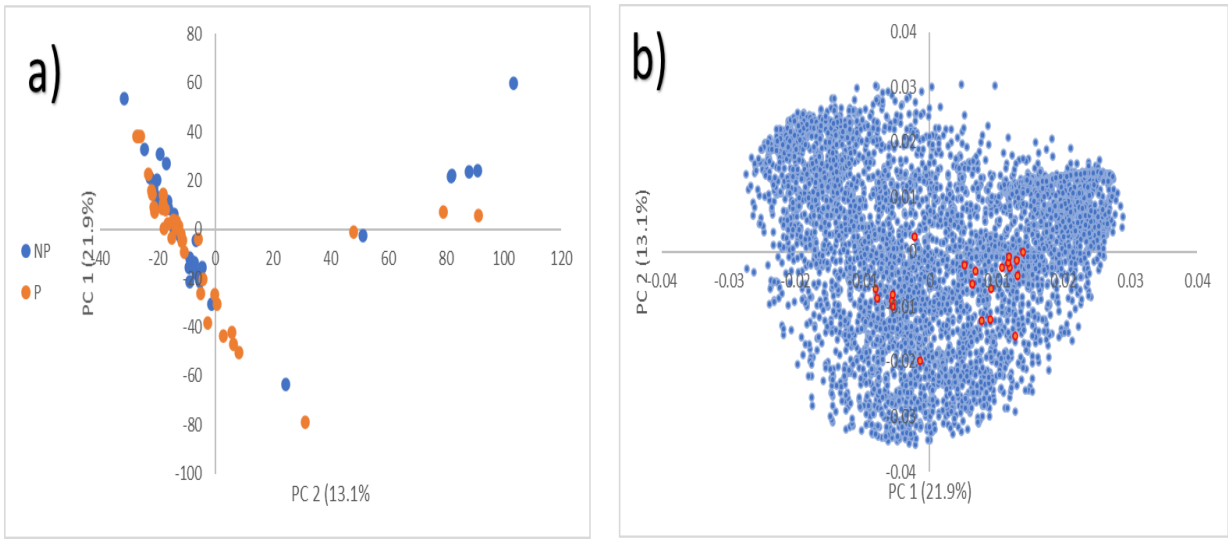


Figure 6A) Principal component analysis (PCA) score plot of both M1M1 and M22 samples (n= 68) of 5089 metabolomic compounds detected by UPLC-OT-FTMS B) PCA Loading plots of 5089 metabolomic compounds detected in all samples.

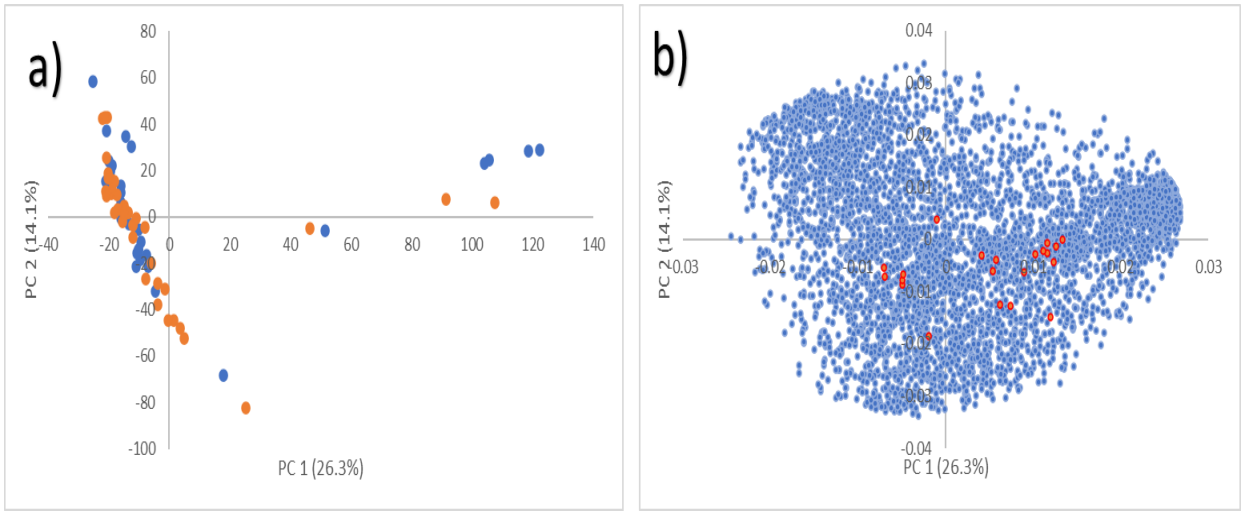


*Figure 7: **:** a) Principal component analysis (PCA) score plot of both M1M1 and M22 samples (n= 65; after removing three outlier samples) of 5089 metabolomic compounds detected by UPLC-OT-FTMS b) PCA Loading plots of 5089 metabolomic compounds detected in all samples.*

3.2 DIFFERENTIAL ANALYSIS (VOLCANO PLOTS)

The differential analysis techniques' power is identifying quantitative changes in compounds between pregnant and non-pregnant sample groups. I used Volcano plots to determine the metabolomic biomarkers that are significantly different between pregnant and non-pregnant. The volcano plots evaluate these differences by plotting significance ($-\log_{10}$ of the p-value from an ANOVA or t-test) on the y-axis versus $\log_2$ of fold-change (FC) on the x-axis (Hur et al., 2018). The compounds with p-values below the chosen significance level and $\log_2$ < FC threshold are considered significantly down-regulated (compounds associated with pregnancy) or significantly specific to the denominator, whereas the compounds meeting the same p-value threshold and $\log_2$ > FC threshold are considered significantly up-regulated (compounds associated with nonpregnancy) or significantly specific to the numerator. Compounds with

significant p-values and FC meeting the upper or lower FC threshold are particular to that group of samples or characteristic of that sample group. For all plots, the p-value was set to 0.05, and FC was set to 1. The larger the $\log_2$ FC, the higher intensity that compound has in that sample group compared to the other. The likelihood that a compound is present is higher with an increasingly more significant $-\log_{10}$ p-value. Thus, compounds in the upper left and upper right portions of the volcano plot are statistically more characteristic of that group and different from the other groups. It follows that these statistically distinct compounds can be isolated to explore how sample groups differ based on compound class and structure. Compound Discoverer 3.1 used the NOVA test to calculate P-value and then use $-\log_{10}$ p-value in the y-axes and $\log_2$ fold in the x-axes. I classified the up-regulated compounds (significant and higher than upper FC threshold) as non-pregnant biomarkers and down-regulated compounds (significant and less than the lower threshold) as pregnant biomarkers. In M2M2 samples, I identified 26 nonpregnant compounds and 72 pregnant compounds (see Figure 10). In M1 samples, before removing the outlier, there were eight nonpregnant compounds and ten pregnant compounds. On the other hand, there were 15 nonpregnant compounds and 34 pregnant compounds (see Figures 11&12). When I ran all the samples without removing the three outliers, there were 14 nonpregnant compounds and 23 pregnant compounds. After removing the outlying sample, there were 48 nonpregnant compounds and 29 pregnant compounds. Which approve that PCA could be a handy tool to identify outliers

which could remove and enhance volcano plots in identifying nonpregnant and pregnant compounds.

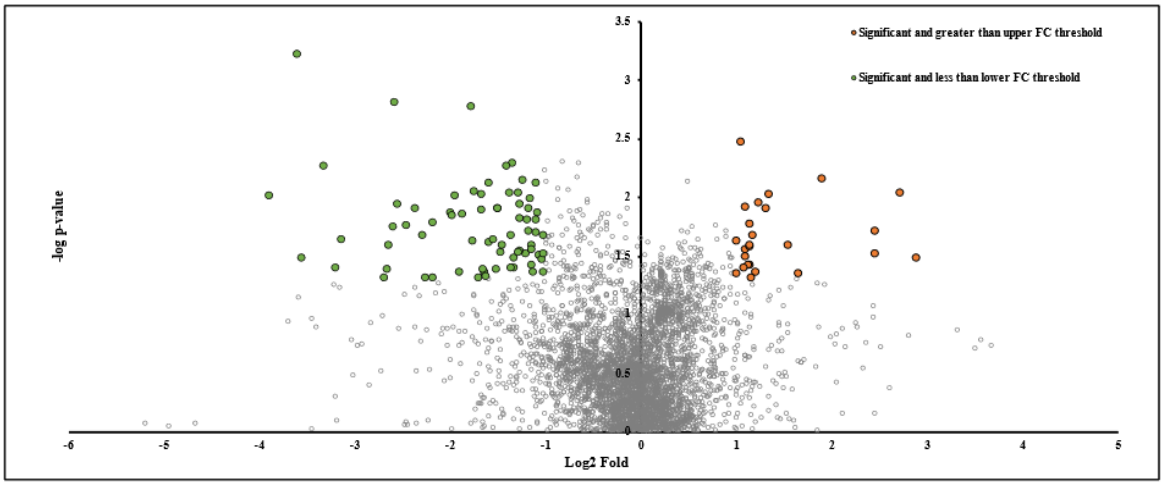


Figure 8: Volcan plots for M2 samples, there were 26 nonpregnant compounds and 72 pregnant compounds.

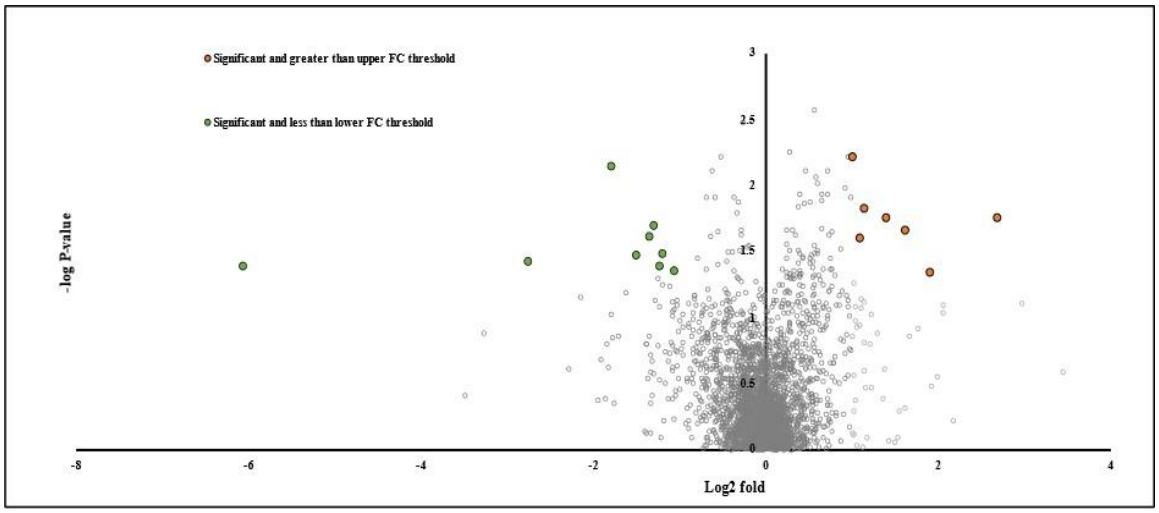


Figure 9: Volcano plots for M1 samples before removing the outlier; there were eight pregnant compounds and ten pregnant compounds.

