CHARACTERIZATION OF MESENCHYMAL STEM CELLS USING METABOLOMICS

by

XUNAN SHEN

(Under the Direction of Arthur S. Edison)

ABSTRACT

Mesenchymal stromal/stem cells (MSCs) are multipotent adult stem cells which show immunomodulation capacity and have been widely used in clinical trials. Despite their immunomodulation potential, to date, there is no FDA-approved MSC therapy available. MSCs show heterogeneity at different levels, including donor-donor variation, tissue-tissue variation and cell subpopulation variation. Therefore, it is difficult to predict MSC immunomodulation effects as well as predict outcome after MSC therapy. The lack of predictive markers which can indicate MSC immunomodulation effects limits its use in clinic.

In this dissertation, New methods were developed to characterize MSC immunomodulation effects using metabolomics. The cellular metabolomics and media cytokines were first investigated to predict MSC immunomodulation effects. The composite functional scores were generated from T cell suppression assay and IDO (Indoleamine 2,3-Dioxygenase) activity results. These composite scores were then predicted using partial least squares (PLSR) model using metabolomics and cytokines profiles. Several metabolites including phosphocreatine, NN-dimethylglycine, asparagine, myo-inositol and serine were found important in predicting composite functional score, therefore, can be used as surrogates to predict MSC immunomodulation effects. These methods are all destructive to cells. A non-destructive, in-process characterization of MSCs is needed. Time course monitoring of culture media can closely reflect cell phenotype and can be used to predict MSC function. Metabolites change rate of first three days culturing in culture media were found can be

used to predict MSC immunomodulation effects. These metabolites include leucine, threonine and alanine. Based on these findings, pathway analysis was also used to reveal some of the underlying mechanisms. The pathway analysis showed that cells consume aspartate to generate asparagine, meanwhile, secrete alanine and glycine to culture media. Based on these findings, Some preliminary results of the validation of the extracellular predictive markers we identified previously were also shown using additional nine bone marrow-derived MSC lines culturing in two different culture media. The bottleneck of MSC research is the application in cell manufacturing. The application of media monitoring using bioreactors and bench-top NMR in industrial setting were also discussed.

INDEX WORDS: [Mesenchymal stem cells, immunomodulation, metabolomics, NMR, predictive marker]

CHARACTERIZATION OF MESENCHYMAL STEM CELLS USING METABOLOMICS

by

XUNAN SHEN

B.S., Nanjing University of Chinese Medicine, China, 2015
M.S., University of Kansas, 2017

A [Dissertation] Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

[DOCTOR] OF [PHILOSOPHY]

ATHENS, GEORGIA

2021

CHARACTERIZATION OF MESENCHYMAL STEM CELLS USING METABOLOMICS

by

XUNAN SHEN

Major Professor: Arthur S. Edison

Committee: Steven L. Stice
Wandaliz Torres-Garcia
Jessica C. Kissinger
Liang Liu

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

DEDICATION

In memory of my maternal grandma, Gufang Liu.

Your love makes this work possible.

&

To my family.

ACKNOWLEDGMENTS

I joined Dr. Edison's lab in December 2017, since then, I've been cared and guided by my advisor Dr. Arthur S. Edison. I am so grateful to have such a wonderful mentor, who is kind, brilliant and critical at the same time. He helped me not only in my research, but also influenced me in my life, from who, I learned how to see and treat the world in a different way. I want to express my deepest appreciation to Dr. Edison, for his constant guidance and support throughout my PhD study, without him, this work would not have been possible. To the entire Edison lab members, I would like to thank you for your mental support in my PhD study.

I also want to express my appreciation to my committee members, Dr. Steven L. Stice, Dr. Wandaliz Torres-Garcia, Dr. Jessica C. Kissinger and Dr. Liang Liu. They provided me valuable advice on my projects throughout my PhD study. For their constant support and help, I say thank you. Special thanks to Dr. Liu, who helped me a lot in my secondary statistics master's program.

I would like to thank my collaborators, Dr. Steve Stice's laboratory and Dr. Ross Marklein's laboratory at University of Georgia, Dr. Facundo Fernandez's laboratory and Dr. Manu Platt's laboratory at Georgia Institute of Technology, for their help and advice in building up these projects.

Last but not least, I want to thank my family. My grandma, Gufang Liu, who raised me up with love, gave me courage to face the world. I want to thank my mom, Shuzhen Li, and my dad, Zijun Shen, for their love and support. I would also like to thank my husband, Xianfeng Liang, for his continued love and support throughout my PhD study.

CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 MSC and its anti-inflammatory effects

1.1.1 Mesenchymal stromal cells

Mesenchymal stem cells (MSCs) are pluripotent stem cells that can be isolated from various tissues including bone marrow, adipose tissue, umbilical cord, muscle, etc. MSCs show self-renewal potential and can differentiate into different cell types such as osteoblast, chondrocyte, myotube, stromal cells, fibroblast and adipocyte [1]. Other than differentiation function, MSCs also show anti-inflammatory capacity thus have been investigated in multiple clinical trials [2]. Until 2020, 732 MSC inflammation related clinical trials have been conducted, data obtained from clinicalgov (Figure 1.1.1). To date, no FDA approved MSC therapy available.

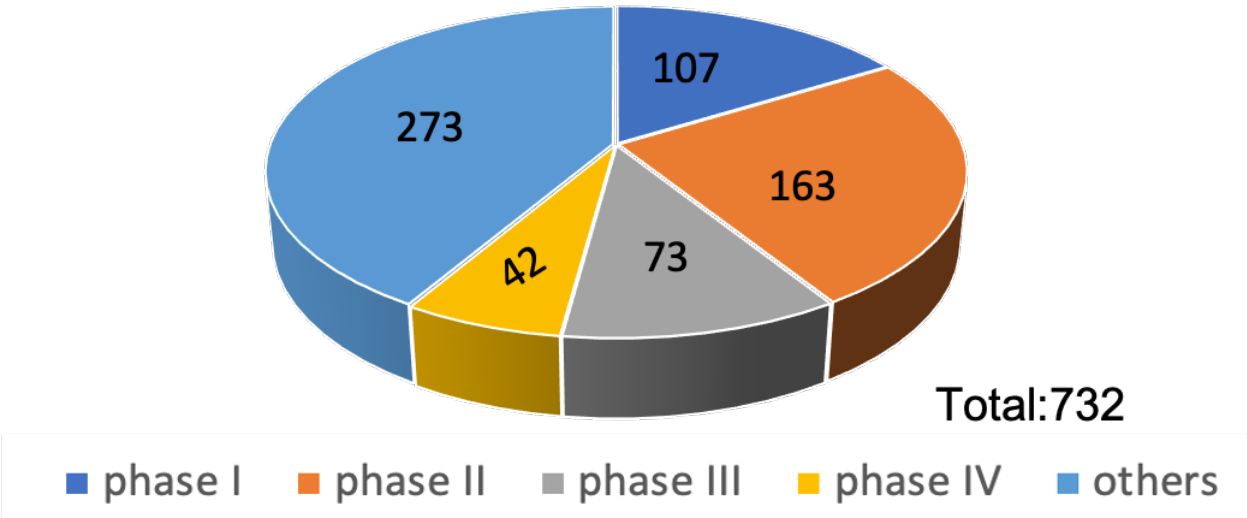


Figure 1.1.1: MSCs inflammation related clinical trials (data until 2020)

In 2006, the International Society for Cellular Therapy (ISCT) proposed minimal criteria of MSCs (Table 1.1), which enable researchers to distinguish MSCs from other type of cells [3].

Table 1.1: ISCT proposed MSCs minimal criteria

MSCs minimal criteria
Plastic-adherent when maintained in standard culture conditions
Express CD105, CD73 and CD90
Lack expression of CD45, CD34, CD14 or CD11b, CD79α or CD19 and HLA-DR
Differentiate to osteoblasts, adipocytes and chondroblasts in vitro

However, these criteria have been shown inadequate for characterizing MSCs [4]. Research indicate the negative expression of CD34 is an artifact of cell culture conditions [5–7]. Studies conducted by Simmons & Torok-Storb show the positive expression of CD34 in uncultured human bone marrow MSCs [8]. Later, they also show that CD34 positive human bone marrow cells can be used as immunogen to generate a panel of hybridomas [9]. In addition, several studies show that MSCs

express CD34 at the time of isolating from adipose tissue, but lose expression in culture [10–12]. The debate of MSC CD34 expression indicate markers may vary depending on different conditions.

1.1.2 MSC heterogeneity

MSCs show heterogeneity at different levels including donors, tissues and cell population [13–15]. Studies prove that MSCs show donor-donor variation because of the differences in donor age [16], donor sex [17], as well as physiological status [18], which result in functional differences in MSC. Advanced age is negatively correlating with MSC function [19–21]. Flyn Mckinnirey et al. found MSC potency differences between donors with different sex. They matched the age and health status of MSC donors, and found that MSCs from female donors are more potent at the immunomodulation effects in vitro than their male counterparts [22, 23].

MSCs from the same donor also show heterogeneity in different tissues. One study conducted by Pieternella S. in 't Anker et al. showed higher expression of $CD34^+$ cells in fetal lung, bone marrow, and spleen than in fetal liver. In fetal lung most of the cells did not express CD45, whereas in fetal spleen, about half of the cells were $CD34^+$ $CD45^-$ [24]. The other paper talked about the differences of MSC proliferatvie rates and differentiation potential between bone marrow MSC and adipose tissue derived MSC [25]. Tissue-dependent differentiation, surface markers and transcriptional and proteomic profiles has been widely studied [26–28].

MSCs display substantial cell-to-cell variation. Bone marrow MSCs isolated from the same donor at a single time point can exhibit different proliferation rate [29]. Even within a single isolation, MSC phenotype are different during culture expansion and downstream use [30].

1.1.3 Mechanism of MSC anti-inflammatory effects

MSC offers a promising therapy treatment for autoimmune diseases, sepsis, and in transplant surgery [31–34]. Various studies have demonstrated immunomodulatory changes after administration of MSCs. However, the underlying mechanism of MSC immunomodulatory capacity are

not fully addressed. This may partially be explained by MSCs heterogeneity from different tissues and different donors [35–38]. MSCs can regulate immune cell proliferation, differentiation and phenotype through cell-to-cell contact and secretion of regulatory molecules, which contain various growth factors and immunomodulatory factors.

MSCs suppress $CD4^+$ and $CD8^+$ T cell proliferation in multiple models. Baboon MSCs showed T cell proliferation suppression and can be applied to graft-versus-host disease (GVHD) model [39]. Human bone-marrow-derived MSCs can inhibit T lymphocytes proliferation in vitro [40]. These suppression effects are mediated by the release of growth factors including transforming growth factor beta (TGF-β) and hepatocyte growth factor (HGF) [41, 42]. On the other hand, MSCs can also regulate T cell differentiation. MSCs induce $CD4^+$ T cell generating regulatory T cell (Tregs), which is essential for immune homeostasis by preventing autoimmunity [43–45]. In addition, upon activation, MSCs also increases the secretion of indoleamine 2,3-dioxygenase (IDO), which can cause tryptophan depletion and lead to allogeneic T cell inhibition [46]. MSCs regulate T cell suppression not only through cell-cell contact, but also through extracellular vesicles (EVs). Studies suggest that EVs can also inhibit T cell proliferation [47], induce T cell apoptosis [48, 49] and promote Treg generation [50].

B cell is effector cell of immune system. MSCs show suppression of B cells proliferation, differentiation and activation [51, 52]. MSCs interact with B cells to induce regulatory B cells (Bregs) and produce Interleukin 10 (IL-10) [53], which suppress inflammation and provide immunological tolerance [54, 55]. In addition, evidence shows a dose-dependent suppression of B cell proliferation and differentiation using EVs [56, 57].

Dendritic cells (DCs) are the most potent antigen-presenting cells (APCs) in the body [58], and are critical in directing the responses of the adaptive immune system [59]. MSCs were shown inhibiting the monocytes/macrophages differentiating into DCs [60–62]. MSCs were also found inhibit $CD34^+$ hematopoietic progenitor cell-derived DCs differentiation and function [63–67]. In addition, MSCs impair DC migration and decrease inflammatory cytokine secretion [68, 69].

Another important immune cell that MSCs can effect is natural killer (NK) cell. Multiple studies suggest MSCs as an inhibitor of NK cells [70–73]. It has been shown that MSCs suppress NK cell proliferation, cytokine production and cytotoxicity [74, 75]. Soluble factors including IDO, Prostaglandin E2 (PGE2), Human leukocyte antigen 5 (HLA-5) and EVs play important roles in NK cells regulation mediated by MSCs [76].

Despite the considerable possibilities of MSCs to treat immune disease, there are still challenges in their clinical use. Various studies showed the therapeutic effects of MSC cannot be obtained after transplanting to patients [77, 78]. These are caused by various mechanisms of action, MSCs heterogeneity, timing, dose and pretreatment of MSC [79–81]. There is a pressing need to develop biomarkers to predict MSC immunosuppression capacity especially in a cell manufacturing setting.

1.2 An overview of metabolomics and MSCs

1.2.1 Metabolomics

Metabolomics is a method to quantitatively analyze metabolites (molecular weight < 1500 Da) in organisms and find relationships between metabolites and physiological and pathological changes [82]. As one type of 'omics' study, metabolomics is considered closely reflecting biological phenotype and environment influence [83, 84]. Nowadays, metabolomics has been widely used to study phenotypes [85], toxicology [86, 87], disease related biomarkers [88] and molecular mechanism [89]. Metabolomics is also used to assess therapy responses. One in vitro study successfully predicted dug-resistant and drug-sensitive groups before treatment with 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea using metabolomics method [90]. This method provide useful information in selecting the drug dose and treatment plan.

Two typical methods in metabolomics are targeted and untargeted metabolomics. Targeted metabolomics focus on identification and quantitation of a pre-defined project-driven set of metabolites in biological samples. To perform targeted analysis, the structure of the target metabolites must be known.

This type of studies often times measures metabolites level in samples and feeds into statistical tools to investigate their contribution to distinguish phenotype of interest from control group [91]. The biggest disadvantage of targeted metabolomics is that it requires the compounds of interest to be known a priori and standard materials should be commercial available. Currently, large number of metabolites cannot be identified using available database, and often times, the purified standards are not available [92].

On the other hand, untargeted studies do not pre-define a set of metabolites, instead, they focus on measuring and comparing all the signals in samples, followed by identification of metabolites using database [93]. This type of study is more interested in identifying unknown metabolites, especially these unknowns have been selected as biomarkers of studies. They usually measure the signal intensities and feed into multivariate modeling to investigate the inherent relations of metabolic profile between different groups. The disadvantage of untargeted metabolomics is that many features are not able to identify.

1.2.2 Metabolomics and biomarker discovery

The term 'biomarker' has been adapted from molecular epidemiology to describe a molecular change in biology [94]. Later, World Health Organization (WHO) defined a biomarker as 'any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease' [95]. Biomarkers are the most objective, quantifiable medical signs laboratory science use to measure reproducibly [96]. In 2002, Henrik Poulsen defined the valid biomarkers with following criteria: identity, linearity, accuracy, selectivity, reproducibility and an appropriate limit of detection [97]. Scientists often times use biomarkers as drug discovery tools and as new targets for therapy, which closely reflect diseases state and could be used for diagnosis as well as therapy monitoring [98].

Metabolomics offers a promising method to characterize diseases and identify biomarkers. As shown in Figure 1.2.2, there is a significant increase of papers using metabolomics to discover

biomarkers since 2010 (data obtained from pubmed). The identification and quantification of key metabolite biomarkers can be used for disease early stage diagnosis, patient stratification and as the treatment response indicators.

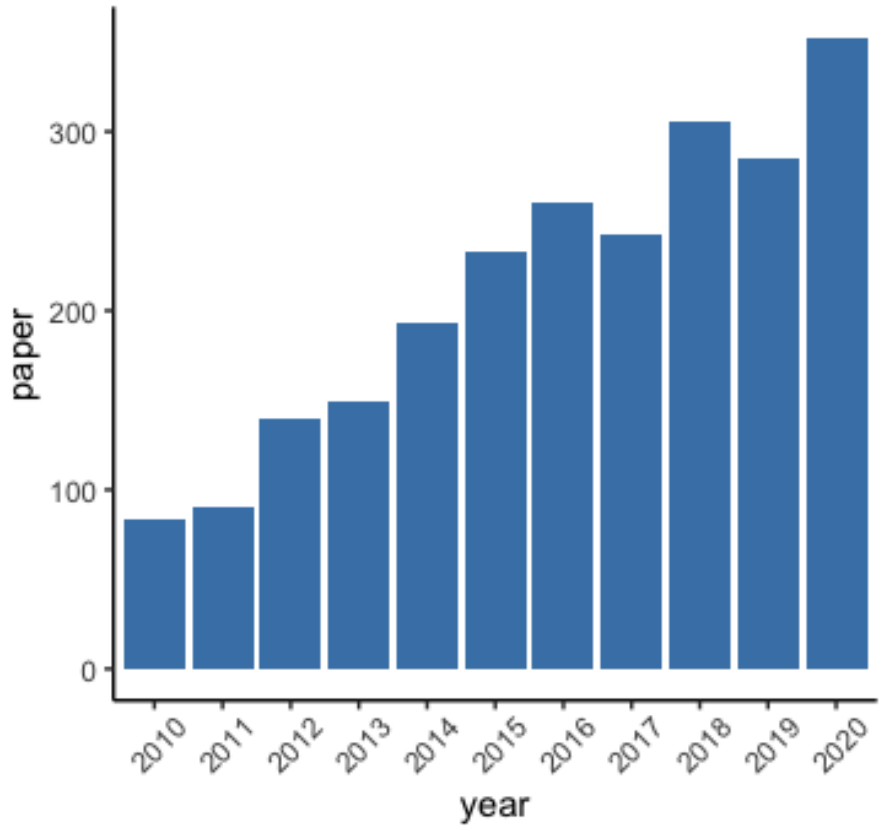


Figure 1.2.1: Number of papers in metabolomics and biomarker discovery (data until 2020)
Data obtained by searching 'metabolomics' and 'biomarker' in pubmed website.

Animal model, human cell cultures, biofluid as well as tissues are common materials used in metabolomic biomarker discovery. One study used serum and tissue samples of colorectal cancer patients to investigate metabolomic profile, and found that ultra-long-chain fatty acids have been significantly reduced in serum samples compared to control healthy group [99]. Another study indicate 11 amino acids were found significantly different between cancer and control groups using serum samples [100]. Urine is another biofluid used in metabolomics biomarker discovery. Cheng

et al. have found that citrate, hippurate, myristate, putrescine and kynurenate are important metabolites in discriminating colorectal cancer from healthy people [101]. Candidate biomarkers such as lysine, glutamine, phenylalanine ect. have been found significant different between pancreatic cancer patients and healthy people [102]. Another study showed a significant decrease of lactate and trimethylamine-N-oxide as well as a significant increase of isoleucine, leucine and creatinine in pancreatic cancer patients (RCC) [103]. Olatomiwa O. Bifarin et al. found that Hippuric acid, lactate, trigonellinamide, and mannitol can be used to predict RCC status [104]. Metabolomics are also used in identifying biomarkers of various types of cancer including breast and ovarian cancers [105–107], urinary cancers [108–111] and lung cancers [112–114].

The disease variability and differences in drug responses in patients indicate the necessity of individualized drug therapy [115]. In this context, metabolomics can fill the gap between phenotype and genotype and raise the possibility of patient stratification and treatment individualization by discovering valuable biomarkers [116].