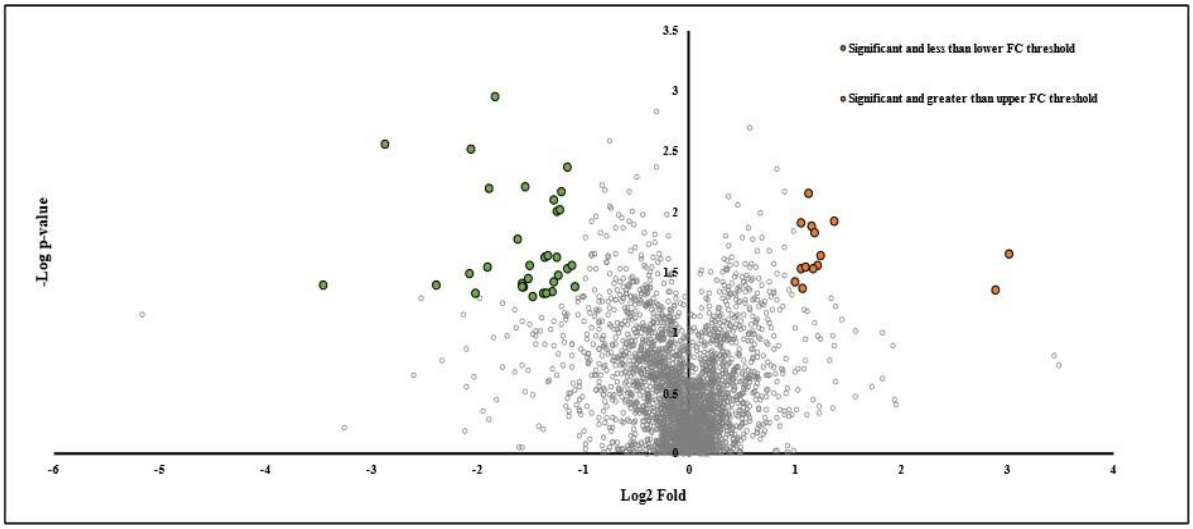


Figure 10: Volcano plots for M1 samples after removing the outlier, there were 15 nonpregnant compounds and 34 pregnant compounds.

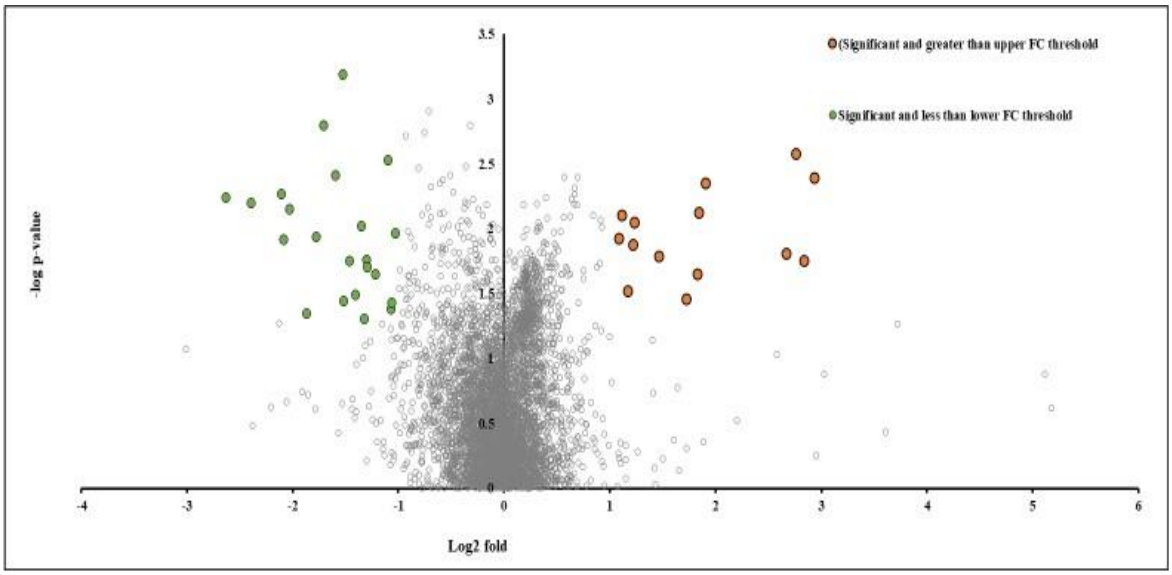


Figure 11: Volcano plots for All samples; before removing the outlier, there were 14 nonpregnant compounds and 23 pregnant compounds.

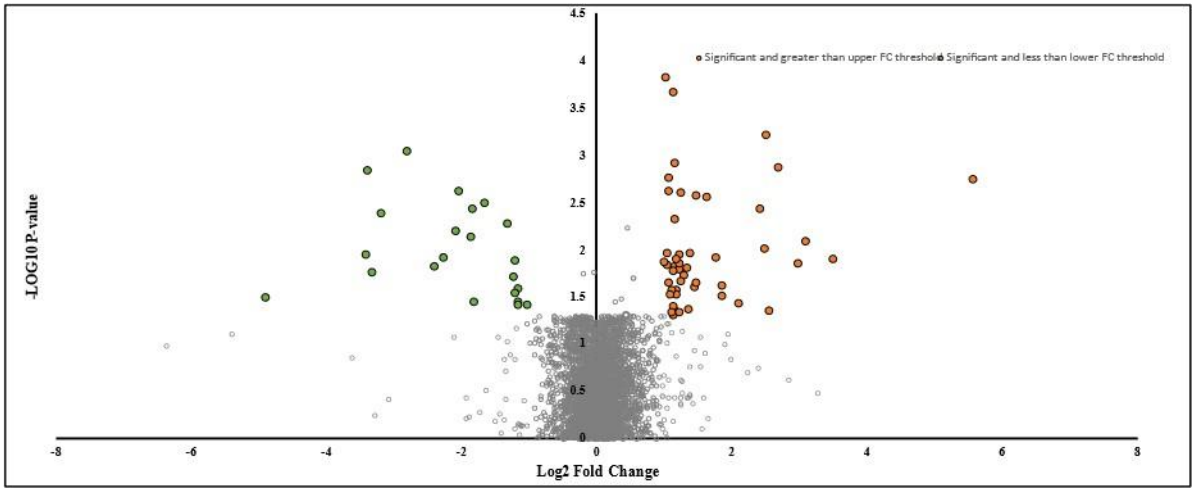


Figure 12: Volcano plots for All samples After removing the outlier, there were 48 nonpregnant compounds and 29 pregnant compounds.

3.2.1 MOLECULAR FORMULA & PYTHON CODE

After identifying pregnant and nonpregnant compounds, we run them through the Molecular Formula calculator. To assure that formulas from the compound discoverer for those compounds are correct, we compared both assigned formula (the MF calculator formula and Compound Discoverer formula) for each compound (See Tables 3 to 12)

3.3 BOX WHISKER CHART

To confirm these biomarkers' identification, we used the Box Whisker chart to verify the statistically significant difference in these biomarkers' concentration between pregnancy and non-pregnancy. Some examples of Box Whisker charts are shown in Figures 13 and 14.

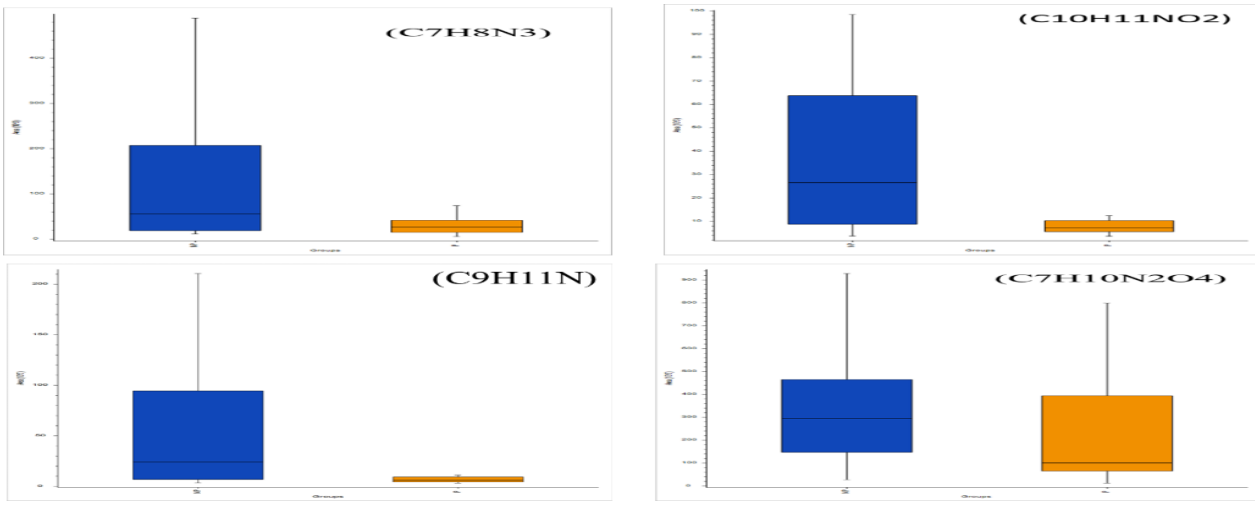


Figure 13: Box Whisker chart of some nonpregnant compounds that are found in all samples run.

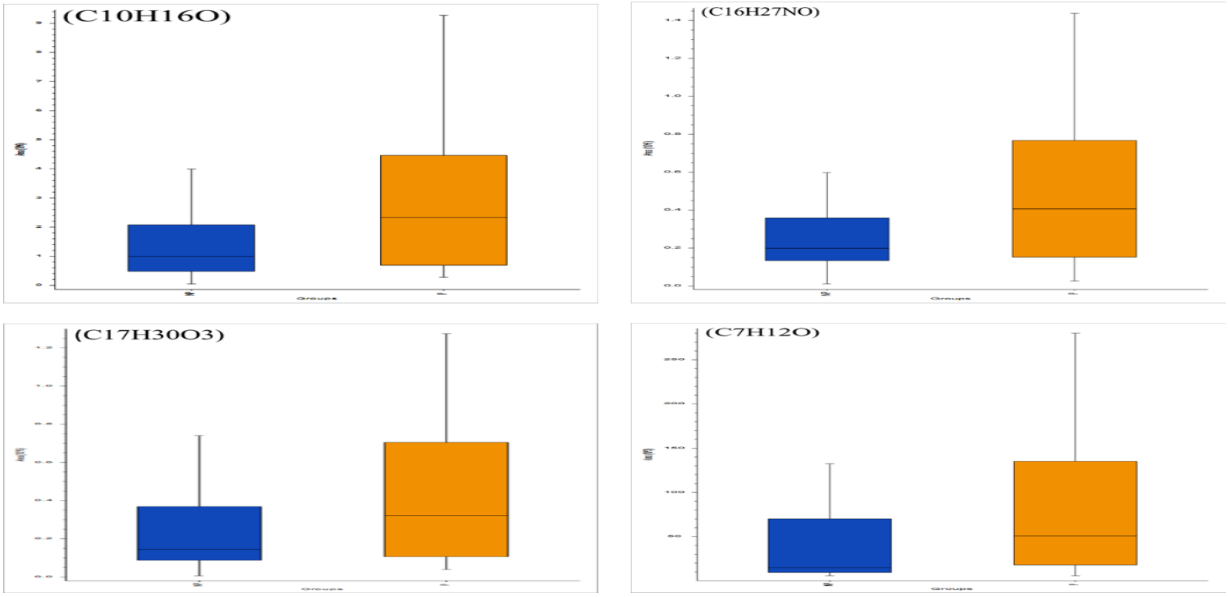


Figure14: Box Whisker chart for some pregnant compounds for All samples

3.4. Identification of the Biomarker Chemical Structures

I compared the MS^2 fragmentation detected for each biomarker formula to either mzCloud and ChemSpider databases to identify the possible chemical structures. Besides, we generated in silico- fragmentation spectra (MS^2) for the proposed structure using Mass Frontier software. These in silico fragmentation spectra were compared to the measured MS^2 spectra for each biomarker by the FISh score node in the Compound Discoverer. The FISh score generates a similarity percentage score based on match and unmatched fragmentations from both in-silico and measured fragmentation spectra (see Figures 15 to 19). A list of the names and the chemical structures of the biomarkers identified is listed in Tables 13-18.