One study conducted by Hisayuki Erabi and colleagues using metabolomics analysis to find biomarkers that can be used as treatment response for depression. In this study, the Hamilton rating scale for depression (HRSD) was measured and plasma metabolites of major depressive disorder (MDD) patients were collected and found that kynurenic acid and kynurenine were significantly and negatively associated with HRSD reduction after six-week treatment [117]. Evidence shows that metabolomics can also be used to discover biomarkers indicating treatment response.

1.2.3 Metabolomics application in MSC research

As stated before, MSCs show differentiation and anti-inflammatory effects. The characterization of MSCs and understanding the biological mechanisms of function can help translate MSCs from laboratory to cell manufacturing and clinical use. Metabolomics can clarify MSC biological mechanisms of function by monitoring cell activity.

The study conducted by Klontzas ME and colleagues indicate that metabolomics can be used to re-

veal differential sensitivity to osteogenic agents [118]. UCB (umbilical cord blood) MSCs induced with BMP-2 shows different metabolic profile comparing to dexamethasone, and dexamethasone is more efficient in inducing MSC differentiation. Moreover, the MSCs induced by dexamethasone showed similar metabolic physiology compared to primary osteoblasts. Thus, metabolomics is a sensitive tool to monitor and optimize MSC bioprocess. Other studies suggest that hypoxia restricts MSC proliferation to maintain self-renewal capacity [119, 120]. IFN-γ (Interferon gamma) induces MSCs to limit inflammation and promote survival. MSCs primed with IFN-γ shows high concentration in three metabolites including beta-alanine, hypotaurine and tyrosine compared to hypoxia conditioned MSCs, indicating metabolites closely relate to MSC anti-inflammatory effects [121]. The secretome of MSCs have been demonstrated effective on several diseases [122–124]. The investigations of MSC secretome activation mechanisms as well as the mechanisms of secretome component functions are important in applying MSC secretome to clinic. Conditional media and EVs are two main MSC secretome types. Conditional media is the media co-cultured with MSCs, and the EVs are the Extracellular vesicles that secreted from MSCs directly. One study compared metabolic profiles between MSC conditional media and human umbilical cord blood plasma (hUCBP) and found that lipids, β-hydroxybutyrate and inositol shows significant differences [125]. LC-MS based culture media metabolite footprinting was used to characterize MSC osteogenic differentiation by Amal Surrati et al [126]. Culture time-course media analysis reveals that citrate, succinate, glycerol and orthophosphate in MSC media were upregulated in ostrogenic treatment group.

Although multiple studies investigated metabolic profiles in MSC culture under different conditions, limited studies connect metabolomics with MSC immunomodulatory functions. This gap needs to be filled to translate MSCs from laboratory research to clinical use and cell manufacturing setting.

1.3 Nuclear magnetic resonance (NMR) in metabolomics

1.3.1 Advantages of NMR in metabolomics

Multiple sophisticated analytical tools can be used in metabolomics studies, including nuclear magnetic resonance (NMR) [127], gas chromatography–mass spectrometry (GC-MS) [128], and liquid chromatography–mass spectrometry (LC–MS) [129]. Figure 1.3.1 shows the steadily increasing trend of NMR-based metabolomics papers for the past 20 years (data until 2020). This growth of NMR in metabolomics reflects the multiple advantages of NMR platform [130].

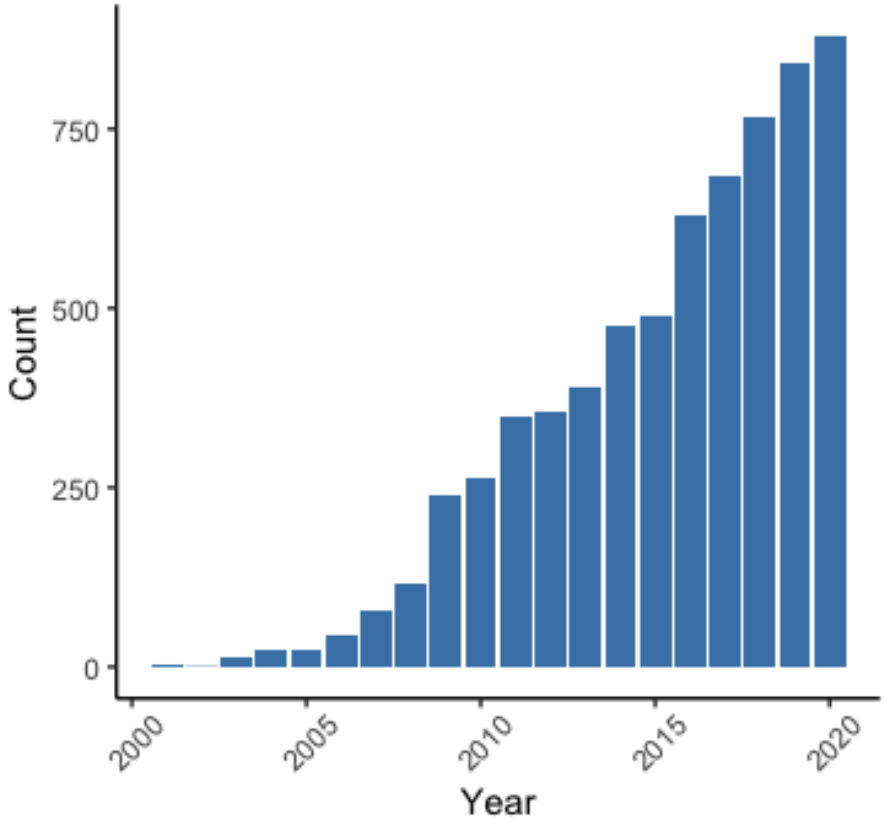


Figure 1.3.1: Number of papers in NMR and metabolomics (data until 2020)

Data obtained by searching 'metabolomics' and 'NMR' in pubmed website.

NMR is highly reproducible in metabolomics studies. One study conducted in UK utilized urine samples from Aito Town, Japan, Chicago, IL and Guangxi, China to investigate the analytical reproducibility of NMR. They found the reproducibility of the NMR screening platform was >98% and most of the classification errors were due to the inhomogeneity of urine samples [131]. In another study, the rat urine sample data were collected at two NMR filed strengths: 500 MHz and 600 MHz, followed by Principal component analysis (PCA), and the results showed near-identical descriptions of the metabolic responses [132]. In addition to reproducibility, NMR is also highly quantitative.

Table 1.2 shows the comparison of common techniques used in metabolomics. The combination of different techniques is often used in metabolomics studies to get a more accurate view of metabolic profile of interest.

Table 1.2: Comparison of NMR and MS in metabolomics

	NMR	MS
Samples	Metabolites with high concentration can be detected	Destructive
	Nondestructive	Need optimization of ionization
	Minimal sample preparation	Tissue samples need extraction
Sensitivity	Relatively low, at μM level	Relatively high, at nM level
Reproducibility	High	Moderate
In vivo study	Yes	No

1.3.2 NMR overview

Nuclear magnetic resonance (NMR) can be used to determine the structure of organic molecules. Nuclei such as ^{1}H, ^{1}3C, 15N, and ^{3}1P possess nuclear spins of 1/2. Each nucleus contains two linearly independent spin states, with m = $\frac{1}{2}$ (spin-up) or m=-$\frac{1}{2}$ (spin-down). In the absence of magnetic filed, these states have the same energy. When an external magnetic field is applied, the different nuclear spin states have different energies, and they are aligned with (low energy state) or

against (high energy state) the magnetic field. When applying an oscillating magnetic field, nuclei generate an electromagnetic signal, and its frequency depends on the magnetic field applied. The frequency can be measured in time domain (Free induction decay, FIDs) and then can be converted into the frequency domain using Fourier transformation (FT) [133]. Figure 1.3.2 is a sample 1D NMR spectrum, with the frequency of resonance as chemical shift in X axis and signal intensities in y axis.

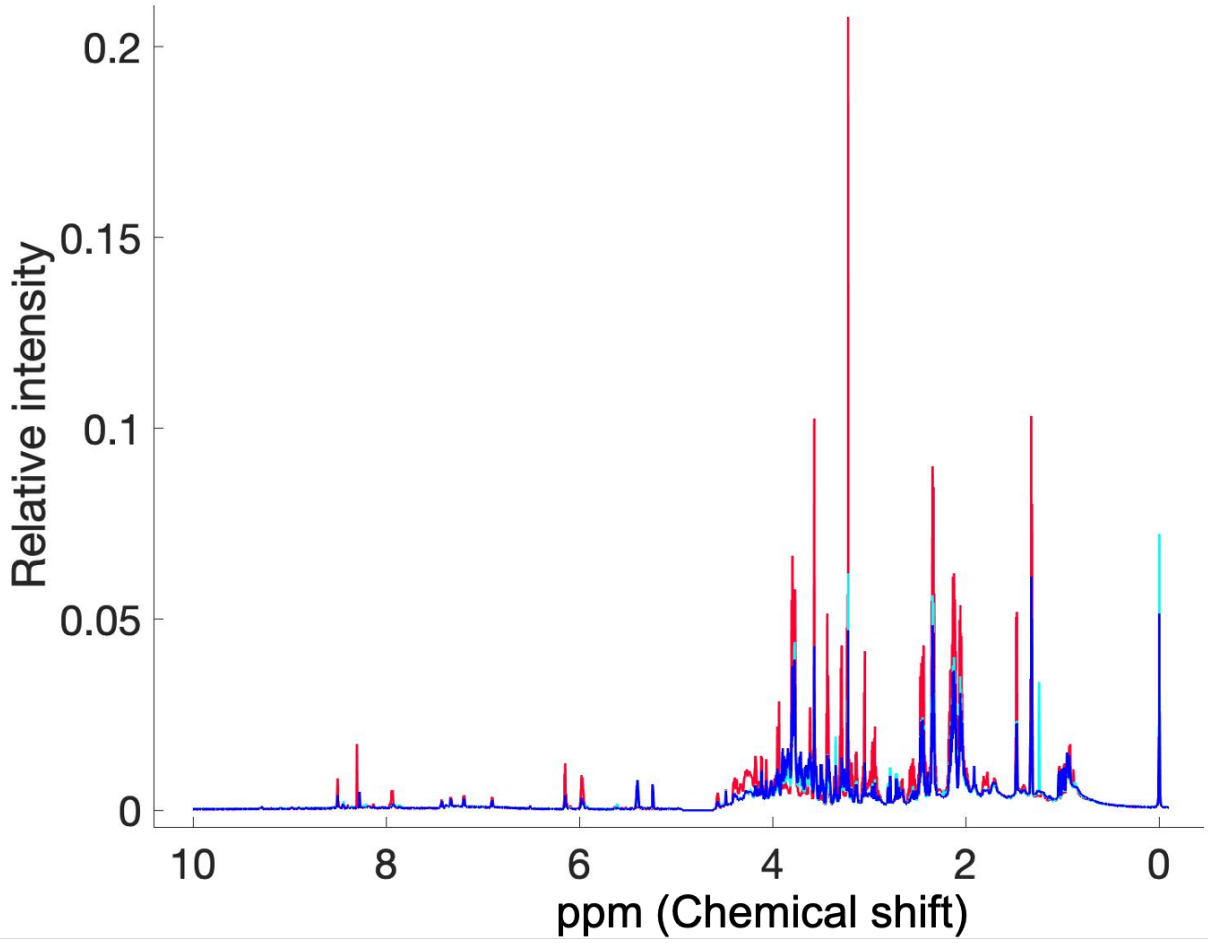


Figure 1.3.2: Sample NMR spectra.

The chemical shift is calculated as following:

$$\delta = \frac{\upsilon_{Sample} - \upsilon_{reference}}{\upsilon_{reference}} \tag{1.1}$$

where the numerator is the difference in Hz between the singal and a reference material, and the denomiator is the operating frequency in MHz.

There are multiple compounds can be used as chemical shift reference materials in ^{1}H-NMR, including Tetramethylsilane (TMS), Sodium-3-Trimethylsiylpropionate-2,2,3,3-d_4 (TMSP-d_4), and Sodium 2,2-Dimethyl-2-silapentane-5-sulfonate (DSS). The reference materials can not only be used to calibrate the chemical shift values, but also used in quantitative NMR to calculate molecules concentrations in samples [134].

In NMR samples, there is more solvent signal (higher concentration) than the analyte. An ordinary proton-contained NMR sample will generate a spectrum with huge solvent absorption, which will distort and dominate the NMR spectrum. Thus, deuterated solvent, like deuterium oxide (D_2O), is often used in NMR samples to keep the magnetic field stable and get high quality data [135]. Moreover, this solvent can be mixed in buffers to control samples PH in order to get consistency chemical shifts between different samples [136].

1.3.3 Sample preparation and experimental design in metabolomics

In metabolomics studies, samples usually need to be frozen in -80° for later analysis in order to prevent rapid degradation of metabolites [137]. The storage of samples needs to be done as soon as possible after sample collection. After sample collection, the next step is sample preparation.

In this dissertation, two types of samples were collected: cell pellets and cell culture media. The cell pellets were thawed on ice and vortexed for one minutes three times. samples were extracted using MeOH:H_2O (80:20) solvent and evaporate using speedvec. Each sample was reconstituted in 80 μl sodium phosphate buffer and vortexed thoroughly and 60 μl was transferred to 1.7-mm

NMR tube for data acquisition. The culture media samples were thawed on ice and centrifuged at 14,000 x g at 4°C for 25 min. Sixty μl of each supernatant was transferred to 1.7-mm NMR tube for data acquisition. After sample preparation, the experiment need to be carefully designed to get reliable and reproducible data.

Quality assurance measurement must be taken in order to ensure the features of interest are not contaminated during sample preparation or data collection. The samples are all randomized to correct any differences caused by sample preparation and instrument alterations through NMR data collection. Buffer blanks are also introduced into studies. The buffer blanks contains high performance liquid chromatography (HPLC) grade water, deuterated solvent and any other buffer added into samples. They are deployed into studies at the beginning and end and one in the middle of each rack. The two extraction blanks are also added after the first buffer blank and before the last buffer blank when samples need extraction procedures. This is to ensure no additional signals introduced to samples during extraction process. In addition, internal pooled controls have also been added and randomized into data collection. They are made by pooling equal small amount of materials from each sample and mixing together. If extraction has been applied, the pooled samples should be generated before sample extraction. These pooled controls are important in determining variations caused by sample preparation and data collection. In addition, they will be used to collect 2D NMR data for metabolite identification (See Section 1.3.5). After 1D data acquisition, the pooled controls should cluster together using principal component analysis (PCA) as shown in figure 1.3.3.

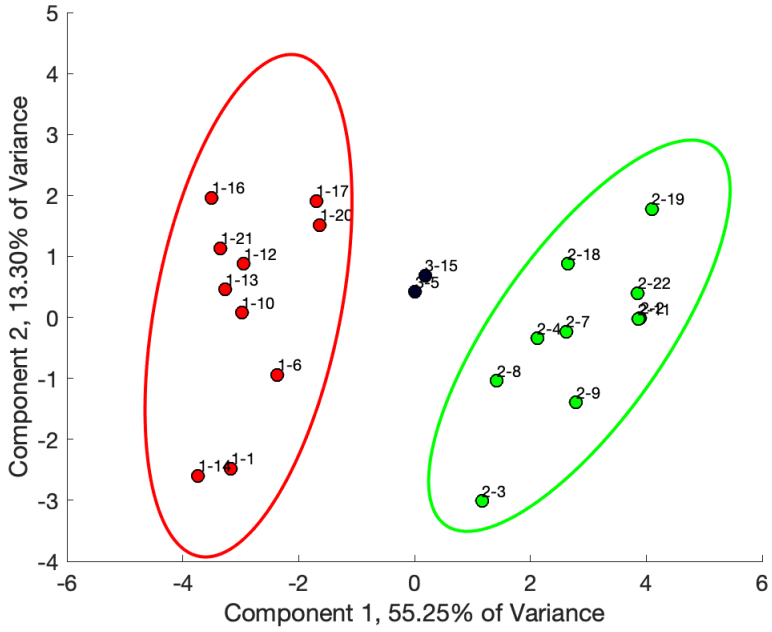


Figure 1.3.3: A sample PCA plot shows the clustering of pooled controls, red and green represents MSCs cultured in complete media and serum-free media respectively. Black points are the pooled controls

1.3.4 NMR data collection and processing

Proton NMR (^{1}H-NMR) is the most common 1D (1-dimensional) NMR experiment in metabolomics because of the high natural abundance (99.98%) of protons as well as high sensitivity. The spectrum is easy and fast to collect, therefore widely used in quantification in metabolomics. In this dissertation, the one dimensional nuclear Overhauser enhancement spectroscopy with water suppression (noesypr1d) were collected on each sample for quantification and downstream analysis. However, this type of spectrum often shows peak overlap, making the identification and quantification of peaks more difficult.

In addition, two common 2D NMR were also collected to further annotate metabolites in samples, they are ^{1}H-^{1}3C heteronuclear single quantum correlation (HSQC) spectroscopy and ^{1}H-^{1}3C heteronuclear single quantum correlation-total correlation spectroscopy (HSQC-TOCSY). ^{1}H-^{1}3C

HSQC is a technique to transfer magnetization on the proton to the second nucleus, in this case, carbon. The result spectrum is a 2-dimensional spectrum with one axis for proton and the other for carbon, which containing peak of each proton and its bonded carbon. It can significantly reduce the complexity of annotation compared to 1D spectrum, thus has been widely used in metabolite annotation. ^{1}H-^{1}3C HSQC-TOCSY combines 1H-13C HSQC with a 1H-TOCSY experiment. A spin system is any group of nuclear spins that interact with one another in a magnetic field, and TOCSY provides proton correlations in a given spin system. TOCSY can show the crosspeaks of protons that belong to the same molecule, thus is very useful in annotation to further reduce the overlap issues. Although 2D NMR can be used to deal with overlap in 1D NMR spectra, they are time-consuming with hours of experimental run time. Thus, it is suggested to collect 2D NMR data on internal pooled samples to get all information from each sample at the same time. The details of spectra metabolites annotation will be discussed in later section.

Before metabolite annotation, spectral processing is needed to get clean and high quality spectra. Several data cleaning steps need to be conducted. In practical, a phase correction on zero-order and first-order is required for Fourier transform in order to obtain the desired appearance of the real part of the spectra [138]. Baseline in 1D NMR is often distorted by the corruption of the first few data points in FID [139]. Correction of baseline is needed because the distortion of baseline can offset signal intensity values, which will cause the peak quantification inaccurate. The referencing is to adjust the chemical shift of a reference compound (i.e DSS) to 0 ppm.

1.3.5 NMR metabolite annotation

As indicated above, although 1D NMR is widely used in quantification, it often has highly overlapping peaks in biological samples, which can cause the identification challenging. 2D NMR, however, contains high resolution information thus can separate most of the overlapping peaks in 1D NMR spectra. This section will briefly discuss techniques for metabolite annotation.

Firstly, peaks from the same metabolite in 1D ^{1}H NMR spectra can be grouped using statistical

correlation spectroscopy (STOCSY). It generates a pseudo-two-dimensional NMR spectrum with the correlation of various peaks intensities across the whole sample based on the multicollinearity of variables in spectra [140]. Thus, peaks with high correlation coefficients are likely belong to the same molecule, which makes 1D NMR annotation easier. Open access databases are also available for 1D NMR peak assignment including Human Metabolome Database (HMDB) [141] and BiomagResBank (BMRB) [142].

2D NMR contains two-dimensional chemical information, thus can reduce the overlapping issue in 1D NMR, and improve the compound identification [143]. HSCQ provides the information of both carbon and proton, and gives their coupled atoms correlation in samples. It shows each proton chemical shift and its bonded carbon chemical shift. TOCSY, on the other hand, provides correlations between all protons in a given spin system [144]. Thus, HSQC-TOCSY combines HSQC and TOCSY together to provide carbon-proton correlation and proton-proton correlation information [145].

Multiple software and website servers are designed for NMR metabolite annotation. One commercial available software is AssureNMR from Bruker company. It matches HSQC and TOCSY spectra to database to identify metabolites in the samples. The list of identified metabolites are ranked from high confidence to low confidence. Another commonly used web server to identify NMR metabolites is Complex Mixture Analysis by NMR (COLMAR). It has HSQC, HSQC-TOCSY and multiple 2D NMR scpectra (COLMARm) queries that matches 2D experiments to database [146–148].

The identification of metabolites in NMR is complicated and challenging, with the risks of false identification because of mismatching to database and missing peaks in our data. Therefore, we designed a confidence score system for metabolites identification in NMR. This system is described as Table 1.3, with the highest confidence score 5 to the lowest confidence score 1:

Table 1.3: The confidence score of NMR metabolite annotation

Confidence score	Criteria
5	Validated by spiking the standard material to the sample
4	Matched to HSQC and validated by HSQC-TOCSY in COLMARm
3	Matched to HSQC
2	Matched to literature and/or BMRB and HMDB database
1	Putatively guessing

1.4 Machine learning and its application in metabolomics

1.4.1 Machine learning overview

With the increase of high throughput data, machine learning is important in metabolomics. A machine learning algorithm is a computational process that uses input data to achieve a desired task [149]. The challenges of metabolomics data, including large feature numbers with small sample size (p»n problem), high noise and batch effects make it important to adopt machine learning algorithms into NMR metabolomics field. Supervised machine learning is one type of machine learning that requires human supervision by either pre-labeling the samples into different groups, or assigning different values to each sample. It involves classification (classify samples into different groups) algorithms and regression (predict continuous outcome values) algorithms.

The following paragraph is from the paper I published with my co-workers in Analytical Chemistry journal entitled 'NMR: Unique Strengths That Enhance Modern Metabolomics Research' [150]. I reprinted here with permission from publisher. Figure 1.4.1 lists popular classification machine learning algorithms in the NMR metabolomics field [151]. Partial least squares (PLS) is a regression method that projects features into linear structures to maximize the explained variance of datasets. PLS-DA is an extension of the PLS algorithm to classify binary classes. RF is a decision tree-based ensemble learning method. By constructing multiple decision trees and combining trees using a

majority voting rank, it can be used for classification. SVM outputs a map of sorted data with the clearest margins to separate the two groups. Sometimes, a kernel method can be used in SVM to transform the margin from linear to non-linear. Artificial neural networks (ANNs) collect the connected units, which allow signal travel between each layer and to model a biological brain [152]. Genetic algorithms are stochastic methods for function optimization based on biological evolution [153].

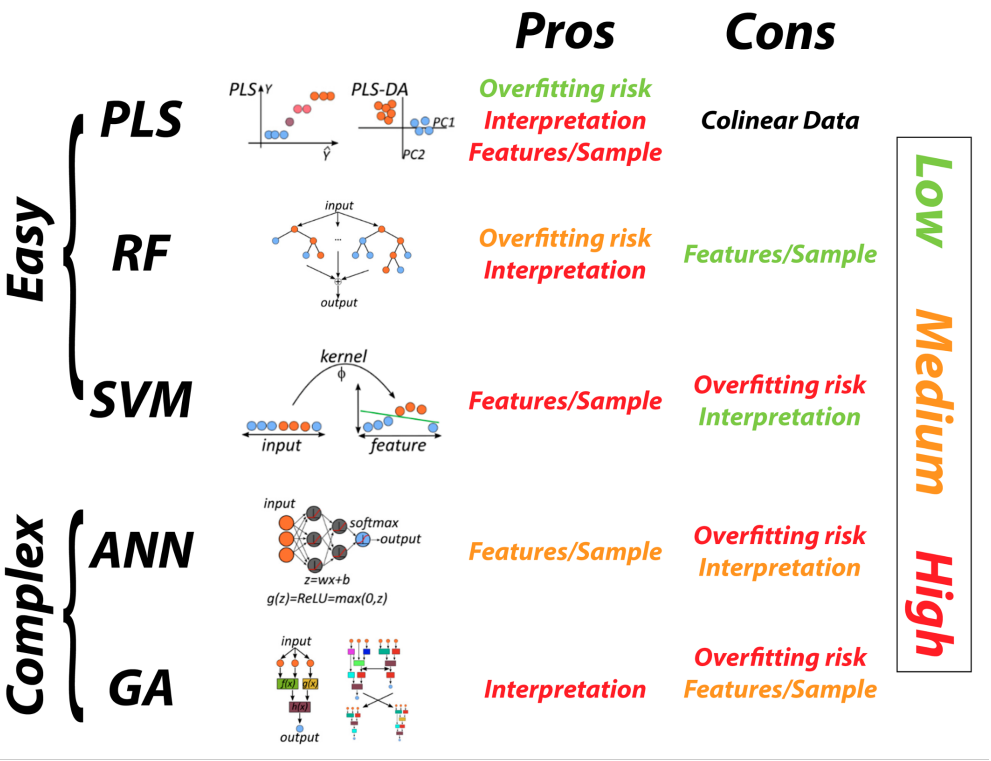


Figure 1.4.1: Advantages and disadvantages of popular ML algorithms. The terms on the left ("Easy" and "Complex") refer to implementation. For more detailed information, please see the reference by Liebal and coworkers. Figure reprinted (adapted) with permission from Edison, Arthur S., et al. "NMR: unique strengths that enhance modern metabolomics research." Analytical Chemistry 93.1 (2020): 478-499. Copyright {2021} American Chemical Society. [150]

Regression is another type of surprised learning. Regression is using features to predict final outcomes. Linear regression is one type of regression methods, and based on this, multiple algorithms

are designed for regressions in metabolomics. Principal components regression (PCR) is a combination of principal component analysis (PCA) and multiple linear regression [154]. SVM (support vector machine) and RF (random forest) are designed for classification but can also applied to regression problems. Table 1.4 shows the advantages and disadvantages of commonly used regression methods in metabolomics.

Table 1.4: Comparison of regression methods in metabolomics

	Pros	Cons
PCR	Solve collinearity	Information loss Data standardization is required
PLSR	Need fewer components for prediction	Overfitting
SVR	Doesn't require grouping or variable selection	Not for large dataset Bad in dealing with noise when p>n, it will underperforme
RFR	Reduce overfitting Automates missing values Data normalization is not required	High computation cost Low interpretability

Unsupervised learning, including clustering, on the other hand, doesn't require human supervision. Unlike supervised learning, unsupervised learning has no label assigned to any pattern. It is used to group data based on similarity among them. Rokach defined clustering as dividing data patterns into subsets in such a way that similar patterns are clustered together [155]. Principal component analysis is one of the common clustering methods used in metabolomics. It is a technique for reducing the dimensionality, by creating uncorrelated variables that maximize the variance [156]. Hierarchical clustering is another clustering algorithm utilizing varied metric and linkage criteria to cluster instances based on data similarities [157]. K-means clustering is a centroid based clustering algorithm. When clusters are specified to K, the algorithm aims to assign each data into the nearest cluster based on distance. Different algorithms can apply to different problems for metabolomics

analysis. However, Metabolomics data often have p»n problem, which makes the feature selection necessary before applying machine learning algorithms.

1.4.2 Feature selection

Though modeling often uses many features, only a few may be related to outcomes in NMR metabolomics. To speed up the learning process as well as increase model performance, feature selection is often performed before the model training process. Feature selection uses specific strategies to select features that contribute most highly to the prediction variable. Many feature selection methods have been developed—from traditional univariate selection methods to modern machine learning algorithms [158].

Although univariate analysis has been rarely solely used to analyze metabolomics data, they are widely used to select features. T-test and ANOVA based feature selection method are called filters. Basically, they set certain thresholds to filter the features. The Weiss group used ANOVA as feature selection algorithm followed by partial least squares (PLS) approach to detect renal cell carcinoma (RCC) in patient's urine [159]. Olatomiwa O. Bifarin et al. used T-test to select features for downstream machine learning analysis in order to distinguish RCC patients from healthy people [104]. Information gain (IG) is another well-known filter strategy. It is an entropy-based feature evaluation method [160] and has been used as a feature selection strategy before applying classification algorithms to identify nonalcoholic fatty liver disease biomarkers [161]. Other filtering methods include chi-square test [162], Fisher score [163], correlation coefficient [164], variance threshold [165] and so on. In this dissertation, variance-based feature selection is used. Variance-based feature selection is one method to select the features based on the overall variances of each feature from all samples. It has the assumption that features with higher variance across different samples should contain more useful information, therefore, need to be kept for downstream analysis. It usually requires to set a threshold to filter all features. The challenge of this method is the choice of threshold, it requires human input and sometimes introduces bias to

dataset. This method is very useful when the data contains a lot of noise [166].

Wrapper is another feature selection strategy. It relies on a specific machine learning algorithm to evaluate and find optimal features. Genetic algorithm (GA) is one of the most popular methods for wrapper feature selection. It selects the features based on a variety of random and local search methods [167]. GA (genetic algorithm) is suitable for when the domain knowledge is hard to provide, thus has been widely used in metabolomics study. A study used GA to select features from 574 NMR metabolites, later, they reduced the feature number to 52. By using these 52 features, the data was then validated using K nearest neighbor. By using GA feature selection, the misclassification rate decreased to zero, indicating this method is efficient. [168]. Other wrapper methods include forward selection, backward elimination and recursive feature elimination. Often times, scientists use hybrid method to combine filter and wrapper, such as using information gain (IG) and GA as a two-stage feature selection method. Comparing to one selection strategy, two-stage selection yields higher accuracy [169, 170]. The third method of feature selection integrates the feature selection as one part of the learning algorithm. The commonly used embedded methods are decision tree and least absolute shrinkage and selection operator (LASSO). Random forest (RF) is the decision tree-based method and often yields low computer cost as well as high accuracy. When training the trees, the machine will compute how much each feature decreased the impurity. The features are more important when they decrease more impurity [171]. Chai group used RF to identify serum biomarkers that can differentiate hepatocellular carcinoma from liver cirrhosis, and successfully selected 7 potential biomarkers metabolites [172].

The following section is from the paper I published with my co-workers in Analytical Chemistry journal entitled 'NMR: Unique Strengths That Enhance Modern Metabolomics Research' [150]. The paper is reprinted here with permission from publisher.

1.4.3 ML algorithms comparison

With the development of high throughput NMR and MS data collection in metabolomics, machine learning algorithms have been widely used in this field for multivariate metabolite analyses in order to diagnose disease [173], predict risk [174], and reveal underlying biological mechanisms between human health and disease [175]. Each method has pros and cons, which makes choosing proper ML algorithms a challenge.

One study compared eight regression methods including ridge regression (RR), LASSO, elastic net (EN), PCR, PLS, sparse PLS (SPLS), SVR, and RFR. The mean square error of prediction and goodness of fit were used to evaluate regression models. In both real metabolomics dataset and simulated dataset, EN performs better than other methods. However, SPLS performs better with respect to the selection of correlated variables [176]. SVR, on the other hand, has been found efficient in multiple studies to reduce batch effects [177, 178]. In metabolomics, another widely used regression model is partial least squares regression (PLSR). It is constructed to maximizing correlation of explanatory variables with response variables. Multiple studies show the advantages of PLSR-based regression methods including: non-linear iterative partial least squares (NIPALS), sparse PLSR, orthogonal PLSR etc. [179–181]

The Broadhurst group has tested eight different linear and nonlinear ML approaches for their performance in binary classification on 10 clinical datasets from metabolomics studies [182–192]. These 10 datasets were acquired either from NMR or MS, dataset size varied from 59 to 968, and the number of metabolite variables of each dataset varied from 29 to 689. The eight ML approaches are partial least squares regression (PLSR), principal component regression (PCR), principal component logistic regression (PCLR), linear kernel support vector machines (SVM-Lin), radial basis function kernel support vector machines (SVM-RBF), RF, linear ANNs, and non-linear ANNs. The results showed that overall the linear ML algorithms and non-linear algorithms achieved similar prediction performance in the binary classification metabolomics scenario. RF performed poorly with a small number of samples when the number of variables was large [193]. It also emphasized

that if applying Occam's razor principle, the PLS-DA remains the first choice in binary classification problems [194].

Similarly, Power's group also evaluated five different ML algorithms (Orthogonal PLS-DA, PLS, SVM, RF, and principal components – linear discriminant analysis (PC-LDA)) in binary classification NMR metabolomics studies [195]. They simulated a 50-metabolite-NMR data set to mimic human urine sample data with known within-group variances, between-group variances, and precisely defined group separation. Thirty-three out of these 50 metabolites were commercially available and used to collect experimental NMR spectra. Overall, equivalent performance was achieved from the five ML algorithms when analyzing high-quality datasets with low noise, small within-group variance, and large between-group variance. When the group separation contributor was limited to one single variable, OPLS-DA and PC-LDA outperformed other models.

From these papers, a simple conclusion can be drawn: PLS-DA and OPLS-DA are currently the gold standard for binary classification. However, things are different in multi-class classification cases. Multi-class classification trains a system to discriminate different classes for various unknown objects. With the increase of class numbers, the complexities of the model also increase, making it more complicated than in a binary classification [196]. Therefore, the strategy of multi-class classification analysis should also be thoroughly addressed. The common strategy to handle multi-class classification is to transform the problem into a binary case, which is called binary decomposition [197]. One-against-all (OAA) and one-against-one (OAO) are two common binary decomposition strategies. OAA divides each class and all other classes into two groups, transforming the K-class classification problem into K parallel binary classifications [198], while OAO generates binary classification between each pair of classes and total K(K-1)/2 parallel binary classifications are conducted [199]. A study of multiclass discrimination in untargeted metabolomics has been conducted by Trainor et al. to evaluate the performance of six different classifiers [200]. In this case, the OAO approach was used in multi-class classification. The simulated data were incorporated with realistic block-wise correlation and partial correlation of structures to mimic the

correlations and metabolite clustering in untargeted metabolomics studies, and then tested with six ML approaches including PLS-DA, Sparse PLS-DA, RF, SVM, ANN, KNN, and naïve Bayes. The results showed that SVM and RF outperformed other models when the studies incorporated non-normal error distributions, unbalanced phenotype allocation, outliers, missing values, and dimension reduction. When training with the three class datasets, SVM and RF also perform better than other algorithms. Therefore, knowing the intricate pattern of the dataset, considering the statistical power of analysis, and choosing an algorithm accordingly are important and essential when applying ML algorithms in metabolomics studies.

1.5 Key challenges and major research questions addressed

Analysis of MSCs is often unpredictable outcomes in clinical trials because of heterogeneity and lacking of predictive biomarkers. Finding predictive biomarkers to indicate MSC anti-inflammatory effects has been pursued by researchers across different science fields. Metabolomics is one type of omics studies that can most closely reflect the phenotype and environmental influences. Thus, in this dissertation, I hypothesized that the intracellular and extracellular metabolites can be used as surrogates to predict MSC anti-inflammatory outcomes.

To test this hypothesis, we performed cell culture on three differnt MSC cell lines, including iMSC (iPS-derived MSC, Cellular Dynamics International, Madison WI, Lot 0003), BM71 (Bone marrow-derived MSC, F, 18-30, RoosterBio, Frederick MD) and BM182 (Bone marrow-derived MSC, F, 26, RoosterBio, Frederick MD). These cells were cultured for three passages using continuous 25, 21, 23 days respectively.

In chapter 2, I will show that intracellular metabolic profile and its relationship with MSC T cell suppression and IDO assays results. Intracellular metaboloites detecting using NMR and MS, along with cytokines profile, were used to predict MSC T cell suppression and IDO assays results. The predictive biomarkers were identified using PLSR model and variable importance in projection (VIP) scores. In addition to regression method, classification methods including KNN, RF, logistic

regression and SVM have also been used to predict outcome classes generated by K-means clustering. The important features of the best model (RF) showed highly overlap with PLSR model. This chapter is a research paper accepted for publication in Cytotherapy titled 'Multi-omics characterization of mesenchymal stem/stromal cells for the identification of putative critical quality attributes'.
In chapter 3, the same study set was used for analysis. Instead of intracellular metabolomics, this chapter utilized culture media to investigate extracellular metabolic profile. A PLSR model was used to regress metabolite change across time against composite functional score generated from T cell suppression and IDO assays. The pathway analysis was conducted to explain how media metabolites can influence cell metabolites change and explain the underlying mechanisms. The GISSMO (guided ideographic spin system model optimization) simulation is used to estimate spectra generated in benchtop NMR and test the feasiability of applying this method to cell manufacturing. This chapter is a manuscript written for submission to the journal Cytotherapy and titled 'Non-destructive characterization of Mesenchymal stem cells'.
In chapter 4, I will show that the composite functional score can be predicted using new cell line first 3 day culture media metabolomics profile. I aim to validate and apply our methods on cell manufacturing. The first generation and second generation of bioreactors were used for MSC culture and culture media was collected and analyzed using NMR. The relations between media metabolite and MSC proliferation was investigated using multivariate statistical analysis. In addition, a large cohort of different type of MSCs were cultured in two types of media and the culture media metabolites were analyzed using NMR to predict the composite functional score we will generate and later validate the biomarkers we identified in previous chapters.
In chapter 5, I concluded the dissertation by summarizing my contributions to this field and offer some possible paths to build on my findings, and ultimately, apply to clinical use and cell manufacturing.

BIBLIOGRAPHY

[1] Catherine M Kolf, Elizabeth Cho, and Rocky S Tuan. "Mesenchymal stromal cells: biology of adult mesenchymal stem cells: regulation of niche, self-renewal and differentiation". In: *Arthritis research & therapy* 9.1 (2007), pp. 1–10.

[2] Li-Tzu Wang et al. "Human mesenchymal stem cells (MSCs) for treatment towards immune-and inflammation-mediated diseases: review of current clinical trials". In: *Journal of biomedical science* 23.1 (2016), pp. 1–13.

[3] MLBK Dominici et al. "Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement". In: *Cytotherapy* 8.4 (2006), pp. 315–317.

[4] S Schachtele, C Clouser, and J Aho. "Markers and Methods to Verify Mesenchymal Stem Cell Identity". In: *Potency, and Quality.[(accessed on 29 January 2020)]* ().

[5] Ching-Shwun Lin et al. "Is CD34 truly a negative marker for mesenchymal stromal cells?" In: *Cytotherapy* 14.10 (2012), pp. 1159–1163.

[6] Ludovic Zimmerlin et al. "Human adipose stromal vascular cell delivery in a fibrin spray". In: *Cytotherapy* 15.1 (2013), pp. 102–108.

[7] S Kaiser et al. "BM cells giving rise to MSC in culture have a heterogeneous CD34 and CD45 phenotype". In: *Cytotherapy* 9.5 (2007), pp. 439–450.

[8] Paul J Simmons and Beverly Torok-Storb. "CD34 expression by stromal precursors in normal human adult bone marrow". In: (1991).

[9] PJ Simmons and B Torok-Storb. "CD34 expression by stromal precursors in normal human adult bone". In: (2011).