In M2M2 media samples, I was able to identify 26 nonpregnant compounds and 72 pregnant compounds. Out of the 26 up-regulated compounds, I assigned the molecular formula of 17 compounds (Table 4) and assigned the molecular structure for nine compounds with at least 75% FISh coverage (Table 13). For the 72 pregnant compounds, I assigned the molecular formula for 13 compounds (Table 3) and identified the chemical structures for ten compounds (Table 14). For M media samples (after removing the outlier), I identified 15 nonpregnant compounds and 34 pregnant compounds. Out of the 15 nonpregnant compounds, I assigned the molecular formula for 11 compounds (Table 8) and identified four chemical structures (Table 15). On the other hand, out of the 34 pregnant compounds, I assigned the molecular formula for 23 compounds and the chemical structure for 20 compounds. However, I identified 48 nonpregnant compounds and 29 pregnant compounds in processing the combined two media sample sets. From 48 nonpregnant compounds, I assigned the molecular formulas to 14 compounds (Table 12) and identified the chemical structure of 4 compounds (Table 17). In comparison, I assigned the molecular formula

for 11 compounds as pregnant compounds (Table 11) and identified the chemical structures for five out of the 11 compounds (Table 18).

Comparing the chemical structures of identified biomarkers in the three culture media data sets (M1M1, M22, and the combined M1M1 and M22), I found that the 9 M2M2's nonpregnant chemical structure compounds were unique biomarkers for M2M2 culture media only. They didn't show a significant nonpregnant in the other media (M11 culture media) or the combined sample set. All the four M1 culture media nonpregnant chemical structure biomarkers were unique compounds for M1 samples only. Out of 10 identified chemical structures in M2M2 media samples for the pregnant biomarkers, six compounds were unique pregnant biomarkers for M2M2 media samples. However, 3 of the nonpregnant biomarkers were also identified as nonpregnant biomarkers for the combined samples set (M1M1 and M22 media). These three compounds are 2-Hexylfuran ($C_{10}H_{16}O$), N-(2-Methoxybenzyl)-1-octanamine ($C_{16}H_{27}NO$), and tetrahydro-2-furanylmethyl 6-cyclohexylhexanoate ($C_{17}H_{30}O_3$). We can consider those compounds as essential biomarkers since we can use them in the clinical application regardless of the used culture media.

The differences in the identified biomarkers between the two-culture media could be attributed to that each media is made from a different mixture of nutrients and essential metabolites. This will lead the embryo to follow slightly different anabolic pathways in one media relative to synthesizing the important large biomolecules (like proteins, lipids, and carbohydrates). For example, suppose one of the media is missing phenylalanine amino acid. In that case, the embryo will use a phenylalanine amino acid anabolism from an organic acid or other amino acids to synthesis it so it can synthesis the required proteins. If phenylalanine is present in the original culture media, the embryo will not need to use a phenylalanine amino acid anabolism. This will reflect in different metabolomic profiles between these two media.

3.5. Metabolomic pathways

I used two metabolomic databases to identify possible pathways for pregnant and non-pregnant media sample sets. For nonpregnant compounds in M2M2 samples, I could not identify any potential metabolomic pathways. However, using the KEGG database, I distinguished a potential metabolomic pathway that involves skatole (C_9H_9N), one of the nonpregnant compounds detected in the M1 media samples set (Figure 21). It's part of the tryptophan amino acid metabolomic pathway. This pathway is associated with many disorders' defects syndrome such as Aromatic L-amino acid decarboxylase deficiency, Primary congenital glaucoma, and 3-Hydroxyacyl-CoA dehydrogenase deficiency. Aromatic L-amino acid decarboxylase deficiency is a human Nervous system disease associated with gene dopa decarboxylase (DDC). romatic L-amino acid decarboxylase (AADC) deficiency is an autosomal recessive disses of monoamine neurotransmitter metabolism, clinically characterized via way of means of vegetative symptoms, oculogyric crises, dystonia, and intense neurologic disorder in infancy. Mutations withinside the gene encoding for the enzyme AADC (DDC) cause a intense mixed deficiency of serotonin and the 2 catecholamines dopamine and norepinephrine (Brun et al., 2010, p. 67). Primary congenital glaucoma (Glaucoma 3) is also Human diseases, and it is Congenital malformations of eye. GLC3A) CYP1B1 and (GLC3D) LTBP2. Primary congenital glaucoma (PCG) is a intense shape of glaucoma that offers early in life. PCG effects from developmental abnormalities that have an effect on the aqueous humor outflow pathway. PCG scientific capabilities consist of multiplied IOP, corneal edema, expansion of the globe (buphthalmos), corneal expansion, rupture of Descemets membrane, and optic nerve damage. Two genes had been said to reason PCG, CYP1B1 and LTBP2. Both genes reason a recessive shape of this disease (Azmanov et al., 2010,

p. 328). 3-Hydroxyacyl-CoA dehydrogenase deficiency known also as HADH deficiency or SCHAD deficiency. It is a human disease and it is a congenital disorders of metabolism and mitochondrial diseases. Gene associated with this pathway is hydroxyacyl-CoA dehydrogenase (HADH). -Hydroxyacyl-CoA dehydrogenase (HADH, SCHAD) deficiency is an autosomal recessive metabolic ailment, on account of mutations withinside the HADH gene. HADH deficiency is one of the mitochondrial fatty acid oxidation ailment that has been the maximum these days defined best in some sufferers. The medical phenotype of maximum sufferers which have been defined is recurrent hypoglycemia related to hyperinsulinism.

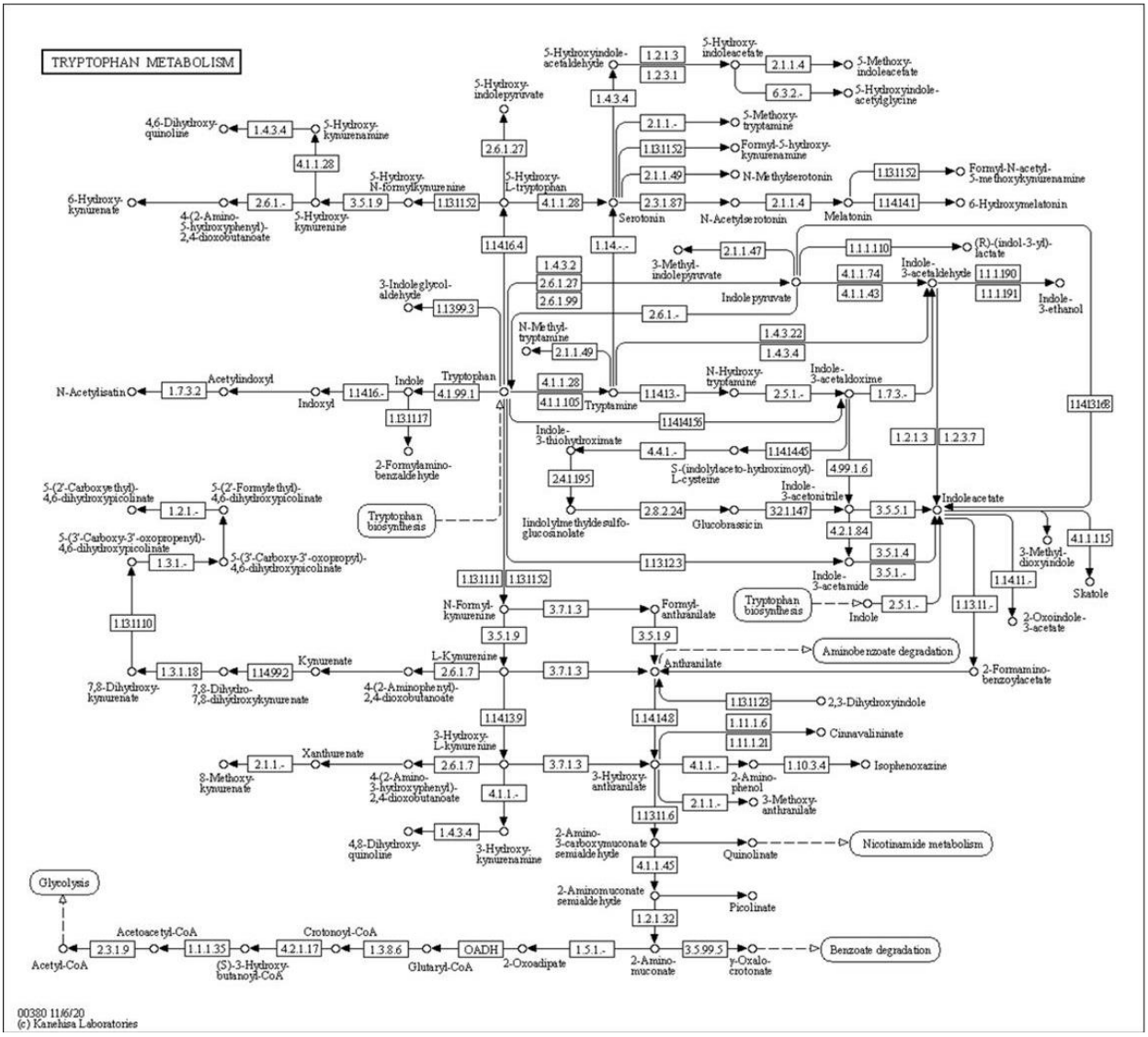


Figure 15: Tryptophan amino acid metabolomic pathway that is associated with non-pregnant samples.

In addition, I was able to identify another metabolomic pathway for pregnant samples using a pregnant compound, ((R)-2,3-dihydroxy-3-methylbutanoic acid, $C_5H_{10}O_4$) that identified in M11 samples only. This metabolomic pathway was identified in both Metabolica and KEGG pathways databases. (R)-2,3-dihydroxy-3-methylbutanoic acid is considered involved in branched chain amino acid biosynthesis pathway like valine, leucine, and isoleucine biosynthesis.

Interestingly we also detected 16 other compounds in our metabolomic profile that also involved in this specific pathway.