[10] Nadia Quirici et al. "Anti-L-NGFR and-CD34 monoclonal antibodies identify multipotent mesenchymal stem cells in human adipose tissue". In: *Stem cells and development* 19.6 (2010), pp. 915–925.

[11] Kotaro Yoshimura et al. "Characterization of freshly isolated and cultured cells derived from the fatty and fluid portions of liposuction aspirates". In: *Journal of cellular physiology* 208.1 (2006), pp. 64–76.

[12] G Pachón-Peña et al. "Stromal stem cells from adipose tissue and bone marrow of age-matched female donors display distinct immunophenotypic profiles". In: *Journal of cellular physiology* 226.3 (2011), pp. 843–851.

[13] CM McLeod and RL Mauck. "On the origin and impact of mesenchymal stem cell heterogeneity: new insights and emerging tools for single cell analysis". In: *European cells & materials* 34 (2017), p. 217.

[14] Donald G Phinney. "Functional heterogeneity of mesenchymal stem cells: implications for cell therapy". In: *Journal of cellular biochemistry* 113.9 (2012), pp. 2806–2812.

[15] Alison Wilson, Andrew Webster, and Paul Genever. "Nomenclature and heterogeneity: consequences for the use of mesenchymal stem cells in regenerative medicine". In: *Regenerative medicine* 14.6 (2019), pp. 595–611.

[16] Susan W Volk, Yanjian Wang, and Kurt D Hankenson. "Effects of donor characteristics and ex vivo expansion on canine mesenchymal stem cell properties: implications for MSC-based therapies". In: *Cell transplantation* 21.10 (2012), pp. 2189–2200.

[17] Georg Siegel et al. "Phenotype, donor age and gender affect function of human bone marrow-derived mesenchymal stromal cells". In: *BMC medicine* 11.1 (2013), pp. 1–20.

[18] Joice Fülber et al. "Comparative study of equine mesenchymal stem cells from healthy and injured synovial tissues: an in vitro assessment". In: *Stem cell research & therapy* 7.1 (2016), pp. 1–13.

[19] Mahmood S Choudhery et al. "Donor age negatively impacts adipose tissue-derived mesenchymal stem cell expansion and differentiation". In: *Journal of translational medicine* 12.1 (2014), pp. 1–14.

[20] Sergey V Tokalov et al. "Age-related changes in the frequency of mesenchymal stem cells in the bone marrow of rats". In: *Stem cells and development* 16.3 (2007), pp. 439–446.

[21] J Justesen et al. "Maintenance of osteoblastic and adipocytic differentiation potential with age and osteoporosis in human marrow stromal cell cultures." In: *Calcified tissue international* 71.1 (2002).

[22] Flyn Mckinnirey et al. "Immune modulation via adipose derived Mesenchymal Stem cells is driven by donor sex in vitro". In: *Scientific Reports* 11.1 (2021), pp. 1–14.

[23] Flyn Mckinnirey et al. "Mesenchymal Stem Cell Mediated Immune Modulation Is Donor Sex Specific". In: (2021).

[24] WA Noort et al. "Mesenchymal stem cells in human second-trimester bone marrow, liver, lung, and spleen exhibit a similar immunophenotype but a heterogeneous multilineage differentiation potential". In: *haematologica* 88.8 (2003), pp. 845–852.

[25] June Seok Heo et al. “Comparison of molecular profiles of human mesenchymal stem cells derived from bone marrow, umbilical cord blood, placenta and adipose tissue”. In: *International journal of molecular medicine* 37.1 (2016), pp. 115–125.

[26] Susanne Kern et al. “Comparative analysis of mesenchymal stem cells from bone marrow, umbilical cord blood, or adipose tissue”. In: *Stem cells* 24.5 (2006), pp. 1294–1301.

[27] Philipp Mattar and Karen Bieback. “Comparing the immunomodulatory properties of bone marrow, adipose tissue, and birth-associated tissue mesenchymal stromal cells”. In: *Frontiers in immunology* 6 (2015), p. 560.

[28] Marius Strioga et al. “Same or not the same? Comparison of adipose tissue-derived versus bone marrow-derived mesenchymal stem and stromal cells”. In: *Stem cells and development* 21.14 (2012), pp. 2724–2752.

[29] Donald G Phinney et al. “Donor variation in the growth properties and osteogenic potential of human marrow stromal cells”. In: *Journal of cellular biochemistry* 75.3 (1999), pp. 424–436.

[30] CM McLeod and RL Mauck. “On the origin and impact of mesenchymal stem cell heterogeneity: new insights and emerging tools for single cell analysis”. In: *European cells & materials* 34 (2017), p. 217.

[31] Katarina Le Blanc et al. “Mesenchymal stem cells for treatment of steroid-resistant, severe, acute graft-versus-host disease: a phase II study”. In: *The Lancet* 371.9624 (2008), pp. 1579–1586.

[32] Geoffrey M Forbes et al. “A phase 2 study of allogeneic mesenchymal stromal cells for luminal Crohn’s disease refractory to biologic therapy”. In: *Clinical Gastroenterology and Hepatology* 12.1 (2014), pp. 64–71.

[33] Manuel A González et al. “Adipose-derived mesenchymal stem cells alleviate experimental colitis by inhibiting inflammatory and autoimmune responses”. In: *Gastroenterology* 136.3 (2009), pp. 978–989.

[34] Marlies EJ Reinders et al. “Autologous bone marrow-derived mesenchymal stromal cells for the treatment of allograft rejection after renal transplantation: Results of a phase I study”. In: *Stem cells translational medicine* 2.2 (2013), pp. 107–111.

[35] Andreas Robert Rudolf Weiss and Marc Hendrik Dahlke. “Immunomodulation by mesenchymal stem cells (MSCs): mechanisms of action of living, apoptotic, and dead MSCs”. In: *Frontiers in immunology* 10 (2019), p. 1191.

[36] Yongkang Wu et al. “Adipose tissue-derived mesenchymal stem cells have a heterogenic cytokine secretion profile”. In: *Stem cells international* 2017 (2017).

[37] Yinan Deng et al. “Umbilical cord-derived mesenchymal stem cells instruct monocytes towards an IL10-producing phenotype by secreting IL6 and HGF”. In: *Scientific reports* 6.1 (2016), pp. 1–9.

[38] Samantha FH De Witte et al. “Aging of bone marrow–and umbilical cord–derived mesenchymal stromal cells during expansion”. In: *Cytotherapy* 19.7 (2017), pp. 798–807.

[39] Amelia Bartholomew et al. “Mesenchymal stem cells suppress lymphocyte proliferation in vitro and prolong skin graft survival in vivo”. In: *Experimental hematology* 30.1 (2002), pp. 42–48.

[40] Sarah Glennie et al. “Bone marrow mesenchymal stem cells induce division arrest anergy of activated T cells”. In: *Blood* 105.7 (2005), pp. 2821–2827.

[41] Massimo Di Nicola et al. “Human bone marrow stromal cells suppress T-lymphocyte proliferation induced by cellular or nonspecific mitogenic stimuli”. In: *Blood, The Journal of the American Society of Hematology* 99.10 (2002), pp. 3838–3843.

[42] Sarah Glennie et al. “Bone marrow mesenchymal stem cells induce division arrest anergy of activated T cells”. In: *Blood* 105.7 (2005), pp. 2821–2827.

[43] Shohei Hori, Takashi Nomura, and Shimon Sakaguchi. “Control of regulatory T cell development by the transcription factor Foxp3”. In: *Science* 299.5609 (2003), pp. 1057–1061.

[44] Alexander Marson et al. “Foxp3 occupancy and regulation of key target genes during T-cell stimulation”. In: *Nature* 445.7130 (2007), pp. 931–935.

[45] Shimon Sakaguchi et al. “Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases.” In: *The Journal of Immunology* 155.3 (1995), pp. 1151–1164.

[46] Wei Ge et al. “Regulatory T-cell generation and kidney allograft tolerance induced by mesenchymal stem cells associated with indoleamine 2, 3-dioxygenase expression”. In: *Transplantation* 90.12 (2010), pp. 1312–1320.

[47] Li Wang et al. “Extracellular vesicles released from human umbilical cord-derived mesenchymal stromal cells prevent life-threatening acute graft-versus-host disease in a mouse model of allogeneic hematopoietic stem cell transplantation”. In: *Stem cells and development* 25.24 (2016), pp. 1874–1883.

[48] Aram Mokarizadeh et al. “Microvesicles derived from mesenchymal stem cells: potent organelles for induction of tolerogenic signaling”. In: *Immunology letters* 147.1-2 (2012), pp. 47–54.

[49] Andrea Del Fattore et al. “Immunoregulatory effects of mesenchymal stem cell-derived extracellular vesicles on T lymphocytes”. In: *Cell transplantation* 24.12 (2015), pp. 2615–2627.

[50] Bin Zhang et al. “Mesenchymal stem cells secrete immunologically active exosomes”. In: *Stem cells and development* 23.11 (2014), pp. 1233–1244.

[51] Anna Corcione et al. “Human mesenchymal stem cells modulate B-cell functions”. In: *Blood* 107.1 (2006), pp. 367–372.

[52] Soraya Tabera et al. “The effect of mesenchymal stem cells on the viability, proliferation and differentiation of B-lymphocytes”. In: *haematologica* 93.9 (2008), pp. 1301–1309.

[53] Min Yang et al. “Cutting edge: novel function of B cell-activating factor in the induction of IL-10–producing regulatory B cells”. In: *The Journal of Immunology* 184.7 (2010), pp. 3321–3325.

[54] Marcella Franquesa et al. “Human adipose tissue-derived mesenchymal stem cells abrogate plasmablast formation and induce regulatory B cells independently of T helper cells”. In: *Stem cells* 33.3 (2015), pp. 880–891.

[55] Elizabeth C Rosser and Claudia Mauri. “Regulatory B cells: origin, phenotype, and function”. In: *Immunity* 42.4 (2015), pp. 607–612.

[56] Mariano Di Trapani et al. “Differential and transferable modulatory effects of mesenchymal stromal cell-derived extracellular vesicles on T, B and NK cell functions”. In: *Scientific Reports* 6.1 (2016), pp. 1–13.

[57] Antonella Conforti et al. “Microvescicles derived from mesenchymal stromal cells are not as effective as their cellular counterpart in the ability to modulate immune responses in vitro”. In: *Stem cells and development* 23.21 (2014), pp. 2591–2599.

[58] Alexander Mildner and Steffen Jung. “Development and function of dendritic cell subsets”. In: *Immunity* 40.5 (2014), pp. 642–656.

[59] Katarina Le Blanc and Dimitrios Mougiakakos. “Multipotent mesenchymal stromal cells and the innate immune system”. In: *Nature Reviews Immunology* 12.5 (2012), pp. 383–396.

[60] Rajesh Ramasamy et al. “Mesenchymal stem cells inhibit dendritic cell differentiation and function by preventing entry into the cell cycle”. In: *Transplantation* 83.1 (2007), pp. 71–76.

[61] Xiao-Xia Jiang et al. “Human mesenchymal stem cells inhibit differentiation and function of monocyte-derived dendritic cells”. In: *Blood* 105.10 (2005), pp. 4120–4126.

[62] Sara M Melief et al. “Multipotent stromal cells induce human regulatory T cells through a novel pathway involving skewing of monocytes toward anti-inflammatory macrophages”. In: *Stem Cells* 31.9 (2013), pp. 1980–1991.

[63] Alma J Nauta et al. “Mesenchymal stem cells inhibit generation and function of both CD34+-derived and monocyte-derived dendritic cells”. In: *The Journal of Immunology* 177.4 (2006), pp. 2080–2087.

[64] Yin-Ping Li et al. “Human mesenchymal stem cells license adult CD34+ hemopoietic progenitor cells to differentiate into regulatory dendritic cells through activation of the Notch pathway”. In: *The Journal of Immunology* 180.3 (2008), pp. 1598–1608.

[65] Hsin-Wei Chen et al. “Mesenchymal stem cells tune the development of monocyte-derived dendritic cells toward a myeloid-derived suppressive phenotype through growth-regulated oncogene chemokines”. In: *The Journal of Immunology* 190.10 (2013), pp. 5065–5077.

[66] FM Abomaray et al. “Human chorionic villous mesenchymal stem cells modify the functions of human dendritic cells, and induce an anti-inflammatory phenotype in CD1+ dendritic cells”. In: *Stem Cell Reviews and Reports* 11.3 (2015), pp. 423–441.

[67] Karen English, Frank P Barry, and Bernard P Mahon. “Murine mesenchymal stem cells suppress dendritic cell migration, maturation and antigen presentation”. In: *Immunology letters* 115.1 (2008), pp. 50–58.

[68] Sabrina Chiesa et al. “Mesenchymal stem cells impair in vivo T-cell priming by dendritic cells”. In: *Proceedings of the National Academy of Sciences* 108.42 (2011), pp. 17384–17389.

[69] Hong Li et al. “Mesenchymal stem cells alter migratory property of T and dendritic cells to delay the development of murine lethal acute graft-versus-host disease”. In: *Stem Cells* 26.10 (2008), pp. 2531–2541.

[70] Sudeepta Aggarwal and Mark F Pittenger. “Human mesenchymal stem cells modulate allogeneic immune cell responses”. In: *Blood* 105.4 (2005), pp. 1815–1822.

[71] Grazia Maria Spaggiari et al. “Mesenchymal stem cell-natural killer cell interactions: evidence that activated NK cells are capable of killing MSCs, whereas MSCs can inhibit IL-2-induced NK-cell proliferation”. In: *Blood* 107.4 (2006), pp. 1484–1490.

[72] Panagiota A Sotiropoulou et al. “Interactions between human mesenchymal stem cells and natural killer cells”. In: *Stem cells* 24.1 (2006), pp. 74–85.

[73] Grazia Maria Spaggiari et al. “Mesenchymal stem cells inhibit natural killer–cell proliferation, cytotoxicity, and cytokine production: role of indoleamine 2, 3-dioxygenase and prostaglandin E2”. In: *Blood, The Journal of the American Society of Hematology* 111.3 (2008), pp. 1327–1333.

[74] Mengmeng Qu et al. “Bone marrow-derived mesenchymal stem cells suppress NK cell recruitment and activation in PolyI: C-induced liver injury”. In: *Biochemical and biophysical research communications* 466.2 (2015), pp. 173–179.

[75] Clive M Michelo et al. “Added effects of dexamethasone and mesenchymal stem cells on early Natural Killer cell activation”. In: *Transplant immunology* 37 (2016), pp. 1–9.

[76] Zohair Selmani et al. “Human leukocyte antigen-G5 secretion by human mesenchymal stem cells is required to suppress T lymphocyte and natural killer function and to induce CD4+ CD25highFOXP3+ regulatory T cells”. In: *Stem cells* 26.1 (2008), pp. 212–222.

[77] Bassem Yamout et al. “Bone marrow mesenchymal stem cell transplantation in patients with multiple sclerosis: a pilot study”. In: *Journal of neuroimmunology* 227.1-2 (2010), pp. 185–189.

[78] HillardM Lazarus et al. “Ex vivo expansion and subsequent infusion of human bone marrow-derived stromal progenitor cells (mesenchymal progenitor cells): implications for therapeutic use.” In: *Bone marrow transplantation* 16.4 (1995), pp. 557–564.

[79] Maciej Kabat et al. “Trends in mesenchymal stem cell clinical trials 2004-2018: Is efficacy optimal in a narrow dose range?” In: *Stem cells translational medicine* 9.1 (2020), pp. 17–27.

[80] Hiromi Hayashita-Kinoh et al. “Improved transduction of canine X-linked muscular dystrophy with rAAV9-microdystrophin via multipotent MSC pretreatment”. In: *Molecular Therapy-Methods & Clinical Development* 20 (2021), pp. 133–141.

[81] Chantal Lechanteur et al. “MSC Manufacturing for Academic Clinical Trials: From a Clinical-Grade to a Full GMP-Compliant Process”. In: *Cells* 10.6 (2021), p. 1320.

[82] Qiang Yang et al. “Metabolomics biotechnology, applications, and future trends: a systematic review”. In: *RSC Advances* 9.64 (2019), pp. 37245–37257.

[83] Oliver Fiehn. “Metabolomics—the link between genotypes and phenotypes”. In: *Functional genomics* (2002), pp. 155–171.

[84] Jeremy K Nicholson, John C Lindon, and Elaine Holmes. “’Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data”. In: *Xenobiotica* 29.11 (1999), pp. 1181–1189.

[85] Adam Hines et al. “Direct sampling of organisms from the field and knowledge of their phenotype: key recommendations for environmental metabolomics”. In: *Environmental Science & Technology* 41.9 (2007), pp. 3375–3381.

[86] Kei Zaitsu et al. “Application of metabolomics to toxicology of drugs of abuse: A mini review of metabolomics approach to acute and chronic toxicity studies”. In: *Drug metabolism and pharmacokinetics* 31.1 (2016), pp. 21–26.

[87] Jun Zeng et al. “Effect of bisphenol A on rat metabolic profiling studied by using capillary electrophoresis time-of-flight mass spectrometry”. In: *Environmental science & technology* 47.13 (2013), pp. 7457–7465.

[88] Yang Zhou et al. "Discovery and validation of potential urinary biomarkers for bladder cancer diagnosis using a pseudotargeted GC-MS metabolomics method". In: *Oncotarget* 8.13 (2017), p. 20719.

[89] Arun Sreekumar et al. "Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression". In: *Nature* 457.7231 (2009), pp. 910–914.

[90] Wael El-Deredy et al. "Pretreatment prediction of the chemotherapeutic response of human glioma cell cultures using nuclear magnetic resonance spectroscopy and artificial neural networks". In: *Cancer research* 57.19 (1997), pp. 4196–4199.

[91] Aalim M Weljie et al. "Targeted profiling: quantitative analysis of 1H NMR metabolomics data". In: *Analytical chemistry* 78.13 (2006), pp. 4430–4442.

[92] Vladimir Shulaev. "Metabolomics technology and bioinformatics". In: *Briefings in bioinformatics* 7.2 (2006), pp. 128–139.

[93] Kerem Bingol. "Recent advances in targeted and untargeted metabolomics by NMR and MS/NMR methods". In: *High-throughput* 7.2 (2018), p. 9.

[94] E Offord. "Markers of oxidative damage and antioxidant protection: current status and relevance to disease". In: *Free Radic Res* 33 (2000), S5–S19.

[95] World Health Organization et al. *Biomarkers in Risk Assessment: Validity and Vlidation-Environmental Health Criteria 222*. 2001.

[96] Kyle Strimbu and Jorge A Tavel. "What are biomarkers?" In: *Current Opinion in HIV and AIDS* 5.6 (2010), p. 463.

[97] Helen R Griffiths et al. "Biomarkers". In: *Molecular aspects of medicine* 23.1-3 (2002), pp. 101–208.

[98] Kewal K Jain and Kewal K Jain. *The handbook of biomarkers*. Springer, 2010.

[99] Shawn A Ritchie et al. "Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: implications for early screening and detection". In: *BMC medicine* 8.1 (2010), pp. 1–20.

[100] Alexander Benedikt Leichtle et al. "Serum amino acid profiles and their alterations in colorectal cancer". In: *Metabolomics* 8.4 (2012), pp. 643–653.

[101] Yu Cheng et al. "Distinct urinary metabolic profile of human colorectal cancer". In: *Journal of proteome research* 11.2 (2012), pp. 1354–1363.

[102] Shiro Urayama et al. "Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer". In: *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* 24.5 (2010), pp. 613–620.

[103] Dong OuYang et al. "Metabolomic profiling of serum from human pancreatic cancer patients using 1 H NMR spectroscopy and principal component analysis". In: *Applied biochemistry and biotechnology* 165.1 (2011), pp. 148–154.

[104] Olatomiwa O Bifarin et al. "Machine Learning-Enabled Renal Cell Carcinoma Status Prediction Using Multiplatform Urine-Based Metabolomics". In: *Journal of Proteome Research* (2021).

[105] Philippe Bougnoux, Nawale Hajjaji, and Charles Couet. "The lipidome as a composite biomarker of the modifiable part of the risk of breast cancer". In: *Prostaglandins, Leukotrienes and Essential Fatty Acids* 79.3-5 (2008), pp. 93–96.

[106] Kai Ren Ong et al. "Biomarkers of dietary energy restriction in women at increased risk of breast cancer". In: *Cancer Prevention Research* 2.8 (2009), pp. 720–731.

[107] M Luısa Dória et al. "Lipidomic approach to identify patterns in phospholipid profiles and define class differences in mammary epithelial and breast cancer cells". In: *Breast cancer research and treatment* 133.2 (2012), pp. 635–648.

[108] Lin Lin et al. "LC-MS-based serum metabolic profiling for genitourinary cancer classification and cancer type-specific biomarker discovery". In: *Proteomics* 12.14 (2012), pp. 2238–2246.

[109] Zhenzhen Huang et al. "Bladder cancer determination via two urinary metabolites: a biomarker pattern approach". In: *Molecular & Cellular Proteomics* 10.10 (2011).

[110] Sheila Ganti et al. "Urinary acylcarnitines are altered in human kidney cancer". In: *International journal of cancer* 130.12 (2012), pp. 2791–2800.

[111] Kyoungmi Kim et al. "Urine metabolomics analysis for kidney cancer detection and biomarker discovery". In: *Molecular & cellular proteomics* 8.3 (2009), pp. 558–570.

[112] Hao Wu et al. "Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry". In: *Analytica chimica acta* 648.1 (2009), pp. 98–104.

[113] Jian Zhang et al. "Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods". In: *PloS one* 7.1 (2012), e30181.

[114] Yumei Guo et al. "Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using Fourier transform ion cyclotron resonance mass spectrometry". In: *Clinica chimica acta* 414 (2012), pp. 135–141.

[115] Leonor Puchades-Carrasco and Antonio Pineda-Lucena. "Metabolomics in pharmaceutical research and development". In: *Current opinion in biotechnology* 35 (2015), pp. 73–77.

[116] Rima Kaddurah-Daouk, Richard M Weinshilboum, and Pharmacometabolomics Research Network. "Pharmacometabolomics: implications for clinical pharmacology and systems pharmacology". In: *Clinical Pharmacology & Therapeutics* 95.2 (2014), pp. 154–167.

[117] Hisayuki Erabi et al. “Kynurenic acid is a potential overlapped biomarker between diagnosis and treatment response for depression from metabolome analysis”. In: *Scientific reports* 10.1 (2020), pp. 1–8.

[118] Michail E Klontzas et al. “Metabolomics analysis of the osteogenic differentiation of umbilical cord blood mesenchymal stem cells reveals differential sensitivity to osteogenic agents”. In: *Stem cells and development* 26.10 (2017), pp. 723–733.

[119] Chenxia Hu et al. “Energy metabolism plays a critical role in stem cell maintenance and differentiation”. In: *International journal of molecular sciences* 17.2 (2016), p. 253.

[120] Lisa B Boyette et al. “Human bone marrow-derived mesenchymal stem cells display enhanced clonogenicity but impaired differentiation with hypoxic preconditioning”. In: *Stem cells translational medicine* 3.2 (2014), pp. 241–254.

[121] Holly M Wobma et al. “The influence of hypoxia and IFN-γ on the proteome and metabolome of therapeutic mesenchymal stem cells”. In: *Biomaterials* 167 (2018), pp. 226–234.

[122] Galya Ivanova et al. “Metabolomic and proteomic analysis of the mesenchymal stem cells’ secretome”. In: *Metabolomics-Fundamentals and Applications. InTech* (2016), pp. 43–66.

[123] Francisco J Vizoso et al. “Mesenchymal stem cell secretome: toward cell-free therapeutic strategies in regenerative medicine”. In: *International journal of molecular sciences* 18.9 (2017), p. 1852.

[124] Laura Frese, Petra E Dijkman, and Simon P Hoerstrup. “Adipose tissue-derived stem cells in regenerative medicine”. In: *Transfusion Medicine and Hemotherapy* 43.4 (2016), pp. 268–274.

[125] Tiago Pereira et al. “MSCs conditioned media and umbilical cord blood plasma metabolomics and composition”. In: *PloS one* 9.11 (2014), e113769.

[126] Amal Surrati et al. “Non-destructive characterisation of mesenchymal stem cell differentiation using LC-MS-based metabolite footprinting”. In: *Analyst* 141.12 (2016), pp. 3776–3787.

[127] Jeremy K Nicholson, John C Lindon, and Elaine Holmes. “’Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data”. In: *Xenobiotica* 29.11 (1999), pp. 1181–1189.

[128] Ute Roessner et al. “Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry”. In: *The Plant Journal* 23.1 (2000), pp. 131–142.

[129] Paul D Fraser et al. “Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids”. In: *The Plant Journal* 24.4 (2000), pp. 551–558.

[130] Panteleimon G Takis et al. “Uniqueness of the NMR approach to metabolomics”. In: *TrAC Trends in Analytical Chemistry* 120 (2019), p. 115300.

[131] Marc-Emmanuel Dumas et al. “Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study”. In: *Analytical chemistry* 78.7 (2006), pp. 2199–2208.

[132] Hector C Keun et al. “Analytical reproducibility in 1H NMR-based metabonomic urinalysis”. In: *Chemical research in toxicology* 15.11 (2002), pp. 1380–1386.

[133] Harald Günther. *NMR spectroscopy: basic principles, concepts and applications in chemistry*. John Wiley & Sons, 2013.

[134] Kihwan Choi et al. “Quantitative NMR as a Versatile Tool for the Reference Material Preparation”. In: *Magnetochemistry* 7.1 (2021), p. 15.

[135] Gregory R Fulmer et al. “NMR chemical shifts of trace impurities: common laboratory solvents, organics, and gases in deuterated solvents relevant to the organometallic chemist”. In: *Organometallics* 29.9 (2010), pp. 2176–2179.

[136] Harry S Rollema and Jan A Brinkhuis. “A 1H-NMR study of bovine casein micelles; influence of pH, temperature and calcium ions on micellar structure”. In: *Journal of Dairy Research* 56.3 (1989), pp. 417–425.

[137] Joana Pinto et al. “Human plasma stability during handling and storage: impact on NMR metabolomics”. In: *Analyst* 139.5 (2014), pp. 1168–1177.

[138] Li Chen et al. “An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization”. In: *Journal of Magnetic Resonance* 158.1-2 (2002), pp. 164–168.

[139] Yuanxin Xi and David M Rocke. “Baseline correction for NMR spectroscopic metabolomics data analysis”. In: *BMC bioinformatics* 9.1 (2008), pp. 1–10.

[140] Olivier Cloarec et al. “Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets”. In: *Analytical chemistry* 77.5 (2005), pp. 1282–1289.

[141] David S Wishart et al. “HMDB: a knowledgebase for the human metabolome”. In: *Nucleic acids research* 37.suppl_1 (2009), pp. D603–D610.

[142] Eldon L Ulrich et al. “BioMagResBank”. In: *Nucleic acids research* 36.suppl_1 (2007), pp. D402–D408.

[143] Jianguo Xia et al. “MetaboMiner–semi-automated identification of metabolites from 2D NMR spectra of complex biofluids”. In: *BMC bioinformatics* 9.1 (2008), pp. 1–16.

[144] Christian Ludwig and Mark R Viant. "Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox". In: *Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques* 21.1 (2010), pp. 22–32.

[145] Wiktor Koźmiński. "Simplified multiplet pattern HSQC-TOCSY experiment for accurate determination of long-range heteronuclear coupling constants". In: *Journal of Magnetic Resonance* 137.2 (1999), pp. 408–412.

[146] Kerem Bingol et al. "Customized metabolomics database for the analysis of NMR 1H–1H TOCSY and 13C–1H HSQC-TOCSY spectra of complex mixtures". In: *Analytical chemistry* 86.11 (2014), pp. 5494–5501.

[147] Kerem Bingol et al. "Unified and isomer-specific NMR metabolomics database for the accurate analysis of 13C–1H HSQC spectra". In: *ACS chemical biology* 10.2 (2015), pp. 452–459.

[148] Kerem Bingol et al. "Comprehensive metabolite identification strategy using multiple two-dimensional NMR spectra of a complex mixture implemented in the COLMARm web server". In: *Analytical chemistry* 88.24 (2016), pp. 12411–12418.

[149] Issam El Naqa and Martin J Murphy. "What is machine learning?" In: *machine learning in radiation oncology*. Springer, 2015, pp. 3–11.

[150] Arthur S Edison et al. "NMR: unique strengths that enhance modern metabolomics research". In: *Analytical Chemistry* 93.1 (2020), pp. 478–499.

[151] Ulf W Liebal et al. "Machine learning applications for mass spectrometry-based metabolomics". In: *Metabolites* 10.6 (2020), p. 243.

[152] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. "Deep learning in bioinformatics". In: *Briefings in bioinformatics* 18.5 (2017), pp. 851–869.

[153] Richard J Gilbert et al. "Genetic programming: A novel method for the quantitative analysis of pyrolysis mass spectral data". In: *Analytical Chemistry* 69.21 (1997), pp. 4381–4389.

[154] Animesh Acharjee et al. "Comparison of regularized regression methods for˜ omics data". In: *Metabolomics* 3.3 (2013), p. 1.

[155] Lior Rokach and Oded Maimon. "Clustering methods". In: *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 321–352.

[156] Ian T Jolliffe and Jorge Cadima. "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150202.

[157] Stephen C Johnson. "Hierarchical clustering schemes". In: *Psychometrika* 32.3 (1967), pp. 241–254.

[158] Jundong Li et al. “Feature selection: A data perspective”. In: *ACM Computing Surveys (CSUR)* 50.6 (2017), pp. 1–45.

[159] Tobias Kind et al. “A comprehensive urinary metabolomic approach for identifying kidney cancer”. In: *Analytical biochemistry* 363.2 (2007), pp. 185–195.

[160] B Azhagusundari, Antony Selvadoss Thanamani, et al. “Feature selection based on information gain”. In: *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 2.2 (2013), pp. 18–21.

[161] Richard D Khusial et al. “Development of a plasma screening panel for pediatric nonalcoholic fatty liver disease using metabolomics”. In: *Hepatology communications* 3.10 (2019), pp. 1311–1321.

[162] Xin Jin et al. “Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles”. In: *International workshop on data mining for biomedical applications*. Springer. 2006, pp. 106–115.

[163] Quanquan Gu, Zhenhui Li, and Jiawei Han. “Generalized fisher score for feature selection”. In: *arXiv preprint arXiv:1202.3725* (2012).

[164] Hui-Huang Hsu, Cheng-Wei Hsieh, et al. “Feature Selection via Correlation Coefficient Clustering.” In: *J. Softw.* 5.12 (2010), pp. 1371–1377.

[165] Roberto Muñoz, Mar Garcıa-Hernández, and Cristina Gómez-Aleixandre. “CVD of carbon nanomaterials: from graphene sheets to graphene quantum dots”. In: *Handbook Carbon Nano Mater* 7.7 (2015), p. 127.

[166] Aedan GK Roberts, Daniel R Catchpoole, and Paul J Kennedy. “Variance-based Feature Selection for Classification of Cancer Subtypes Using Gene Expression Data”. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2018, pp. 1–8.

[167] Riccardo Leardi, R Boggia, and M Terrile. “Genetic algorithms as a strategy for feature selection”. In: *Journal of chemometrics* 6.5 (1992), pp. 267–281.

[168] Hyun-Woo Cho et al. “Genetic algorithm-based feature selection in high-resolution NMR spectra”. In: *Expert systems with applications* 35.3 (2008), pp. 967–975.

[169] Harun Uğuz. “A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm”. In: *Knowledge-Based Systems* 24.7 (2011), pp. 1024–1032.

[170] Shang Lei. “A feature selection method based on information gain and genetic algorithm”. In: *2012 International Conference on Computer Science and Electronics Engineering*. Vol. 2. IEEE. 2012, pp. 355–358.

[171] Bjoern H Menze et al. “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data”. In: *BMC bioinformatics* 10.1 (2009), pp. 1–16.

[172] Yue Liu et al. “NMR and LC/MS-based global metabolomics to identify serum biomarkers differentiating hepatocellular carcinoma from liver cirrhosis”. In: *International journal of cancer* 135.3 (2014), pp. 658–668.

[173] Hayley Abbiss, Garth L Maker, and Robert D Trengove. “Metabolomics approaches for the diagnosis and understanding of kidney diseases”. In: *Metabolites* 9.2 (2019), p. 34.

[174] Jordi Merino et al. “A decade of genetic and metabolomic contributions to type 2 diabetes risk prediction”. In: *Current diabetes reports* 17.12 (2017), pp. 1–14.

[175] Caroline H Johnson, Julijana Ivanisevic, and Gary Siuzdak. “Metabolomics: beyond biomarkers and towards mechanisms”. In: *Nature reviews Molecular cell biology* 17.7 (2016), pp. 451–459.

[176] Animesh Acharjee et al. “Comparison of regularized regression methods for˜ omics data”. In: *Metabolomics* 3.3 (2013), p. 1.

[177] Xiaotao Shen et al. “Normalization and integration of large-scale metabolomics data using support vector regression”. In: *Metabolomics* 12.5 (2016), pp. 1–12.

[178] Ángel Sánchez-Illana et al. “Model selection for within-batch effect correction in UPLC-MS metabolomics using quality control-Support vector regression”. In: *Analytica chimica acta* 1026 (2018), pp. 62–68.

[179] Miaomiao Jiang et al. “Sparse partial-least-squares discriminant analysis for different geographical origins of Salvia miltiorrhiza by 1H-NMR-based metabolomics”. In: *Phytochemical Analysis* 25.1 (2014), pp. 50–58.

[180] Haiwei Gu et al. “Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer”. In: *Analytica chimica acta* 686.1-2 (2011), pp. 57–63.

[181] Katsuaki Nitta et al. “Orthogonal partial least squares/projections to latent structures regression-based metabolomics approach for identification of gene targets for improvement of 1-butanol production in Escherichia coli”. In: *Journal of bioscience and bioengineering* 124.5 (2017), pp. 498–505.

[182] Andrea Ganna et al. “Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease”. In: *PLoS genetics* 10.12 (2014), e1004801.

[183] Andrea Ganna et al. “Large-scale non-targeted metabolomic profiling in three human population-based studies”. In: *Metabolomics* 12.1 (2016), pp. 1–13.

[184] Mika Hilvo et al. “Monounsaturated fatty acids in serum triacylglycerols are associated with response to neoadjuvant chemotherapy in breast cancer patients”. In: *International Journal of Cancer* 134.7 (2014), pp. 1725–1733.

[185] Victoria L Stevens et al. “Serum metabolomic profiles associated with postmenopausal hormone use”. In: *Metabolomics* 14.7 (2018), pp. 1–14.

[186] Christopher W Armstrong et al. "Metabolic profiling reveals anomalous energy metabolism and oxidative stress pathways in chronic fatigue syndrome patients". In: *Metabolomics* 11.6 (2015), pp. 1626–1639.

[187] Etienne A Thévenot et al. "Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses". In: *Journal of proteome research* 14.8 (2015), pp. 3322–3335.

[188] Xiaojiao Zheng et al. "Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice". In: *BMC biology* 15.1 (2017), pp. 1–15.

[189] Johannes F Fahrmann et al. "Investigation of metabolomic blood biomarkers for detection of adenocarcinoma lung cancer". In: *Cancer Epidemiology and Prevention Biomarkers* 24.11 (2015), pp. 1716–1723.

[190] M Kuboniwa et al. "Prediction of periodontal inflammation via metabolic profiling of saliva". In: *Journal of dental research* 95.12 (2016), pp. 1381–1386.

[191] Eric A Franzosa et al. "Gut microbiome structure and metabolic activity in inflammatory bowel disease". In: *Nature microbiology* 4.2 (2019), pp. 293–305.

[192] Angela W Chan et al. "1 H-NMR urinary metabolomic profiling for diagnosis of gastric cancer". In: *British journal of cancer* 114.1 (2016), pp. 59–62.

[193] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "The elements of statistical learning. Springer series in statistics". In: *:* Springer, 2001.

[194] Kevin M Mendez, Stacey N Reinke, and David I Broadhurst. "A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification". In: *Metabolomics* 15.12 (2019), pp. 1–15.

[195] Thao Vu et al. "Evaluation of multivariate classification models for analyzing NMR metabolomics data". In: *Journal of proteome research* 18.9 (2019), pp. 3282–3294.

[196] Hai-Yan Fu et al. "Challenges of large-class-number classification (LCNC): a novel ensemble strategy (ES) and its application to discriminating the geographical origins of 25 green teas". In: *Chemometrics and Intelligent Laboratory Systems* 157 (2016), pp. 43–49.

[197] Mikel Elkano et al. "Fuzzy rule-based classification systems for multi-class problems using binary decomposition strategies: on the influence of n-dimensional overlap functions in the fuzzy reasoning method". In: *Information Sciences* 332 (2016), pp. 94–114.

[198] Peter Clark and Robin Boswell. "Rule induction with CN2: Some recent improvements". In: *European Working Session on Learning*. Springer. 1991, pp. 151–163.

[199] Carl Brunner et al. "Pairwise support vector machines and their application to large scale problems". In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 2279–2292.

[200] Patrick J Trainor, Andrew P DeFilippis, and Shesh N Rai. “Evaluation of classifier performance for multiclass phenotype discrimination in untargeted metabolomics”. In: *Metabolites* 7.2 (2017), p. 30.

CHAPTER 2

METABOLOMICS AND CYTOKINE PROFILING OF MSCs IDENTIFIES MARKERS PREDICTIVE OF T CELL SUPPRESSION

[1]

[1]Shen, X.*,Maughon, T. S.*, Huang, D., Michael, A. O. A., Shockey, W. A., Andrews, S. H., ... Marklein, R. A. (2021). Metabolomics and cytokine profiling of mesenchymal stromal cells identify markers predictive of T-cell suppression. Cytotherapy. Reprinted here with permission from the publisher. *: co-first authors

Authors contributions:

This paper is published in journal Cytotherapy. I am a co-first author with Ty Maughon. In this project, I participated in conceptual study design with Ty Maughon, Arthur S Edison, and Steven L Stice. Ty Maughon and Seth H Andrews cultured the cells, collected cell pellets at the end of culture as well as media for cytokines analysis and performed T cell suppression and IDO assay. I prepared the cells for NMR and MS analysis, performed statistical power analysis and optimized NMR sample preparation protocol. I also collected NMR metabolomics data, and performed data pre-processing. Danning Huang collected MS metabolomics data in Facundo Fernandez's laboratory. William A Shockey performed cytokine analysis. I performed the data modeling using NMR and MS metabolomics data, cytokines data, T cell suppression and IDO activity results under the guidance provided by Manu O Platt and Ross A Marklein. I participated in writing the manuscript draft and revising process together with Ty S Maughon, Danning Huang, Adeola O Adebayo, and other authors.