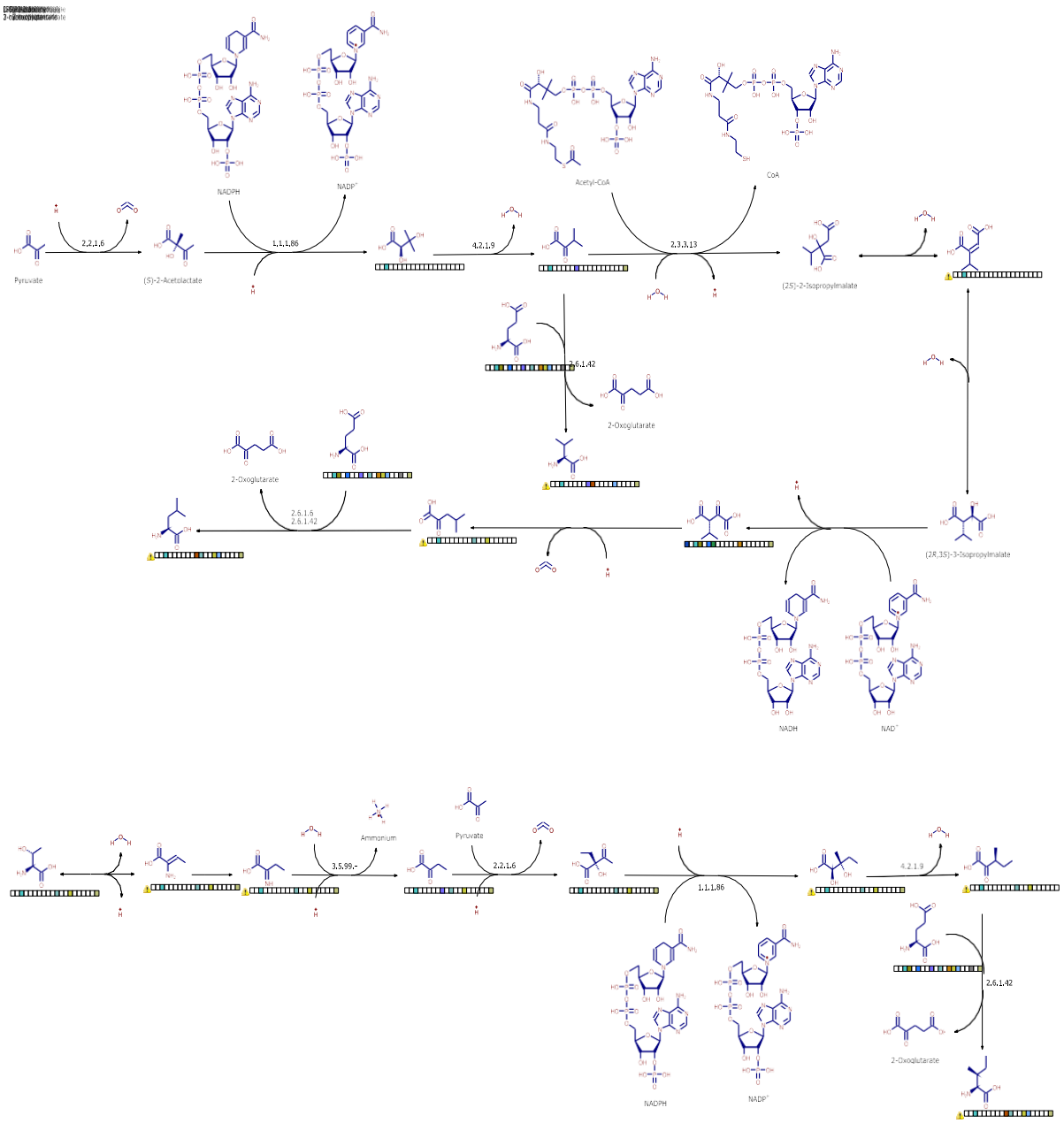


Figure 16: branched chain amino acid biosynthesis pathway like valine, leucine, and isoleucine biosynthesis that is associated with the pregnant sample that has 16 identified pathway compounds

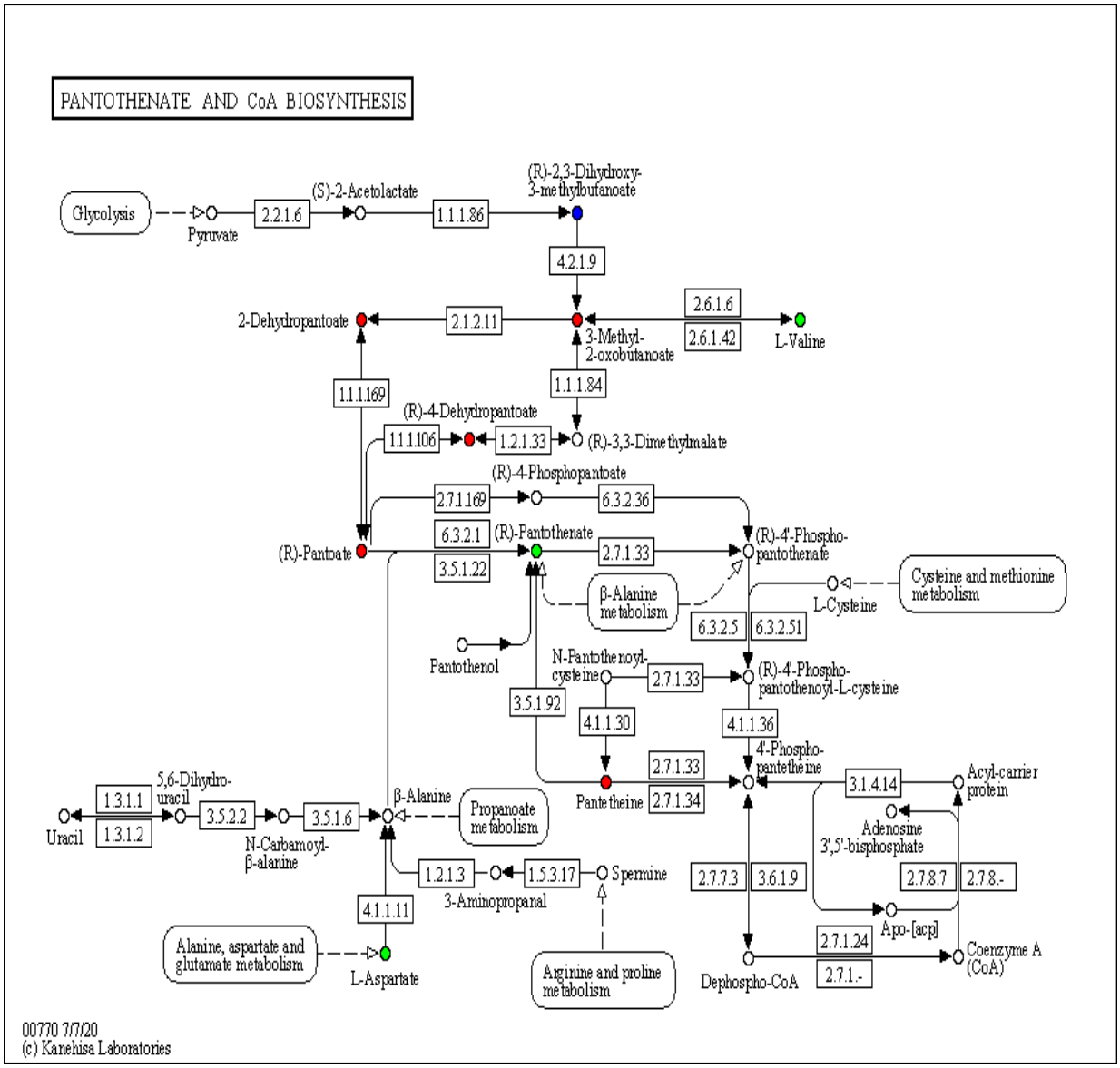


Figure 17: Metabolism of cofactors and vitamins

CHAPTER IV: CONCLUSION

In this study, I used a non-invasive method to evaluate the embryo quality using the culture media. I used UHPLC- Orbitrap Fusion Tribrid mass spectrometer to analyze the metabolomic samples, which enhanced the metabolite separation, mass accuracy, and resolution. We used Dynamic Data Acquisition to get fragmentation for 72% of the total compound detected. For our statistical analysis, I used a combination of PCA, volcano plots, and Box Whisker chart. Using PCA was not enough analysis since there are not any clustered samples for pregnant non-nonpregnant. However, it helps us in identifying three outlier samples. Volcano plots enabled me to recognize nonpregnant compounds associated with non-pregnant samples and down-regulated compounds associated with pregnant samples. I used the Box Whisker chart to confirm those identified compounds are significant biomarkers. Then, I structure elucidated those biomarkers using mzClould and FISh coverage.

One significant finding of this study is detecting one pathway associated with a down-regulated compound, (R)-2,3-dihydroxy-3-methylbutanoic acid. The metabolic result shows that 16 other compounds were found and identified in our samples: (3-methyl-2-oxobutatanoate, L-Glutamate, L-valine, 2-Isopropylmaleate, (2S)-2-Isopropyl-3oxosuccinate, 4-Methyl-2-oxopentatnoate, and the final product, which is L-leucine), (2-3-Dihydroxy-3-methylpentanoate, (S)-2-Aceto-2-hydroxybutanoate, 2-Oxobutanoate, 2-iminobutanoate, (2Z)-2-Aminobut-2-enoate and final product L-threonine) and L-Isoleucine. When I use the KEGG database, it provides us with the same information and more pathways (2S)-2-Isopropyl-3oxosuccinate, 4-Methyl-2-oxopentatnoate in like pantothenate and CoA biosynthesis. This could imply that this compound is a possible key biomarker of pregnancy.

One of my study limitations is that I had only 71 samples and only two types of culture media. So, it is not sufficient to call pregnant and nonpregnant compounds that I detected universal biomarkers for all IVF culture media. Especially that we cannot compare the similarities and differences between different kinds of culture media because IVF companies who make the culture media, they do not reveal the culture media compositions.

Another limitation to this study is that metabolites differ in polarity. Some are very polar compounds and others are nonpolar. Since we run our samples in positive mode and in a reversed-phase column, we can detect compounds that are nonpolar. This technique is a very poor technique to detect polar compounds. However, using UPLC in the negative mode is not beneficial due to the UPLC column is made from silica which will be affected if we use (0.1%) volatile base to promote deprotonation.

Also, that the scarcity of knowledge about human embryo metabolomic pathways and limited metabolomic pathways in databases make it hard for us to detect possible pathways. And even harder to connect the embryos' metabolomic to the embryos' genomics.

In conclusion, the metabolomic profiling of preimplantation of human IVF could be a potential tool to evaluate the embryo capability of implantation and also could be used to predict the aneuploidy of the embryo which is a major cause of pregnancy loss. Also, new methods can be improved to detect polar metabolites. Allowing us to detect polar and nonpolar metabolomic biomarkers.

REFERENCES

Ahlström, A., Wikland, M., Rogberg, L., Barnett, J. S., Tucker, M., & Hardarson, T. (2011). Cross-validation and predictive value of near-infrared spectroscopy algorithms for day-5 blastocyst transfer. Reproductive BioMedicine Online (Reproductive Healthcare Limited), 22(5), 477–484.

Anderlid, B., & Bui, T. (2011). Clinical perinatal genetics. Seminars in Fetal and Neonatal Medicine, 16(2), 69-69. doi:10.1016/j.siny.2011.01.002.