Abstract

Background: Mesenchymal stromal cells (MSCs) have shown great promise in the field of regenerative medicine as many studies have shown that MSCs possess immunomodulatory function. Despite this promise, no MSC therapies have been licensed by the FDA. This lack of successful clinical translation is due in part to MSC heterogeneity and a lack of critical quality attributes (CQAs). While MSC Indoleamine 2,3-dioxygnease (IDO) activity has been shown to correlate with MSC function, multiple predictive markers may be needed to better predict MSC function.

Methods: Three MSC lines (two bone marrow, one iPSC) were expanded to three passages. At the time of harvest for each passage, cell pellets were collected for nuclear magnetic resonance (NMR) and ultra-performance liquid chromatography mass spectrometry (UPLC-MS), and media was collected for cytokine profiling. Harvested cells were also cryopreserved for assessing function using T cell proliferation and IDO activity assays. Linear regression was performed on functional data against NMR, MS, and cytokines to reduce the number of important features,

and partial least squares regression (PLSR) was used to obtain predictive markers of T cell suppression based on variable importance in projection (VIP) scores.

Results: Significant functional heterogeneity (in terms of T cell suppression and IDO activity) was observed between the three MSC lines, as well as donor-dependent differences based on passage. Omics characterization revealed distinct differences between cell-lines using principal component analysis (PCA). Cell-lines separated along principal component 1 based on tissue source (bone marrow vs. iPSC-derived) for NMR, MS, and cytokine profiles. PLSR modeling of important features predicts MSC functional capacity with NMR (R2=0.86), MS (R2=0.83), cytokines (R2=0.70), and a combination of all features (R2=0.88).

Discussion: The work described here provides a platform for identifying markers for predicting MSC functional capacity using PLSR modeling that could be used as release criteria and guide future manufacturing strategies for MSCs and other cell therapies.

2.1 Introduction

Mesenchymal stem/stromal cells (MSCs) have been explored as a cell therapy in clinical trials due to their immunomodulatory properties [1]. Although MSCs have shown great promise in preclinical studies for treatment of immune diseases, there have been challenges translating MSCs into approved therapies. This lack of translation can be attributed to MSC heterogeneity and no well-established critical quality attributes (CQAs i.e. limits or ranges of MSC biological properties) used to monitor MSC functional capacity among cell-lines or even within MSC cultures [2–4]. T cell suppression is one of the most commonly used assays to assess MSC immunomodulatory capacity, but it is not standardized, has low throughput, and there is donor-donor variability among peripheral blood mononuclear cell (PBMC) responses to stimulation [5].

The International Society for Cell and Gene Therapy (ISCT) has proposed several candidate protein properties to use for predicting MSC functional capacity [2]. Indoleamine 2,3-dioxygenase (IDO) is one indicator of potency that has been proposed that plays a major role in the mechanisms by which

MSCs modulate immune cells, such as T cells [6]. Although IDO correlates with T cell suppression, there are several other mechanisms through which MSCs exert their potent immunomodulatory effects [7–9]. Therefore, a matrix-based approach (i.e. a combination of predictive markers) is likely needed to ensure a high quality MSC product [2]. Chinnadurai et al. examined the relationship between various RNAs and secreted molecules with T cell proliferation and showed that several secreted factors and RNAs had strong correlations with T cell proliferation [10]. Although these factors correlated with MSC functional capacity, measurement of these cellular properties require stimulation of sample MSC product acquired at the end of the manufacturing process and thus cannot be performed in-process.

MSCs secrete a broad repertoire of immunomodulatory cytokines that may have therapeutic potential; however, these immunomodulatory functions have not yet been characterized. Several studies have shown that MSC secreted cytokines can be modulated by physiologic conditions such as hypoxia, and pharmacological conditions such as targeted small molecule and growth factor conditioning [11–13]. MSCs from different sources may differ in their immunomodulatory capacities, and the ability to assess cytokine secretion in MSCs using a non-destructive approach such as cytokine assays will allow for a standardized metric for MSC potency during manufacturing [14]; cytokines that are secreted into the media can be measured without interference with cell growth and other conditions. Therefore, combination of specific cytokines secreted by MSCs, together with metrics of T cell suppression and IDO may enable identification of non-destructive, in-process markers for predicting potency.

Measures of cellular metabolism are also promising for assessing MSC quality due to the high abundance of metabolites in cells, and their importance in stem cell fate [15]. Studies have shown that extended in vitro culture of MSCs shifts their metabolism from glycolysis towards oxidative phosphorylation (OXPHOS). MSCs in their native environment have a more glycolytic metabolism, and it has been shown that glycolytic MSCs have improved immunomodulatory effects in vivo [15–17]. Therefore, assessing metabolite profiles in MSCs during expansion could be used as a predictor

of their immunomodulatory capacity. Non-targeted metabolomics enables a detailed profiling of therapeutic cells, providing opportunities towards a more precise understanding of cellular therapeutic mechanisms [18]. Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are the two most commonly used techniques in metabolomics [19]. NMR requires minimal sample preparation, making it highly reproducible. NMR can also provide information in assigning metabolites based on chemical shifts and J-coupling patterns. As an analytical platform, MS also has certain advantages. Its higher sensitivity enables the detection of low abundance metabolites that are below the NMR detection thresholds [20], whereas its high resolution greatly reduces spectral overlap. When coupled with separation techniques such as gas chromatography (GC) or ultra-performance liquid chromatography (UPLC), spectral complexity is greatly reduced, and metabolic chemical properties can be revealed [21–23]. The combination of NMR and UPLC-MS metabolic profiling provides an even more in-depth measurement of MSC cellular metabolism, potentially leading to the discovery of CQAs. To date, NMR- and MS-based metabolic profiling was used to characterize cellular metabolism, leading to the discovery of biomarkers or pathways different between cell-lines or cellular responses of treatment [24–26]. However, to the best of our knowledge, no studies have been reported where NMR- and/or UPLC-MS- based metabolomics are used to establish predictive markers associated with MSC immunomodulatory potency.

In this study, we measured MSC metabolites and cytokine levels during cell manufacturing that could serve as predictors of MSC immunomodulation and as potential potency assays. These metabolites and cytokines are referred to in this work as predictive markers. Intracellular metabolites and secreted cytokines from three MSC lines at multiple passages were studied to determine their correlation with MSC immunosuppressive capacity post thaw. NMR, UPLC-MS, and cytokine data sets were also merged and filtered to identify predictive markers. Using partial least squares regression (PLSR) and variable importance projection (VIP) scores, we identified a panel of candidate cytokines and metabolites that could be used to predict MSC functional capacity and inform future manufacturing strategies.

2.2 Methods

2.2.1 Cell culture

Two bone marrow-derived MSC lines (RoosterBio, Frederick MD) lot 0071 (F, 18-30) and 0182 (F, 26) (which the manufacturer has both research and clinical-grade lots available), and one induced pluripotent stem cell derived MSC cell-line (Cellular Dynamics International, Madison WI) (Lot 0003, also prequalified) were used and referred here as BM71, BM182, and iMSC, respectively. MSCs were thawed and allowed to recover for 24 hours in complete medium (MSC-GM) (Alpha-Minimum Essential Medium (Gibco, Waltham MA), 10% fetal bovine serum (Hyclone, Logan UT), 2mM L-glutamine, 50 U/mL penicillin, 50 μg/mL streptomycin (Gibco)) before being seeded at 500 cells/cm^2. Cells were expanded in 10 150 mm plates with a negative, media-only control plate and 14 T175 flasks for expansion. After reaching approximately 80% confluency, cells in 150 mm dishes were washed with PBS three times. Cells were then scraped and collected in 80:20 methanol:water solution and stored at -80$^{\circ}C$ for NMR and MS analysis. Cells grown in T175 flasks for expansion were harvested using 0.05% trypsin (Gibco) and counted using a Cellometer K2 cell counter (Nexcelom, Lawrence MA). Cells were either cryopreserved for functional assays or reseeded in dishes/flasks for continued expansion (3 total passages for each cell-line, 9 total experimental groups). Population doubling level (PDL) for each cell-line/passage was determined using formula (2.1):

$$PDL = X + 3.322(logY - logI) \tag{2.1}$$

Where X = initial PDL, I = initial cell seeding number and Y = final number of cells.

2.2.2 T cell suppression assay

MSCs from each cell-line/passage were thawed and allowed to recover for 48 hours with a media change at 24 hours. MSCs were trypsinized (0.05% Trypsin-EDTA, Gibco) then seeded at three densities (10,000, 5,000 and 2,000 cells/well) in a 96 well plate and cultured for 24 hours. PBMCs (AllCells, Alameda CA) were thawed in Roswell Park Memorial Institute (RPMI) media (RPMI, 20% FBS, 2mM L-glutamine, 50 U/mL penicillin, 50 μg/mL streptomycin) and cultured overnight at 37°C and 5% CO_2. Prior to co-culture, PBMCs were labeled with carboxyfluorescein succinimidyl ester (CFSE) (Table 2.1, Biolegend, San Diego CA) according to the manufacturer's protocol, and 100,000 PBMCs were added to each well at final MSC:PBMC ratios of 1:10, 1:20, or 1:50 as well as control wells containing only PBMCs. Following PBMC addition, stimulating anti-CD3/CD28 Dynabeads (Thermo Fisher Scientific, Waltham MA) were added at 100,000 beads per well to the appropriate wells (positive controls and all MSC groups). MSCs and PBMCs were co-cultured for 72 hours at 37°C, 5% CO_2.

Following co-culture, PBMCs were collected and stained using APC/Fire anti-CD4 and APC anti-CD8 (Table 2.1) (Biolegend, San Diego CA). PBMCs were first washed and stained with Zombie Yellow (Table 2.1) (Biolegend, San Diego CA) viability dye and blocked using 2% FBS. PBMCs were then washed again and stained for CD4 and CD8 in the dark at room temperature. Following staining, the antibodies were blocked using 2% FBS and washed. Cells were then fixed with 4% PFA for 30 minutes at 4°C. Cells were then washed and re-suspended in PBS containing 2% FBS. Cells were stored overnight at 4°C in the dark until flow analysis.

2.2.3 Flow cytometry

All flow cytometry experiments were performed using a CytoFLEX S (Beckman Coulter, Hialeah FL) with 20,000 events collected per sample. All flow cytometry data were analyzed using FlowJo (Treestar, Inc., Ashland OR). Briefly, cell debris, doublets, and Dynabeads were gated out using

scatter principles. Then, single stained controls were used for compensation, and fluorescence minus one controls were used in order to determine positive populations (Supplemental Figure 2.3.1).

Table 2.1: PBMC co-culture assay antibodies and flow cytometer information.

Laser	EM Filter	Marker	Color/Format	Host/Target	Isotype	Clone	Company	Catalog
638	780/30	CD4	APC/Fire	Mouse anti-Human	IgG1 κ	RPA-T4	BioLegend	67-0047-T025
638	660/20	CD8	APC	Mouse anti-Human	IgG1	CN9V1	BioLegend	orb248718
488	525/40	CFSE	CFSE	N/A	N/A	N/A	BioLegend	423801
405	610/20	Zombie Yellow	Zombie Yellow	N/A anti-All Species	N/A	N/A	BioLegend	423104

2.2.4 IDO activity assay

MSCs from each cell-line/passage experimental group were thawed and cultured for 48 hours with a media change at 24 hours. MSCs were trypsinized then seeded at a density of 40,000 cells/cm^2 in a 96 well plate in MSC-GM. After 24 hours, the medium was replaced in each well with MSC-GM containing 10 ng/mL interferon gamma (IFN-) (Life Technologies). After an additional 24 hours, conditioned media was collected and frozen at -20°C, and cells were fixed with 4% PFA. Media was thawed and 100 μL was transferred into a 96 well plate. Trichloroacetic acid was used to precipitate excess protein. 75 μL of the supernatant was collected and transferred to a separate 96 well plate. Ehrlich's Reagent was then added to each well to detect L-kynurenine levels using a SpectraMax iD5 (Molecular Devices) plate reader. Levels of L-kynurenine were determined using a standard curve. To normalize L-kynurenine values to cell numbers for each experimental group, we performed automated image analysis to quantify cell nuclei in the wells from which conditioned medium was collected. Following fixation, MSCs were washed with PBS twice, and stained with Hoechst. Cells were then imaged on a Cytation 5 high content imaging system (Biotek, Winooski VT) and cell counts determined using CellProfiler[27] to normalize the amount of L-kynurenine per cell.

2.2.5 Metabolomics sample preparation

MSC samples stored in 80:20 methanol:water solution were thawed at 4°C and vortexed three times for 1 min. Then samples were centrifuged at 14,000 x g for 5 min at 4°C. For each sample, 30% (300 μl) and 60% (600 μl) of the supernatant were transferred to pre-labeled Eppendorf tubes for LC-MS and NMR spectroscopy, respectively. For each PDL, 67μl of each sample was pooled together to generate 2 internal PDL quality control (QC) samples. The remaining 33 μl supernatant of each sample was pooled together to generate 2 internal overall QC samples. In each PDL, the extraction blank sample was added using extraction solvent (methanol: water 80:20). Samples were then evaporated in a Speedvac for 6 hours and stored in -80°C until future analysis. NMR samples with QC controls and 2 buffer blank samples were used for data acquisition. Samples were randomized, with a total 49 samples in each cell-line. The LC-MS sample randomization was identical to NMR.

2.2.6 NMR

The NMR buffer solution was prepared by dissolving 928.6 mg of anhydrous NaH_2PO_4 and 320.9 mg of Na_2HPO_4 in 80 ml D_2O (Cambridge Isotope Laboratory) in a volumetric flask. Sodium trimethylsilylpropanesulfonate (DSS) was used as a chemical shift and concentration reference standard by adding 333.3 μL of 1.0 M DSS-D6 (Cambridge Isotope Laboratory) stock solution to the buffer for a final DSS concentration of 1/3 mM. The pH was adjusted to 7.4 (uncorrected for isotope effects) and was brought to a volume of 100 ml with D2O and mixed well. The pH was rechecked, and the buffer stored at 4°C until use.

The NMR samples were reconstituted in 80 μL of the NMR buffer and vortexed thoroughly. Sixty μL of each sample was transferred to racks of 96 1.7-mm NMR tubes for data acquisition using a SamplePro Tube robotic system (Bruker Biospin, Billerica, MA, USA). Samples were run on a Bruker NEO 800 MHz NMR spectrometer equipped with a 1.7-mm cryoprobe and Bruker

SampleJet cooled to 5.6°C. One dimensional nuclear Overhauser enhancement spectroscopy with water suppression (1D-NOESY PR) was collected on all samples. The spectra were processed using NMRPipe[28] and in-house MATLAB metabolomics toolbox. The spectra were aligned using Correlation Optimized Wrapping (COW) algorithm[29] and normalized with a Probabilistic quotient normalization (PQN)[30] algorithm. The non-overlapped peaks were manually binned, and the area under curve was calculated for each non-overlapped feature.

Two-dimensional ^{1}H-^{1}3C Heteronuclear Single Quantum Coherence (HSQC) and ^{1}H-^{1}3C HSQC-Total Correlated Spectroscopy (HSQC-TOCSY) spectra on internal pooled samples were collected for metabolite identification. The spectra were processed using NMRPipe, and the in-house MATLAB metabolomics toolbox. The 2D NMR spectra were matched to a metabolite database using the COLMARm [31]. The metabolites were assigned a confidence level ranging from 1 to 5 according to published criteria [32].

After metabolite identification, peaks were further normalized by the number of protons in each functional group corresponding to a given peak. A total of 28 metabolites were obtained from binning of the whole spectra. The unknown binned features were used to perform correlation analysis. The features with correlation coefficient values greater or equal to 0.8 were then grouped together as tentative unknown metabolites. A total of 29 tentative unknown features were extracted from 100 features (Figure 2.5.2).

2.2.7 UPLC-MS

Cell extract samples and a sample blank were resuspended in 50 μL methanol/water (80:20 v/v), and a pooled quality control (QC) sample was created by mixing a 10 μL aliquot of each cell extract sample. Both the sample blank and the pooled sample were processed with the same procedure, and were analyzed together with the cell extract samples. All samples were run in randomized order on consecutive days. QC samples were analyzed every 10 runs to assess UPLCMS system stability and correct time-dependent batch effects with a QC-based regression curve.

UPLC-MS analyses were performed using a Vanquish Horizon UPLC (Thermo Fisher Scientific, Inc., Waltham, MA) system coupled to a Orbitrap ID-X Tribrid mass spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA). Hydrophilic interaction (HILIC) chromatography was performed with a Waters ACQUITY UPLC BEH HILIC, 2.1 × 75 mm, 1.7 μm particle column. Mobile phase A was water/acetonitrile (95:5 v/v), 10 mM ammonium acetate, and 0.05% ammonium hydroxide. Mobile phase B was acetonitrile with 0.05% ammonium hydroxide. Chromatographic gradients can be found in Supplemental Information (Table 2.2). The column temperature was 55 °C, while samples were maintained at 5 °C in the autosampler. Two μL of each sample were injected, and the mass spectrometer was operated in positive ion mode. For metabolite identification purposes, data-dependent acquisition (DDA) experiments were used to collect MS/MS spectra at stepped normalized collision energy (NCE) of 10, 30, and 50.

Table 2.2: Methods information. Chromatographic gradient for HILIC method: mobile phase A was water/acetonitrile (95:5 v/v), with 10 mM ammonium acetate and 0.05% ammonium hydroxide, and B was acetonitrile with 0.05% ammonium hydroxide. Both ionization modes utilized identical mobile phases and identical chromatographic gradients.

HILIC Separation Method				
Time (min)	A	B	Flow rate (ml/min)	Curve
0	5	95	0.4	5
0.5	5	95	0.4	6
9	70	30	0.4	6
9.4	70	30	0.4	6
9.5	5	95	0.4	6
11	5	95	0.4	6
12	5	95	0.4	6

2.2.8 Cytokine profiling

Conditioned media from each cell-line/passage (at the time of cell harvest) were collected and prepared for cytokine profiling using the Human Premixed Multi-Analyte Magnetic Luminex Assay (RD Systems, Minneapolis MN) that was customized for this study quantifying cathepsin S, cystatin C, GCSF, angiopoietin 1, VEGF, PDGF-AA, bFGF, osteopontin, IL-8, IL-6, CCL12/MCP-1, and CCL5/RANTES (Table 2.3). Cytokine profiling was performed following a standard protocol. Briefly, 50 μL standard or media sample was added to a 96-well plate, followed by 50 μL diluted microparticle cocktail and incubated for 2-hours at room temperature on an orbital microplate shaker. Using a magnetic device attached to the bottom of the microplate, each well was washed with 50 μL of wash buffer. Then 50 μL diluted biotin-antibody cocktail was added to wells and incubated for 1-hour at room temperature on an orbital microplate shaker. After washing, 50 μL of streptavidin-PE was added to each well and incubated for 30 min. 100 μL of wash buffer was added to wells to resuspend microparticles and samples were then loaded into a BioPlex 200 Luminex system (Bio-Rad) and analyzed. The assay measures the intensity of PE emission, which is correlated to cytokine concentration standards.

Table 2.3: List of all cytokines measured and their corresponding function.