Azmanov, D. N., Dimitrova, S., Florez, L., Cherninkova, S., Draganov, D., Morar, B., Saat, R., Juan, M., Arostegui, J. I., Ganguly, S., Soodyall, H., Chakrabarti, S., Padh, H., López-Nevot, M. A., Chernodrinska, V., Anguelov, B., Majumder, P., Angelova, L., Kaneva, R., . . . Kalaydjieva, L. (2010). LTBP2 and CYP1B1 mutations and associated ocular phenotypes in the Roma/Gypsy founder population. *European Journal of Human Genetics*, *19*(3), 326–333. https://doi.org/10.1038/ejhg.2010.181

Boivin, J., Bunting, L., Collins, J. A., & Nygren, K. G. (2009). Reply: International estimates on infertility prevalence and treatment seeking: potential need and demand for medical care. Human Reproduction, 24(9), 2380–2383.

Brun, L., Ngu, L. H., Keng, W. T., Ch'ng, G. S., Choy, Y. S., Hwu, W. L., Lee, W. T., Willemsen, M. A. A. P., Verbeek, M. M., Wassenberg, T., Regal, L., Orcesi, S., Tonduti, D., Accorsi, P., Testard, H., Abdenur, J. E., Tay, S., Allen, G. F., Heales, S., . . . Blau, N. (2010). Clinical and biochemical features of aromatic L-amino acid decarboxylase deficiency. *Neurology*, *75*(1), 64–71. https://doi.org/10.1212/wnl.0b013e3181e620ae

Dettmer, K. (2007). *Mass spectrometry-based metabolomics.* John Wiley & Sons Inc. doi:10.1002/mas.20108

Fountain, S. (2011). In-Vitro Fertilization, 3rd edition edited by Kay Elder, Brian Dale. Obstetrician & Gynaecologist, 13(4), E279.

Góth, L. (2001). A New Type of Inherited Catalase Deficiencies: Its Characterization and Comparison to the Japanese and Swiss Type of Acatalasemia. Blood Cells, Molecules, and Diseases, 27(2), 512–517. https://doi.org/10.1006/bcmd.2001.0415

Kovac, J. R., Pastuszak, A. W., & Lamb, D. J. (2013). The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility. Fertility & Sterility, 99(4), 998–1007.

Kupka, M. S., Ferraretti, A. P., de Mouzon, J., Erb, K., D'Hooghe, T., Castilla, J. A., … Goossens, V. (2014). Assisted reproductive technology in Europe, 2010: results generated from European registers by ESHRE†. Human Reproduction, 29(10), 2099–2113.

Makarov, A., & Scigelova, M. (2010). Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A, 1217(25), 3938–3945.

Martins, E., Cardoso, M. L., Rodrigues, E., Barbot, C., Ramos, A., Bennett, M. J., Teles, E. L., & Vilarinho, L. (2011). Short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency: the clinical relevance of an early diagnosis and report of four new cases. Journal of Inherited Metabolic Disease, 34(3), 835–842.

Ménézo, Y., Lichtblau, I., & Elder, K. (2013). New insights into human pre-implantation metabolism in vivo and in vitro. Journal of Assisted Reproduction & Genetics, 30(3), 293–303.

Nicolson, M., & Fleming, J. E. E. (2013). Imaging and imagining the fetus: The development of obstetric ultrasound. Baltimore: Johns Hopkins University Press.

Pagidas, K., Ying, Y., & Keefe, D. (2008). Predictive value of pre-implantation genetic diagnosis for aneuploidy screening in repeated IVF-ET cycles among women with recurrent implantation failure. Journal of Assisted Reproduction & Genetics, 25(2/3), 103–106.

Qin, J., Wang, H., Sheng, X., Liang, D., Tan, H., & Xia, J. (2015). Pregnancy-related complications and adverse pregnancy outcomes in multiple pregnancies resulting from assisted reproductive technology: a meta-analysis of cohort studies. Fertility & Sterility, 103(6), 1492–1508.e7.

Rødgaard, T., Heegaard, P. M. H., & Callesen, H. (2015). Non-invasive assessment of in-vitro embryo quality to improve transfer success. *Reproductive BioMedicine Online (Elsevier Science)*, *31*(5), 585–592.

Ropers, H. H., & Hamel, B. C. J. (2005). X-linked mental retardation. Nature Reviews Genetics, 6(1), 46–57. https://doi.org/10.1038/nrg1501

Sanchez, T., Seidler, E. A., Gardner, D. K., Needleman, D., & Sakkas, D. (2017). Will non-invasive methods surpass invasive for assessing gametes and embryos? *Fertility & Sterility*, *108*(5), 730–737.

Seli, E., Bruce, C., Botros, L., Henson, M., Roos, P., Judge, K., … Sakkas, D. (2011). Receiver operating characteristic (ROC) analysis of day 5 morphology grading and metabolomic Viability Score on predicting implantation outcome. Journal of Assisted Reproduction & Genetics, 28(2), 137–144.

Shi, H., Enriquez, A., Rapadas, M., Martin, E. M., Wang, R., Moreau, J., Lim, C. K., Szot, J. O., Ip, E., Hughes, J. N., Sugimoto, K., Humphreys, D. T., McInerney-Leo, A. M., Leo, P. J., Maghzal, G. J., Halliday, J., Smith, J., Colley, A., Mark, P. R., . . . Dunwoodie, S. L.

(2017a). NAD Deficiency, Congenital Malformations, and Niacin Supplementation. New England Journal of Medicine, 377(6), 544–552.

Spinella, F., Fiorentino, F., Biricik, A., Bono, S., Ruberti, A., Cotroneo, E., Baldi, M., Cursio, E., Minasi, M. G., & Greco, E. (2018). Extent of chromosomal mosaicism influences the clinical outcome of in vitro fertilization treatments. Fertility and sterility, 109(1), 77 –83. https://doi.org/10.1016/j.fertnstert.2017.09.025

Thompson, J. G., Brown, H. M., & Sutton-McDowall, M. L. (2016). Measuring embryo metabolism to predict embryo quality. Reproduction, Fertility & Development, 28(1/2), 41–50.

Thompson, J. G., Brown, H. M., & Sutton-McDowall, M. L. (2016). Measuring embryo metabolism to predict embryo quality. Reproduction, Fertility & Development, 28(1/2), 41–50.

udakalakatti, S. M. (2013). NMR studies of pre-implantation embryo metabolism in human assisted reproductive techniques: A new biomarker for assessing embryo implantation potential assessment of embryo implantation potential. John Wiley & Sons Inc.

LIST OF APPENDICES

APPENDICES PAGE

Appendix 1: M2 M2 samples pregnant compounds that we can identify using MF calculator and python code.....35

Appendix 2: M2M2 samples nonpregnant compounds that we're able to identify using MF calculator and python code, compounds that are highlighted are matching Compound Discoverer results.....36

Appendix 3: M11 samples pregnant compounds before removing the outlier that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....37

Appendix 4: M11 samples nonpregnant compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....38

Appendix 5: M11 samples pregnant compounds after removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....39

Appendix 6: : M11 samples nonpregnant compounds After removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....40

Appendix 7: All samples Down-regulated compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....41

Appendix 8: All samples UP regulated compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....42
Appendix 9: All samples Down-regulated compounds after removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....43
Appendix 10: All samples UP-regulated compounds after removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.....44
Appendix 11: Structures for up-regulated compounds M2M2 samples.....45
Appendix 12: Structures for down-regulated compounds M2M2 samples.....46
Appendix 13: Structures for up-regulated compounds M11 samples.....47
Appendix 14: Structures for down-regulated compounds M12 samples.....48
Appendix 15: Structures for up-regulated compounds all samples.....50
Appendix 16: Structures for down-regulated compounds all samples.....51
Appendix 17: FISh spectra for two identified nonpregnant compounds in M2M2 samples.....52
Appendix 18: FISh spectra for two identified pregnant compounds in M2M2 samples.....53
Appendix 19: FISh spectra for two identified nonpregnant compounds in M11 samples.....54
Appendix 20: FISh spectra for two identified pregnant compounds in M11 samples.....55
Appendix 21: : FISh spectra for two identified nonpregnant compounds in all samples.....56

APPENDICES

Appendix 1: M2M2 samples pregnant compounds that we can identify using MF calculator and python code.

Compound number	Molecular weight	Molecular Formula	Compound number	Molecular weight	Molecular Formula
1	279.0538	$C_{16}H_9NO_4$	8	190.1727	$C_{14}H_{22}$
2	160.0892	$C_{11}H_{12}O$	9	178.0998	$C_{11}H_{14}O_2$
3	240.9929	$C_{12}H_4NO_3P$	10	152.1205	$C_{10}H_{16}O$
4	142.0634	$C_7H_{10}O_3$	11	208.1832	$C_{14}H_{24}O$
5	649.628	$C_{44}H_{79}N_3$	12	282.2202	$C_{17}H_{30}O_3$
6	651.6244	$C_{36}H_{77}N_9O$	13	249.2099	$C_{16}H_{27}NO$
7	292.9968	$C_{16}H_7NOS_2$			

Appendix 2: M2M2 samples nonpregnant compounds that we're able to identify using MF calculator and python code, which are highlighted, match compound discoverer result.

Compound number	Molecular weight	Molecular Formula	Compound number	Molecular weight	Molecular Formula
1	194.1157	$C_8 H_{18} O_5$	10	175.1001	$C_{11} H_{13} N O$
2	899.8059	$C_{49} H_{105} N_9 O_3 S$	11	201.1732	$C_{11} H_{23} N O_2$
3	297.1793	$C_{12} H_{27} N O_7$	12	261.1371	$C_{15} H_{19} O_3 S_3$
4	145.1106	$C_7 H_{15} N O_2$	13	148.1103	$C_7 H_{16} O_3$
5	275.2102	$C_{14} H_{29} N O_4$	14	222.0570	$C_9 H_{10} N_4 O$
6	300.1555	$C_{19} H_{24} O S$	15	201.1732	$C_{11} H_{23} N O_2$
7	832.8220	$C_{51} H_{104} N_6 O_2$	16	278.0909	$C_{14} H_{18} N_2$
8	304.0418	$C_{16} H_8 N_4 O S$	17	214.0747	$C_{12} H_{10} N_2 O_2$
9	281.0330	$C_{16} H_{11} N$			

Appendix 3: M11 samples pregnant compounds before removing the outlier that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.

Compound number	Molecular weight	Molecular Formula
1	283.2002	$C_{13} H_{25} N_5 O_2$
2	284.2035	$C_{14} H_{28} N_4 S$
3	142.0633	$C_7 H_{10} O_3$
4	269.1846	$C_{12} H_{23} N_5 O_2$
5	239.1739	$C_{11} H_{21} N_5 O$
6	243.2569	$C_{15} H_{33} N$

Appendix 4: M11 samples nonpregnant compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.

Compound number	Molecular weight	Molecular Formula
1	118.0786	$C_9 H_{10}$
2	131.0739	$C_9 H_9 N$
3	229.1320	$C_{11} H_{19} N O_4$
4	336.0728	$C_{11} H_{17} N_2 O_8 P$
5	275.1741	$C_{14} H_{29} N S_2$

Appendix 5: M11 samples pregnant compounds After removing the outliers, we can identify using the MF calculator and python code, which are highlighted match compound discoverer results.

Compound number	Molecular weight	Molecular Formula	Compound number	Molecular weight	Molecular Formula
1	283.2001	$C_{13}H_{25}N_5O_2$	13	207.1629	$C_{13}H_{21}NO$
2	218.0797	$C_9H_{14}O_6$	14	173.1421	$C_9H_{19}NO_2$
3	297.1796	$C_{12}H_{27}NO_7$	15	142.0633	$C_7H_{10}O_3$
4	239.1739	$C_{11}H_{21}N_5O$	16	245.1634	$C_{12}H_{23}NO_4$
5	311.2317	$C_{15}H_{29}N_5O_2$	17	211.1578	$C_{12}H_{21}NO_2$
6	388.1433	$C_{13}H_{29}N_2O_7SP$	18	225.1734	$C_{13}H_{23}NO_2$
7	157.1106	$C_8H_{15}NO_2$	19	229.1685	$C_{12}H_{23}NO_3$
8	244.1681	$C_{13}H_{24}O_4$	20	284.2035	$C_{14}H_{28}N_4S$
9	239.1892	$C_{14}H_{25}NO_2$	21	243.1840	$C_{13}H_{25}NO_3$
10	269.1846	$C_{12}H_{23}N_5O_2$	22	166.0999	$C_{10}H_{14}O_2$
11	169.1107	$C_9H_{15}NO_2$	23	389.1274	$C_{13}H_{28}NO_8SP$
12	173.1421	$C_9H_{19}NO_2$			

\

Appendix 6: M11 samples nonpregnant compounds After removing the outliers, we can identify using the MF calculator and python code, which are highlighted match compound discoverer results.

Compound number	Molecular weight	Molecular calculator	Compound number	Molecular weight	Molecular calculator
1	118.0786	$C_9 H_{10}$	7	302.0673	$C_9 H_{18} O_9 S$
2	131.0739	$C_9 H_9 N$	8	299.0589	$C_{19} H_9 N O_3$
3	401.1199	$C_{12} H_{24} N_3 O_{10} P$	9	336.0728	$C_{11} H_{17} N_2 O_7 P$
4	229.1320	$C_{11} H_{19} N O_4$	10	360.1328	$C_{16} H_{25} O_7 P$
5	300.0673	$C_{13} H_{16} O_6 S$	11	316.1092	$C_{13} H_{20} N_2 O_5 S$
6	172.0641	$C_{10} H_8 N_2 O$			

Appendix 7: All samples pregnant compounds before removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching compound discoverer result.

Molecular weight	Molecular weight	Molecular Formula
1	244.1293	$C_{16} H_{20} S$
2	282.2202	$C_{17} H_{30} O_3$
3	152.1205	$C_{10} H_{16} O$

Appendix 8: All samples of nonpregnant compounds before removing the outliers that we're able to identify using the MF calculator and python code, which are highlighted, match compound discoverer results.

Compound number	Molecular weight	Molecular Formula	Compound number	Molecular weight	Molecular Formula
1	118.0786	$C_9 H_{10}$	6	205.9884	$C_6 H_6 O_6 S$
2	217.1067	$C_8 H_{15} N_3 O_4$	7	177.0794	$C_{10} H_{11} N O_2$
3	133.0895	$C_9 H_{11} N$	8	252.1216	$C_{11} H_{24} O_2 S_2$
4	118.0786	$C_9 H_{10}$	9	277.0381	$C_{16} H_7 N O_4$
5	140.0477	$C_7 H_8 O_3$	10	297.1796	$C_{12} H_{27} N O_7$

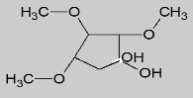
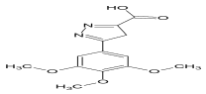
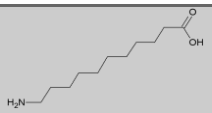
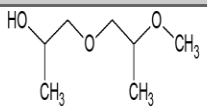
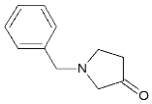
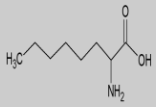
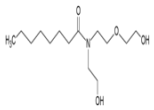
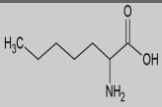
Appendix 9: All samples pregnant compounds After removing the outliers that we can identify using MF calculator and python code, highlighted compounds match compound discoverer result.

Compound number	Molecular weight	Molecular Formula	Compound number	Molecular weight	Molecular Formula
1	120.0943	C_9H_{12}	7	152.1205	$C_{10}H_{16}O$
2	244.9488	$C_{16}H_{29}NO_4$	8	112.0892	$C_7H_{12}O$
3	249.2100	$C_{16}H_{27}NO$	9	193.9731	$C_3H_3N_2O6P$
4	282.2202	$C_{17}H_{30}O_3$	10	183.9788	$C_4HN_4O_3P$
5	191.9727	$C_5H_4O_6S$	11	140.9729	CH_3NO_5S
6	249.2100	$C_{16}H_{27}N\,O$			

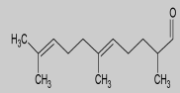
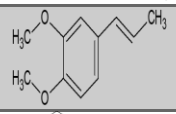
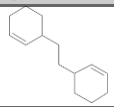
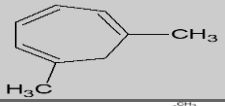
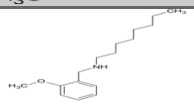
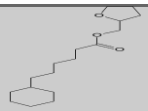
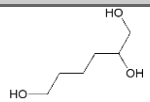
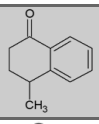
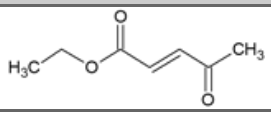
Appendix 10: All samples nonpregnant compounds After removing the outliers that we're able to identify using MF calculator and python code, compounds that are highlighted are matching

Compound number	Molecular weight	Molecular Formula	Compound number	Molecular weight	Molecular Formula
1	278.0909	$C_8H_{11}N_{10}P$	7	277.0375	$C_{16}H_7NO_4$
2	297.1796	$C_{12}H_{27}NO_7$	8	118.0786	$C9H_{10}$
3	186.0645	$C_7H_{10}N_2O_4$	9	133.0895	$C_9H_{11}N$
4	252.1215	$C_{11}H_{24}O_2S_2$	10	194.1159	$C_8H_{18}O_5$
5	140.0477	$C_7H_8O_3$	11	118.0786	C_9H_{10}
6	177.0794	$C_{10}H_{11}NO_2$			

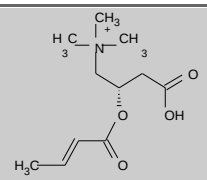
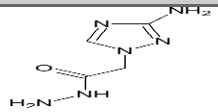
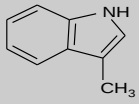
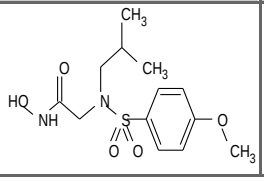
Appendix 11: Structures for nonpregnant compounds M2M2 samples

Compound name	Compound formula	Compound structure	ΔMass(ppm)	FISh scoring%
2,3,4-Tri-O-methyl-L-three-pentitol	$C_8H_{18}O_5$		1.41	75.86
5-(3,4,5-Trimethoxyphenyl)-4H-pyrazole-3-carboxylic acid	$C_{13}H_{14}N_2O_5$		2.15	75.26
11-Aminoundecanoic acid	$C_{11}H_{23}NO_2$		1.54	94.44
1-(2-Methoxypropoxy)-2-propanol	$C_7H_{16}O_3$		2.30	76.92
3-(4-Butoxyphenyl)-3-methyl-2,5-pyrrolidinedione	$C_{15}H_{19}NO_3$		2.21	89.80
N-benzyl-3-pyrrolidinone	$C_{11}H_{13}NO$		2.03	77.78
(±)-2-amino-octanoic acid	$C_8H_{17}NO_2$		3.00	93.65
N-[2-(2-Hydroxyethoxy)ethyl]-N-(2-hydroxyethyl)octanamide	$C_{14}H_{29}NO_4$		1.83	96.15
3-Aminoheptanoic acid	$C_7H_{15}NO_2$		2.83	88.73

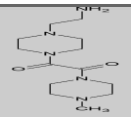
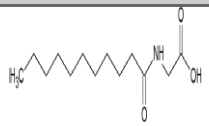
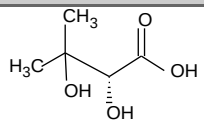
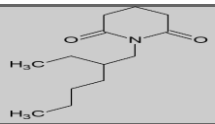
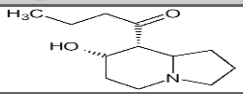
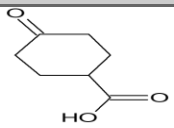
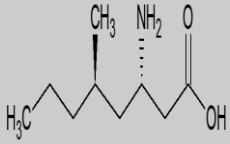
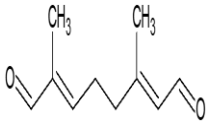
Appendix 12: Structures for pregnant compounds M2M2 samples

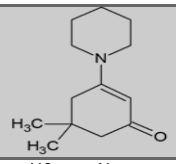
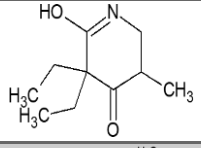
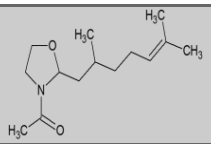
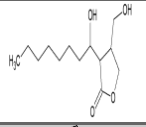
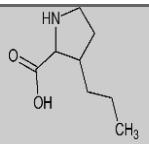
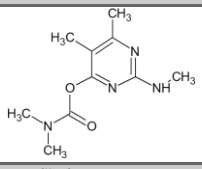
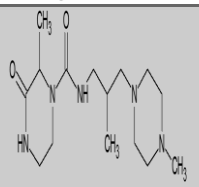
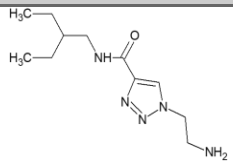
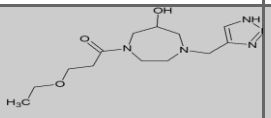
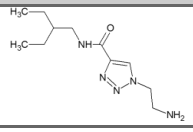
Compound name	Compound formula	Compound structure	ΔMass(ppm)	FISh scoring
2,6,10-Trimethyl-5,9-undecadienal	$C_{14}H_{24}O$		2.46	84.81
2-Hexylfuran	$C_{10}H_{16}O$		3.0	89.47
(E)-methyl isoeugenol	$C_{11}H_{14}O_2$		2.47	82.35
3,3'-Ethane-1,2-diyldicyclohexene	$C_{14}H_{22}$		2.61	70.31
1,6-Dimethyl-1,3,5-cycloheptatriene	C_9H_{12}		3.19	76.67
N-(2-Methoxybenzyl)-1-octanamine	$C_{16}H_{27}N_O$		2.43	86.81
Tetrahydro-2-furanylmethyl 6-cyclohexylhexanoate	$C_{17}H_{30}O_3$		2.50	87.32
MO4650000	$C_6H_{14}O_3$		2.92	85.82
4-Methyl-1-Tetralone	$C_{11}H_{12}O$		2.54	77.78
Ethyl (2E)-4-oxo-2-pentenoate	$C_7H_{10}O_3$		2.66	97.96

Appendix 13: Structures for nonpregnant compounds M11 samples

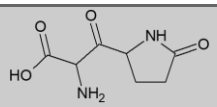
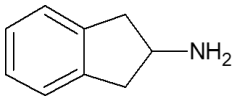
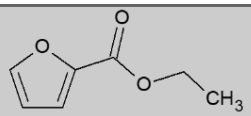
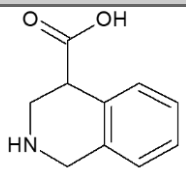
Compound name	Compound formula	Compound structure	ΔMass(ppm)	FISh scoring
Butenylcarnitine	$C_{11}H_{20}NO_4$		2.77	95.56
L-Prolyl-L-leucine	$C_4H_8N_6O$		3.72	100
Skatole	C_9H_9N		2.94	48.35
N-Isobutyl-N-(4-methoxyphenylsulfonyl)glycyl hydroxamic acid	$C_{13}H_{20}N_2O_5S$		-0.17	82.19