Analyte	Function
Angiopoietin-1	Promote angiogenesis and vascular growth
CCL2/MCP-1	Recruits immune cells to sites of inflammation
Cystatin C	Inhibitor of cysteine cathepsin proteases
IL-8/CXCL8	Stimulates neutrophil migration and phagocytosis
PDGF-AA	Cell growth, differentiation and survival
Cathepsin S	Proteolysis, antigen presentation, extracellular functions
CCL5/RANTES	Recruits immune cells to sites of inflammation
FGF basic	Numerous, cell growth and development, embryonic development
IL-6	Numerous, pro-inflammatory signaling
Osteopontin/OPN	Cell migration, adhesion and survival
VEGF	Encourages formation of blood vessels
G-CSF	Stimulates granulocyte and stem cell production and release

2.2.9 Data analysis and statistical methods

Principal component analysis (PCA) was conducted using JMP (version 15). All suppression assays, linear regression models, and IDO assays were analyzed using Prism (GraphPad, San Diego CA). Comparison of T cell proliferation to the positive control, comparison of T cell proliferation among all groups at the 1:10 dilution, and IDO activity were analyzed using one-way ANOVA with Tukey's post-hoc test. Comparisons of T cell suppression against passage 1 and against the 1:10 dilution were analyzed using a two-way ANOVA with Tukey's post hoc test. Error bars represent standard error on all bar graphs.

NMR and MS features were mapped to $CD4^+$ and $CD8^+$ proliferation rates using simple linear regression. The top 20 metabolites of each dataset with the lowest p-values ($p<0.05$) were selected

for downstream analysis. The selected 20 NMR metabolites, 20 MS metabolites and all cytokine data were used to train a partial least squares regression (PLSR) model. The response was the PCA loading score generated by using principal component 1 of all five functional measures termed, composite functional score. The predictors with variable importance in projection (VIP) scores greater or equal to 1 were selected as predictive markers.

2.3 Results

2.3.1 Cell-line and passage-dependent differences in T cell suppression

PDL was used to assess MSC age and growth characteristics for each cell-line throughout expansion (Table 2.4). PDL, doublings per day, and log growth of cell expansions were recorded (Figure 2.3.1B-D). Following expansion, MSC cell-lines at three different passages were co-cultured with anti-CD3/CD28 stimulated PBMCs in order to assess their immunosuppressive capacity with two PBMC donors. MSCs were cultured at three different MSC:PBMC ratios as well as positive (no MSCs with stimulation) and negative (no MSCs, no stimulation) controls. After 3 days, $CD4^+$ and $CD8^+$ T cell proliferation was assessed based on CFSE dilution (Figure 2.3.2A-F). A dose response was observed when increasing the MSC:PBMC ratio on the ability of the iMSC and BM182 cells to suppress $CD4^+$ and $CD8^+$ proliferation ($p<0.05$). $CD4^+$ and $CD8^+$ T cell proliferation decreased at the 1:50 dilution compared to the 1:10 dilution ($p<0.05$) in BM71 except for P3. The 1:10 MSC:PBMC ratio had the greatest variation in function across all cell-lines and passages and was used for comparison and predictive function. All passages of the iMSCs and BM182 lines suppressed $CD4^+$ and $CD8^+$ T cell proliferation when compared to the positive control ($p<0.05$) (Figure 2.3.2A, C, D, F). BM71 increased $CD4^+$ proliferation at all passages and increased $CD8^+$ proliferation at P1 ($p<0.05$) and had no significant differences at P2 and P3 when compared to the positive control (Figure 2.3.2B, E). For BM182, $CD4^+$ and $CD8^+$ T cell proliferation ($p<0.05$) was suppressed at P3 compared to earlier passages and had similar suppression as the iMSCs (Figure

2.3.2G). These studies were repeated using PBMCs harvested from a second donor (Figure 2.5.2). Similarly, there was a MSC dose dependent response, as the 1:10 MSC:PBMC ratio had significantly lower ($p<0.05$) $CD4^+$ and $CD8^+$ proliferation when compared with 1:20 and 1:50 ratios (Figure 2.5.2 A-F). Although BM71 did suppress both $CD4^+$ (except P1) and $CD8^+$ T cells when compared to the positive control ($p<0.05$), they possessed different immunomodulatory capacities than the iMSCs and BM182 ($CD4^+$ and $CD8^+$ $p<0.05$) at the 1:10 ratio (Figure 2.5.2G,H). Again, iMSCs consistently suppressed both $CD4^+$ and $CD8^+$ proliferation across all passages. BM182 P3 had significantly less $CD4^+$ proliferation compared to P1 ($p<0.05$) and P2 ($p<0.05$). BM182 P3 also had significantly less $CD8^+$ proliferation compared to P1 ($p<0.05$), was less than P2 although not technically significant ($p=0.06$) (Figure 2.5.2G, H).

Table 2.4: PDL and doublings per day values of each cell-line at each passage.

Cell-Line/Passage	Population Doubling Level (PDL)	Doublings per Day
iMSC P1	5.7	0.63
iMSC P2	11.7	0.74
iMSC P3	17.1	0.68
BM71 P1	3.9	0.56
BM71 P2	8.6	0.67
BM71 P3	12.7	0.59
BM182 P1	4.1	0.59
BM182 P2	7.9	0.47
BM182 P3	11.5	0.45

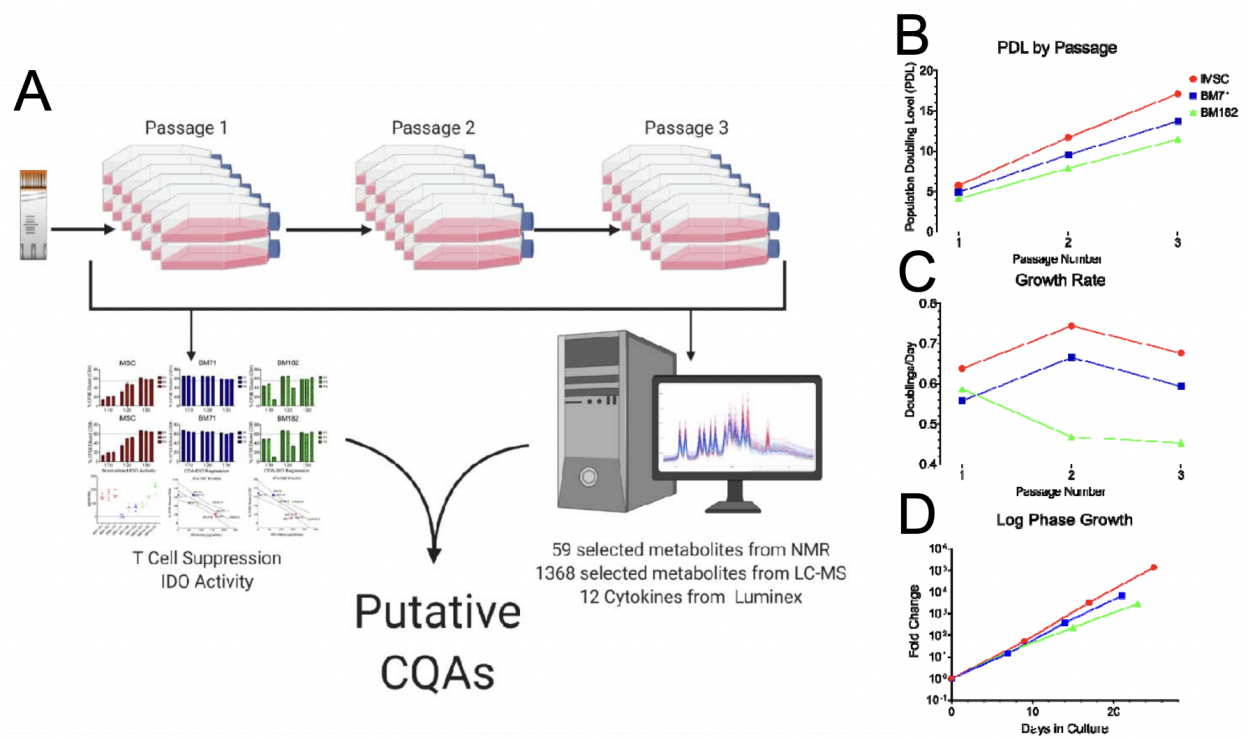


Figure 2.3.1: MSC expansion workflow and growth characteristics. (A) MSC characterization workflow from expanded cells to discover predictive markers. (B) Population doubling level changes of three MSC lines over three passages. (C) MSC growth rate in doublings per day over three passages. (D) Log phase growth characteristics of each MSC line over the number of days in culture.

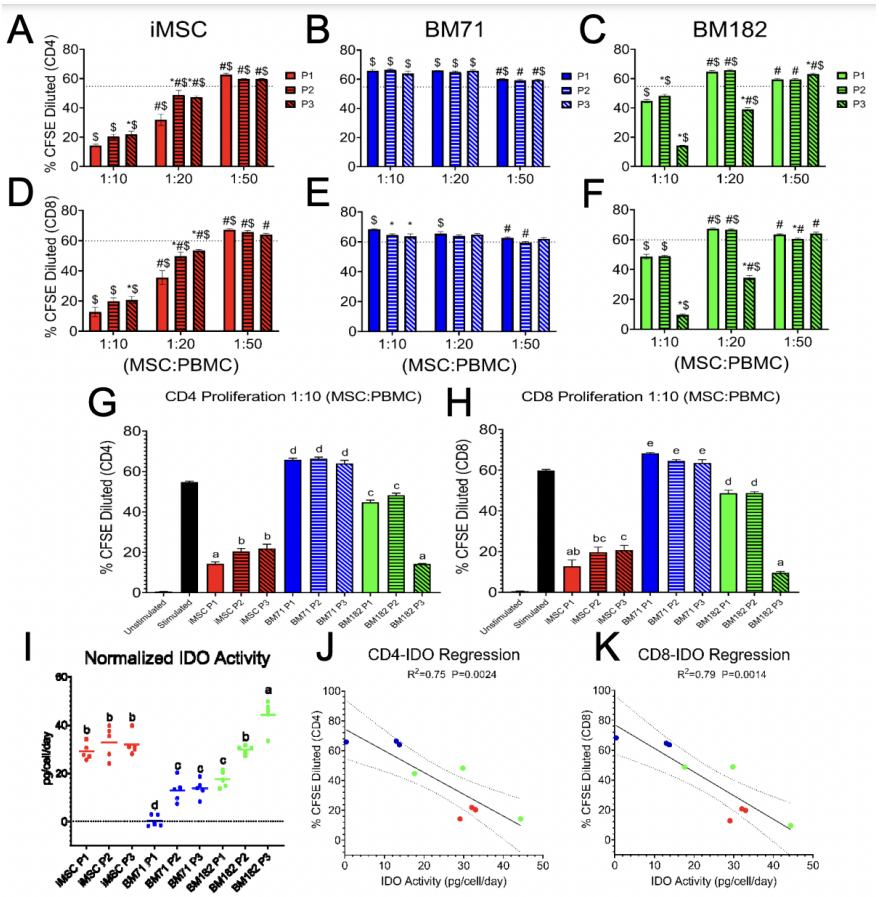


Figure 2.3.2: MSC functional capacity characterized by T-cell proliferation and IDO activity. MSCs were co-cultured with stimulated PBMCs at three different MSC:PBMC ratios (1:10, 1:20, 1:50). (A-C) $CD4^+$ and (D-F) $CD8^+$ T cell proliferation was assessed at each passage and ratio by % CFSE dilution. A 2-way ANOVA was used in order to determine if there was a significant difference from P1 within a ratio (*), a significant difference from the 1:10 ratio within a passage (), or a significant difference from the stimulated control ($) with $p<0.05$ for all significance. (G) $CD4^+$ and (H) $CD8^+$ T cell proliferation comparing all cell-lines and passages at the 1:10 ratio, and (I) IDO activity measured by L-kynurenine levels normalized to cell number and days in culture were analyzed using a one-way ANOVA with Tukey's post hoc test to determine significance ($p<0.05$) . Linear regression of the relationship between IDO activity and (J) $CD4^+$ and (K) $CD8^+$ T cell proliferation. Groups not sharing the same lettering in G, H, and I represents a statistical difference ($p<0.05$).

2.3.2 MSC IDO activity correlates with T cell proliferation

IDO activity for each cell-line/passage group was quantified as another functional readout of MSC immunomodulatory capacity (Figure 2.3.2I). Similar to the T cell suppression results, iMSCs displayed high IDO activity (indicated by high levels of L-kynurenine) for all passages and BM71 had the lowest IDO activity for all cell-lines at each passage. IDO activity of BM182 increased from P1 to P2 ($p<0.05$) with P3 displaying the highest levels of L-kynurenine ($p<0.05$) compared to all other cell-lines. Linear regression was performed to determine the relationship between IDO activity and $CD4^+$/$CD8^+$ T cell proliferation (Figure 2.3.2J, K). IDO activity correlated with both $CD4^+$ and $CD8^+$ T cell proliferation (R^2=0.75 and R^2=0.79, respectively) for the 1:10 MSC:PBMC ratio. The second PBMC donor was consistent with the first PBMC donor, IDO activity and MSC suppression of $CD4^+$/$CD8^+$ T cells (R^2=0.85 and R^2=0.81, respectively) were positively correlated (i.e higher IDO activity means higher T cell suppression) (Figure 2.5.3).

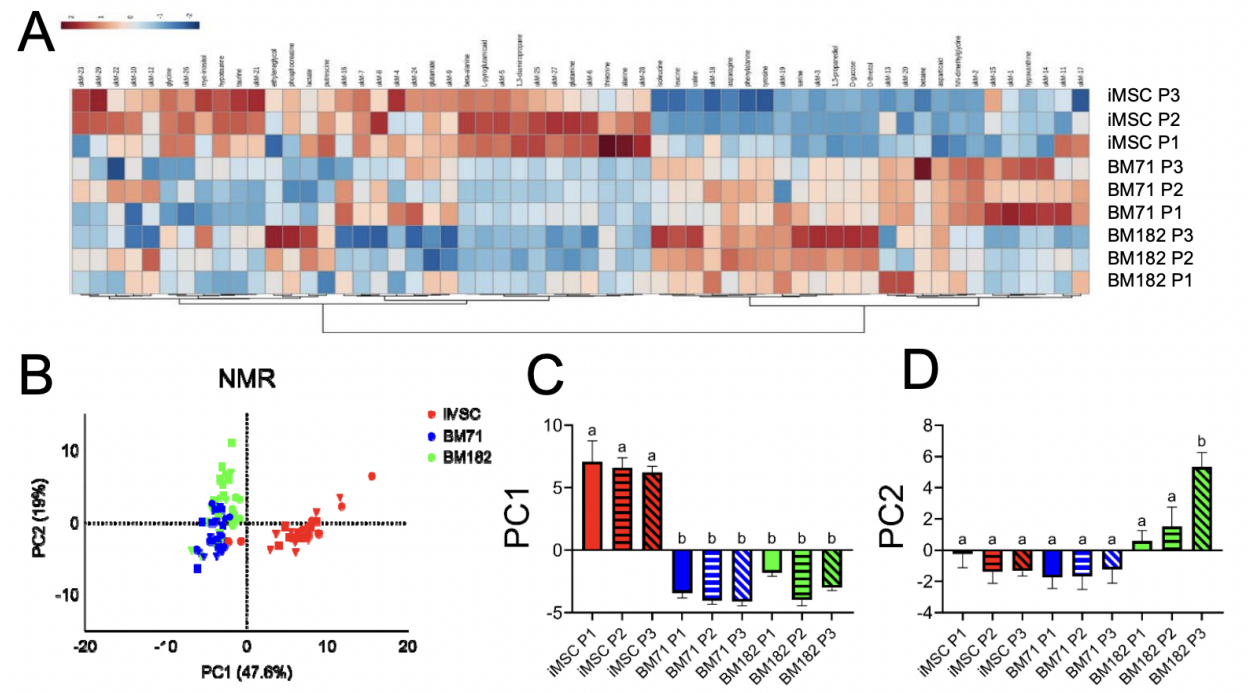


Figure 2.3.3: MSC metabolic profile from NMR analysis. (A) Heat map of all 58 metabolites using Euclidian distance measure and ward clustering method. (B) Unsupervised PCA of all metabolites for each cell-line at three passages (P1=circle, P2=triangle, P3=square). (C) One-way ANOVA comparison of average PC1 value. (D) One-way ANOVA comparison of average PC2 value. Groups not sharing the same lettering in C and D represents a statistical difference ($p<0.05$).

2.3.3 NMR-based identification of MSC metabolic signatures

For each cell-line/passage, the intracellular products were analyzed using NMR metabolomics profiling. The heat map of all detected metabolites shows that the overall metabolite signature was cell-line specific with some passage differences observed within cell-lines (Figure 2.3.3A). It provides intuitive visualization of a data table. Each colored cell on the map corresponds to a concentration value in our data table, with samples in rows and features/compounds in columns. The data was auto scaled and normalized. As an unbiased way of looking at the metabolic signature

of MSCs, we used unsupervised PCA to examine differences in the cell-lines. Unsupervised PCA revealed distinct metabolic differences between the cell-lines (Figure 2.3.3B). Different tissue donor sources clearly separated along PC1 with iMSC samples grouped together on the positive side of PC1 and the both BM cell-lines on the negative side. iMSCs were significantly different ($p<0.05$) from BM71 and BM182 along PC1, but there were no observable differences between BM71 and BM182 (Figure 2.3.3C). BM182 separated from BM71 along PC2 at P3 ($p<0.05$), which is the passage at which BM182 showed significant functional improvement (Figure 2.3.3D).

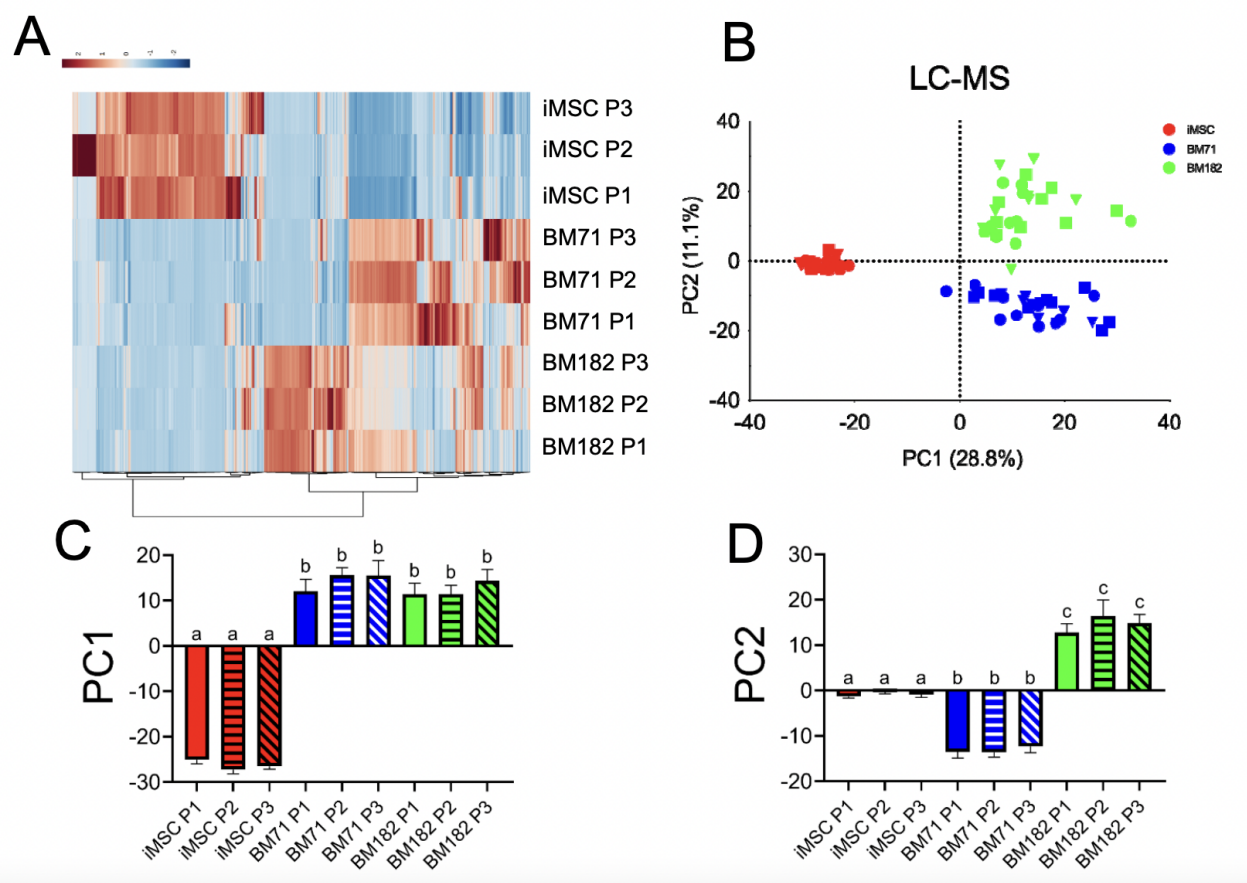


Figure 2.3.4: MSC metabolic profile from MS analysis. (A) Heat map of 1386 metabolites using Euclidian distance measure and ward clustering method. (B) Unsupervised PCA of all metabolites for each cell-line at three passages (P1=circle, P2=triangle, P3=square). (C) One-way ANOVA comparison of average PC1 value. (D) One-way ANOVA comparison of average PC2 value. Groups not sharing the same lettering in C and D represents a statistical difference ($p<0.05$).