Appendix 14: Structures for pregnant compounds M1 samples

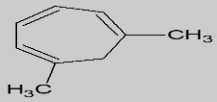
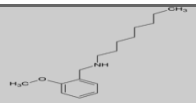
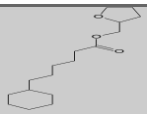
Compound name	**Compound formula**	**Compound structure**	**ΔMass(ppm)**	**FISh scoring**
1-[4-(2-Aminoethyl)-1-piperazinyl]-2-(4-methyl-1-piperazinyl)-1,2-ethanedione	$C_{13}H_{25}N_5O_2$		-2.44	73.08
N-Undecanoylglycine	$C_{13}H_{25}NO_3$		2.44	83.87
2-(decanoylamino)acetic acid	$C_{12}H_{23}NO_3$		2.95	88.34
(R)-2,3-dihydroxy-3-methylbutanoic acid	$C_5H_{10}O_4$		3.63	60.53
1-(2-Ethylhexyl)-2,6-piperidinedione	$C_{13}H_{23}NO_2$		2.54	70.59
Elaeokanine C	$C_{12}H_{21}NO_2$		2.78	84.72
2,4-Heptadien-1-ol, (E, E)-	$C_7H_{12}O$		3.48	80.95
4-Oxocyclohexanecarboxylic acid	$C_7H_{10}O_3$		2.33	90.48
Imagabalin	$C_9H_{19}NO_2$		2.76	93.55
(6E)-8-oxogeranial	$C_{10}H_{14}O_2$		2.84	81.03

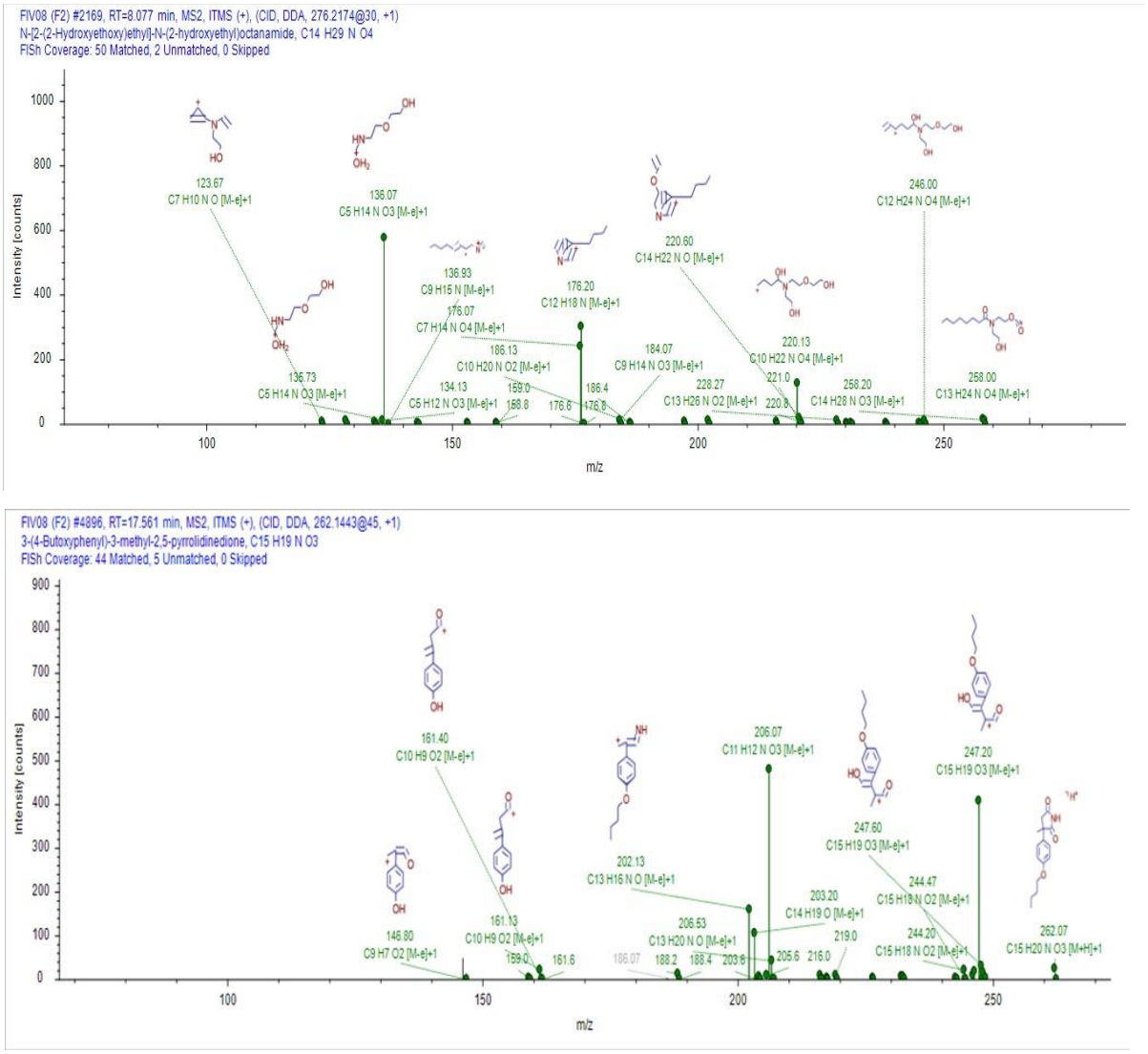
5,5-Dimethyl-3-piperidinocyclohex-2-enone	$C_{13}H_{21}NO$		2.82	85.37
Methyprylon	$C_{10}H_{17}NO_2$		3.01	96.30
3-acetyl-2-(2,6-dimethyl-5-heptenyl)oxazolidine	$C_{14}H_{25}NO_2$		2.78	87.23
2-(1'-Hydroxyoctyl)-3-hydroxymethylbutanolide	$C_{13}H_{24}O_4$		2.81	86.67
3-Propylproline	$C_8H_{15}NO_2$		2.56	81.90
MFCD01940516	$C_{10}H_{16}N_4O_2$		3.03	75.41
2-Methyl-2-propanyl (2,3-dimethyl-2-{[(1-methyl-1H-1,2,3-triazol-4-yl)methyl]amino}butyl)carbamate	$C_{15}H_{29}N_5O_2$		-1.34	85.00
1-(2-Aminoethyl)-N-(2-ethylbutyl)-1H-1,2,3-triazole-4-carboxamide	$C_{11}H_{21}N_5O$		-2.94	83.75
3-Ethoxy-1-[6-hydroxy-4-(1H-1,2,3-triazol-4-ylmethyl)-1,4-diazepan-1-yl]-1-propanone	$C_{13}H_{23}N_5O_3$		-1.60	100.00
Tricarballylic acid trimethyl ester	$C_9H_{14}O_6$		2.82	85.51

Appendix 15: Structures for nonpregnant compounds all samples

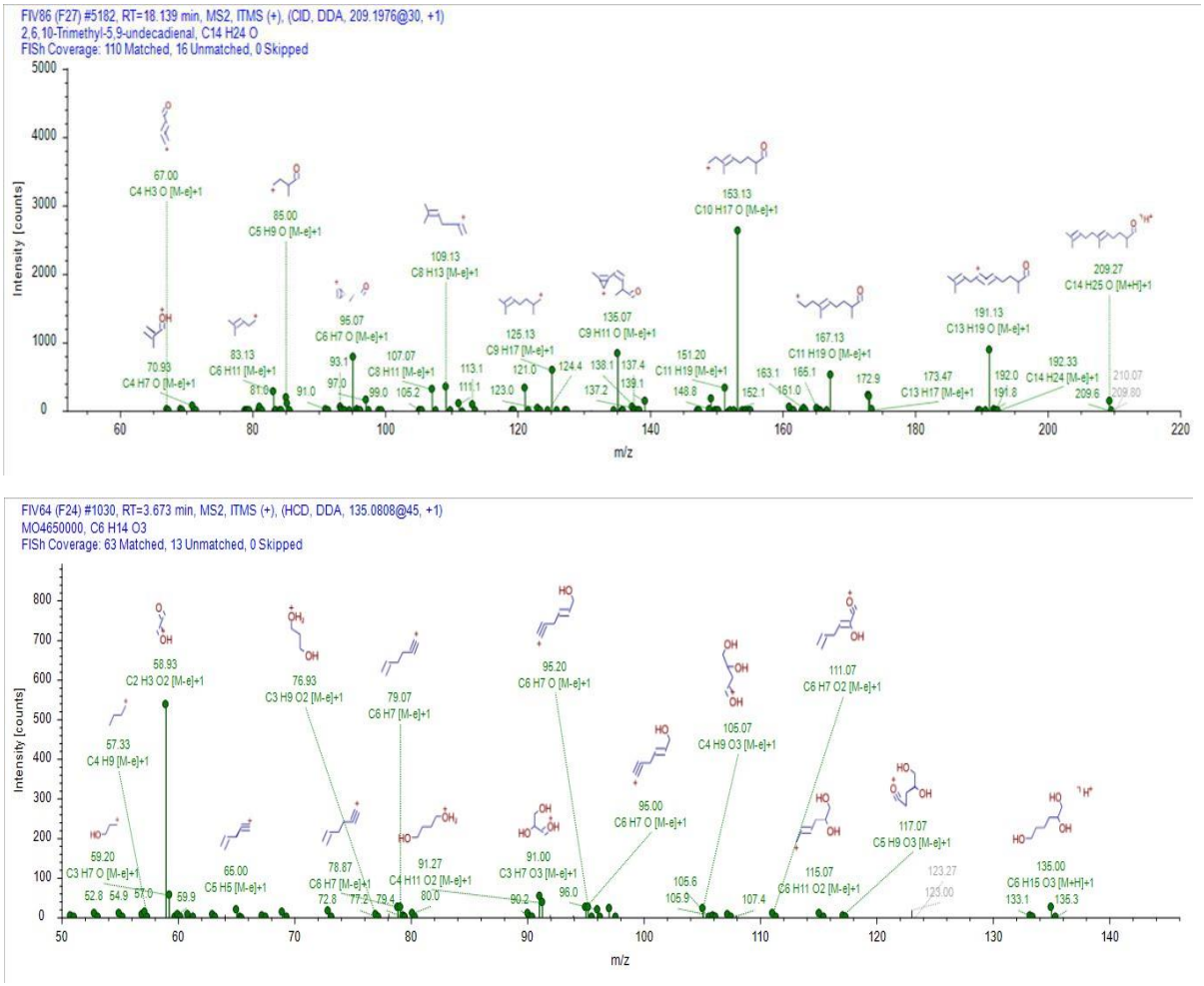
Compound name	Compound formula	Compound structure	ΔMass(ppm)	FISh scoring
Pyroglutamylglycine	$C_7H_{10}N_2O_4$		2.42	90.16
2-Aminoindan	$C_9H_{11}N$		2.70	77.78
LV1850000	$C_7H_8O_3$		2.62	85.88
1,2,3,4-Tetrahydro-4-isoquinolinecarboxylic acid	$C_{10}H_{11}NO_2$		2.34	83.64

Appendix 16*: All the pregnant compounds in All samples group*

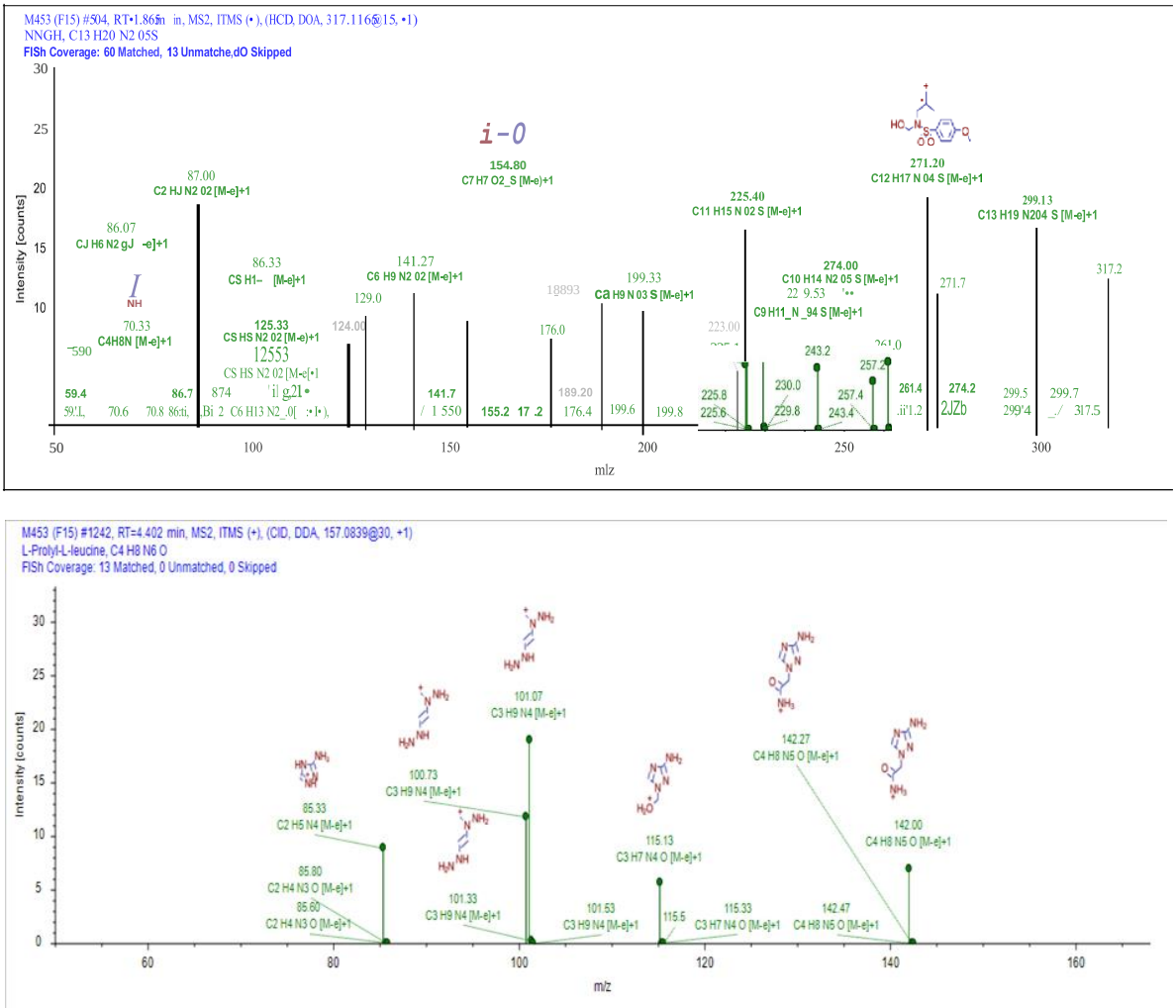
Compound name	Compound formula	Compound structure	ΔMass(ppm)	FISh scoring
1,6-Dimethyl-1,3,5-cycloheptatriene	C_9H_{12}		3.19	76.67
2-Hexylfuran	$C_{10}H_{16}O$		3.0	89.47
N-(2-Methoxybenzyl)-1-octanamine	$C_{16}H_{27}NO$		2.43	86.81
2,4-Heptadien-1-ol, (E, E)-	$C_7H_{12}O$		3.48	80.95
Tetrahydro-2-furanylmethyl 6-cyclohexylhexanoate	$C_{17}H_{30}O_3$		2.50	87.32



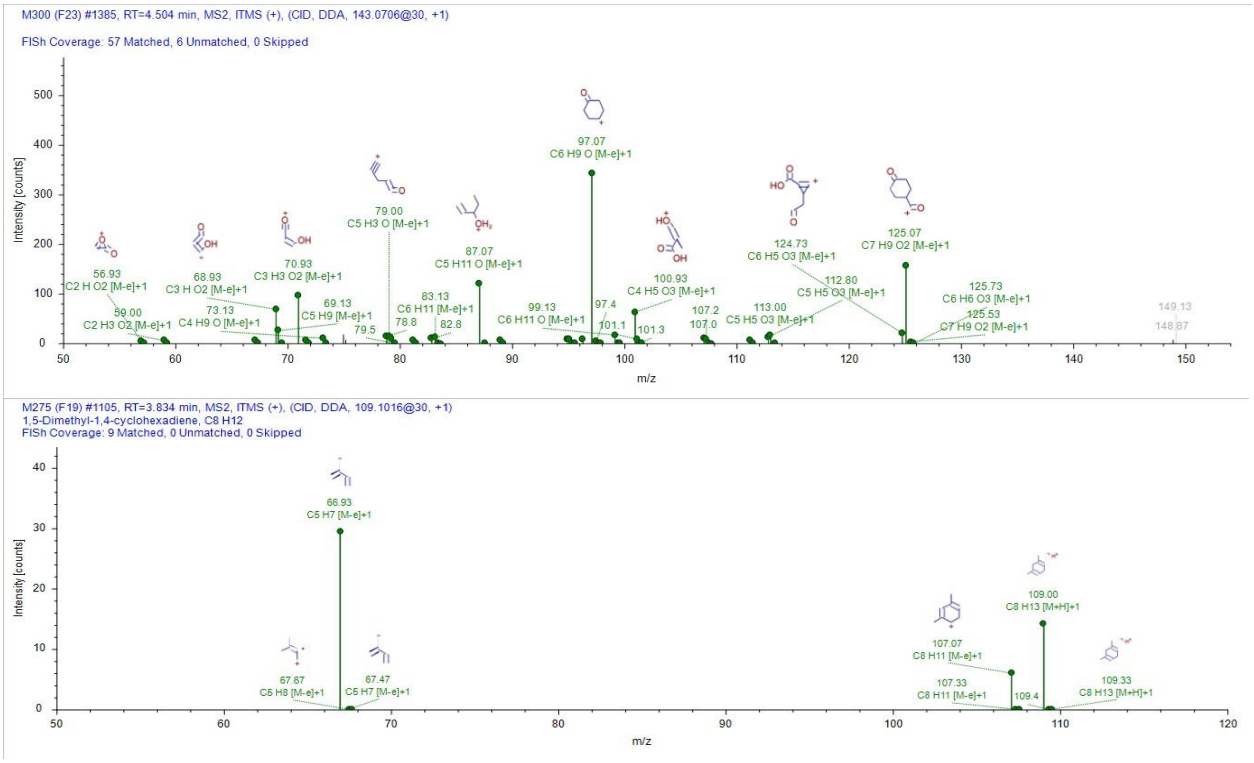
Appendix 17: FISh spectra for two identified nonpregnant compounds in M2M2 samples with high FISh coverage.



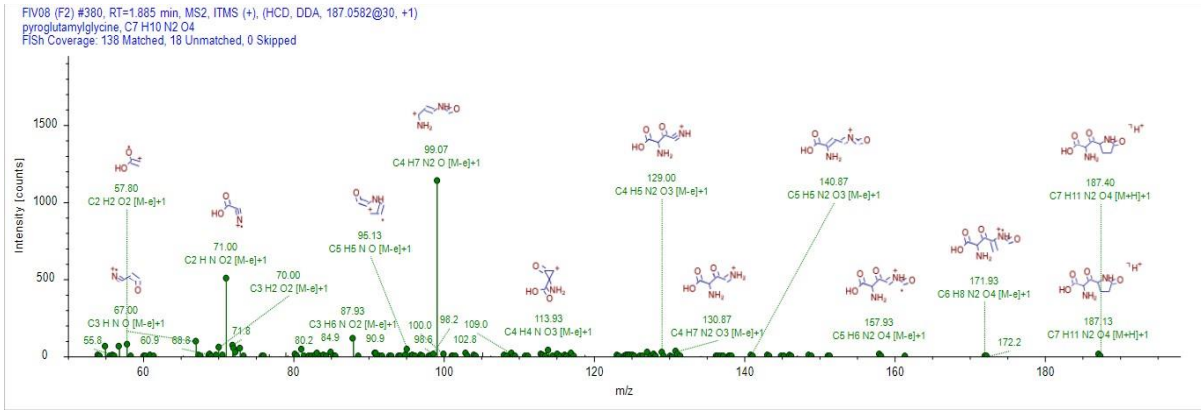
Appendix 18: FISh spectra for pregnant compounds in M2 M2 samples

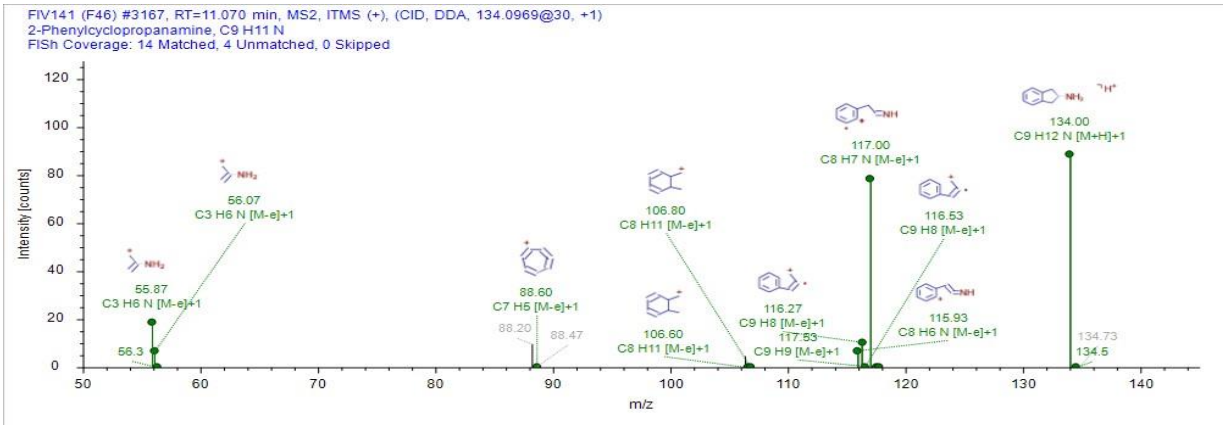


Appendix 19*: FISh spectra for nonpregnant compounds in M11 samples*



Appendix 20: some of FISh spectra for pregnant compounds in M11 samples





Appendix 21: FISh spectra for two nonpregnant compounds in all samples.