2.3.4 UPLC-MS-based identification of MSC metabolic signatures

Additional cell pellet samples collected at the time of cell harvest for each cell-line/passage were analyzed using UPLC-MS. As with NMR metabolic profiling results, the heat map for the 1368 metabolite features showed clear differences in the metabolic profile between BM and iMSC cell-

lines, and between different passages within a given cell-line (Figure 2.3.4A). Similar to what was observed in NMR, PCA of the 1368 UPLC-MS metabolites displayed a clear separation of all cell-lines (Figure 2.3.4B). iMSCs separated from both BM71 and BM182 ($p<0.05$) along the PC1 axis, with no differences between BM71 and BM182 (Figure 2.3.4C). All cell-lines were significantly different ($p<0.05$) from one another along PC2 (Figure 2.3.4D). There were no metabolomic profile differences between passages within a cell-line.

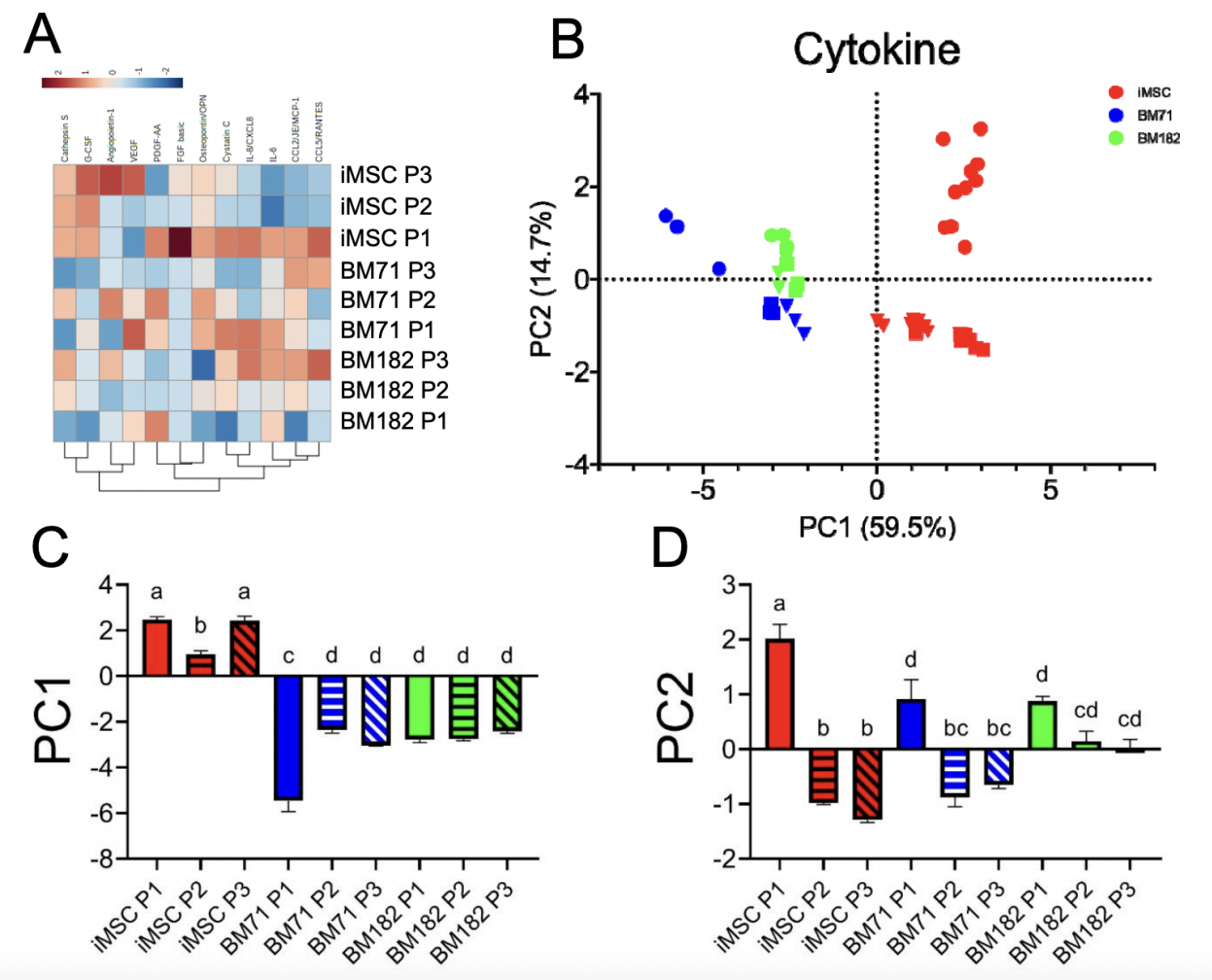


Figure 2.3.5: MSC cytokine profile. (A) Heat map of all 12 measured cytokines using Euclidian distance measure and ward clustering method. (B) Unsupervised PCA of all cytokines for each cell-line at three passages (P1=circle, P2=triangle, P3=square). (C) One-way ANOVA comparison of average PC1 value. (D) One-way ANOVA comparison of average PC2 value. Groups not sharing the same lettering in C and D represents a statistical difference ($p<0.05$).

2.3.5 Secreted cytokine profile of MSC conditioned media

Conditioned media collected at the time of MSC harvest for each cell-line/passage were analyzed using Multi-Analyte Magnetic Luminex assay. The heatmap revealed variable cytokine expression from all cell-lines (Figure 2.3.5A). Cytokines with similar functions clustered together, for example, CCL5 and CCL2, both chemotactic cytokines for T-cells are clustered. Similarly, VEGF and angiopoietin, angiogenic growth factors, are clustered on the heat map. Cathepsin S, a protease expressed and secreted by cells during angiogenesis clustered with these cytokines [33, 34]. IL-8, IL-6 and cystatin C clustered closely with the chemotactic cytokines (Figure 2.3.5A). To examine cytokine signature differences among the cell-lines using an unbiased approach, we performed PCA. The scores plot showed a distinct separation of cytokine profiles among cell-lines (Figure 2.3.5B). The cytokine profiles of the iMSCs grouped together at PC1 on the positive side, while the two BM cell-lines were clustered on the negative side ($p<0.05$) (Figure 2.3.5C). Differences were also observed within the iMSC line (P2 vs. P1/P3) and BM71 line (P1 vs. P2/P3)($p<0.05$) (Figure 2.3.5C). The first passage of iMSCs and BM71 were significantly different ($p<0.05$) from the later passages within their respective cell-line along PC2 (Figure 2.3.5D). BM182 showed no significant differences along PC2 (Figure 2.3.5D).

Table 2.5: NMR identified metabolites, confidence scores and p- values from linear regression model. The metabolites were assigned a confidence level ranging from 1 to 5 according to published criteria. [32]

Metabolite	Confidence Score	CD4+ Proliferation PBMC D1 p value	CD8+ Proliferation PBMC D1 p value	CD4+ Proliferation PBMC D2 p value	CD8+ Proliferation PBMC D2 p value
1,3-diaminopropane	3	0.0455	0.0554	0.3778	0.0565
1,3-propanediol	3	0.7056	0.7567	0.7913	0.5208
alanine	4	0.0304	0.0389	0.3416	0.0535
asparagine	4	0.0021	0.0026	0.0963	0.029
aspartic acid	4	0.4647	0.5197	0.9916	0.3624
beta-alanine	4	0.0243	0.0293	0.2331	0.0225
betaine	4	0.5911	0.616	0.7034	0.6191
glucose	4	0.5637	0.6228	0.8347	0.4985
threitol	3	0.044	0.0572	0.3073	0.0579
glutamate	4	0.039	0.0496	0.3276	0.0542
glutamine	4	0.0016	0.0024	0.0612	0.0019
glycine	4	0.0205	0.023	0.1952	0.0152
hypotaurine	4	0.0162	0.0177	0.1662	0.0106
hypoxanthine	4	0.4121	0.3892	0.2808	0.6632
isoleucine	4	0.1163	0.1265	0.5603	0.2372
L-pyroglutamic acid	3	0.156	0.1464	0.1692	0.3623
lactate	4	0.048	0.0597	0.346	0.0472

Table 2.5 continued from previous page

leucine	4	0.0005	0.0005	0.0238	0.001
myo-inositol	4	0.0028	0.0027	0.0394	0.003
NN-dimethylglycine	4	0.0895	0.0929	0.1941	0.2148
phenylalanine	4	0.0043	0.0068	0.0359	0.0033
phosphocreatine	4	0.0061	0.0077	0.1935	0.0173
putrescine	3	0.0169	0.0193	0.0434	0.103
serine	4	0.0124	0.0133	0.1136	0.0066
taurine	4	0.5886	0.6509	0.8277	0.5039
threonine	4	0.0102	0.0129	0.2188	0.0264
tyrosine	4	0.1816	0.1779	0.203	0.3715
valine	4	0.1328	0.1206	0.0546	0.2482
ukNMR-1	/	0.0751	0.0736	0.0017	0.0282
ukNMR-2	/	0.0008	0.002	0.003	0.0002
ukNMR-3	/	0.6171	0.5474	0.3233	0.8212
ukNMR-4	/	0.0389	0.0484	0.3705	0.0662
ukNMR-5	/	0.0422	0.0524	0.2953	0.0353
ukNMR-6	/	0.1612	0.1993	0.597	0.1639
ukNMR-7	/	0.2288	0.2586	0.7817	0.2074
ukNMR-8	/	0.245	0.2823	0.6771	0.1873
ukNMR-9	/	0.1499	0.1824	0.6002	0.1693
ukNMR-10	/	0.0442	0.055	0.1646	0.0148
ukNMR-11	/	0.9659	0.9898	0.2329	0.6817
ukNMR-12	/	0.1519	0.1857	0.3297	0.0739
ukNMR-13	/	0.6154	0.7216	0.9032	0.536
ukNMR-14	/	0.1316	0.1298	0.0288	0.0916

Table 2.5 continued from previous page

ukNMR-15	/	0.2037	0.2246	0.7016	0.2187
ukNMR-16	/	0.0748	0.094	0.3894	0.08
ukNMR-17	/	0.1686	0.2018	0.7266	0.2693
ukNMR-18	/	0.5793	0.5728	0.831	0.4587
ukNMR-19	/	0.0916	0.1357	0.229	0.2903
ukNMR-20	/	0.9493	0.9817	0.8399	0.7014
ukNMR-21	/	0.0133	0.0159	0.099	0.0076
ukNMR-22	/	0.0219	0.0259	0.1993	0.0209
ukNMR-23	/	0.0699	0.078	0.1697	0.0222
ukNMR-24	/	0.1738	0.2047	0.7233	0.2166
ukNMR-25	/	0.035	0.0434	0.358	0.0482
ukNMR-26	/	0.0083	0.0106	0.0621	0.0019
ukNMR-27	/	0.0821	0.1035	0.4364	0.0831
ukNMR-28	/	0.0397	0.0516	0.2819	0.0379
ukNMR-29	/	0.0638	0.0706	0.1925	0.0244

2.3.6 Regression of identified metabolites to determine markers for predictive function

Metabolites identified from NMR and MS were regressed with $CD4^+$ and $CD8^+$ T cell proliferation (Figure 2.3.6). Metabolites with strong positive or negative correlations (R^2>0.50) were then used in order to reduce the data set and ordered based on the strongest correlation. Regression with PBMC Donor 1 revealed strong correlations with multiple NMR metabolites. Myo-inositol had the highest R^2 values for both $CD4^+$ (R^2=0.84) and $CD8^+$ (R^2=0.84) T cell proliferation as followed by unknown NMR metabolite 2 (ukNMR-2) (R^2=0.82 and R^2=0.77, respectively)

(Figure 2.3.6A-C). To confirm these results, NMR metabolites were then regressed with PBMC Donor 2. Myo-inositol and ukNMR-2 correlated (R^2=0.54 and R^2=0.74, respectively) with $CD4^+$ proliferation (Figure 2.5.5). Regression with $CD8^+$ proliferation showed very similar results to PBMC Donor 1 with myo-inositol and ukNMR-2 having strong correlations (R^2=0.81 and R^2=0.88, respectively). Linear Regression with UPLC-MS metabolites also revealed multiple metabolites that correlated with function. From PBMC donor 1, the top two correlated metabolites for $CD4^+$ and $CD8^+$ proliferation were unknown MS metabolite 1349 (ukM-1349) (R^2=0.88 and R^2=0.85, respectively) and phosphatidylcholine, PC(O-38:4) (R^2=0.81 and R^2=0.88, respectively) (Figure 2.3.6E-H). ukM-1349 and PC(O-38:4) both correlated with PBMC donor 2 $CD4^+$ (R^2=0.64 and R^2=0.55, respectively) and $CD8^+$ (R^2=0.87 and R^2=0.83, respectively) T cell proliferation as well (Figure 2.5.5 and Table 2.6).

Table 2.6: NMR and MS important identified metabolites with R^2 values from the linear regression model for both $CD4^+$ and $CD8^+$ proliferation of two PBMC donors, and indication of whether metabolite intensity increases or decreases with higher potency.

Metabolite	$CD4^+$ Proliferation PBMC D1 R^2 value	$CD8^+$ Proliferation PBMC D1 R^2 value	$CD4^+$ Proliferation PBMC D2 R^2 value	$CD8^+$ Proliferation PBMC D2 R^2 value	Increasing (+) or Decreasing (-) levels with higher potency
Phosphocreatine	0.71	0.67	0.49	0.73	+
Asparagine	0.76	0.75	0.35	0.52	+
NN-dimethylglycine	0.74	0.75	0.48	0.74	-
Myo-inositol	0.84	0.84	0.54	0.81	-
Serine	0.58	0.57	0.46	0.33	+
PC(O-36:4)	0.84	0.81	0.52	0.84	-
PC(O-32:0)	0.85	0.81	0.55	0.85	-
PC(O-38:4)	0.88	0.84	0.55	0.83	-
PC(O-34:0)	0.87	0.83	0.58	0.84	-

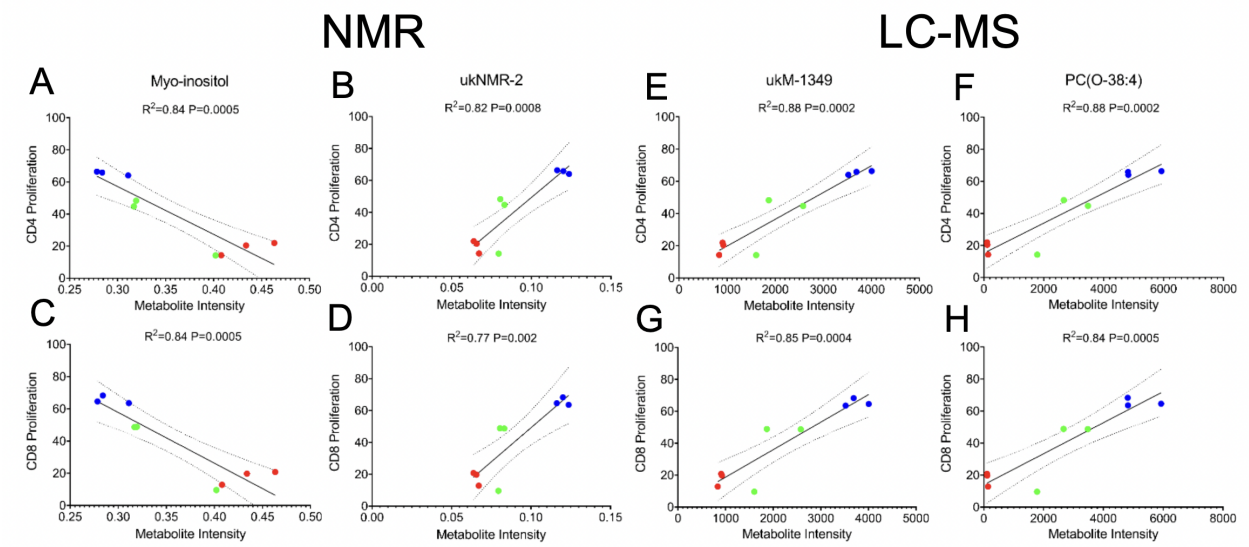


Figure 2.3.6: Linear regression analysis of the top correlated metabolites. (A,C) Myo-inositol and (B,D) ukNMR-2 had the highest R^2 values for $CD4^+$ and $CD8^+$ T cell proliferation from NMR metabolites. (E,G) ukM-1349 and (F,H) PC(O-38:4) had the highest R^2 values for $CD4^+$ and $CD8^+$ T cell proliferation from MS metabolites. The top 20 metabolites with the highest R^2 values were then chosen for further modeling to discover predictive markers. Red=iMSCs, Blue=BM71, Green=BM182

2.3.7 PLSR modeling of predictive markers enables prediction of MSC functional capacity

A composite functional score was created in order to assess each cell-line's overall functional capacity due to varying levels of T cell activation from different PBMC donors (Figure 2.3.7A). All five functional outputs were plotted using PCA, and PC1 was then used as the composite functional score which accounted for 92.1% of the variance (Figure 2.3.7B, C). The top 20 correlated metabolites from linear regression analyses were then used to train the PLSR model. Features with VIP scores greater than 1 were selected: 6 NMR metabolites (Figure 2.3.7D), 7 MS metabolites

(Figure 2.3.7E), 4 cytokines (Figure 2.3.7F), and 10 total predictive markers when combining all data sets (Figure 2.3.7G) from each model respectively. PLSR was then retrained on each omics data set. NMR had the greatest R^2 value (R^2=0.86) followed by MS (R^2=0.83) and cytokines (R^2=0.70) (Figure 2.3.7H-J). Combining these data sets showed high predictability (R^2=0.88), but not higher than NMR or MS alone. Furthermore, we used a leave-one-out cross validation approach to perform some additional level of validation. In this approach, one of the samples is removed from the dataset and used as a "known unknown". A new model is then built and used to predict the removed sample. This process is repeated for all samples and then the average error computed.

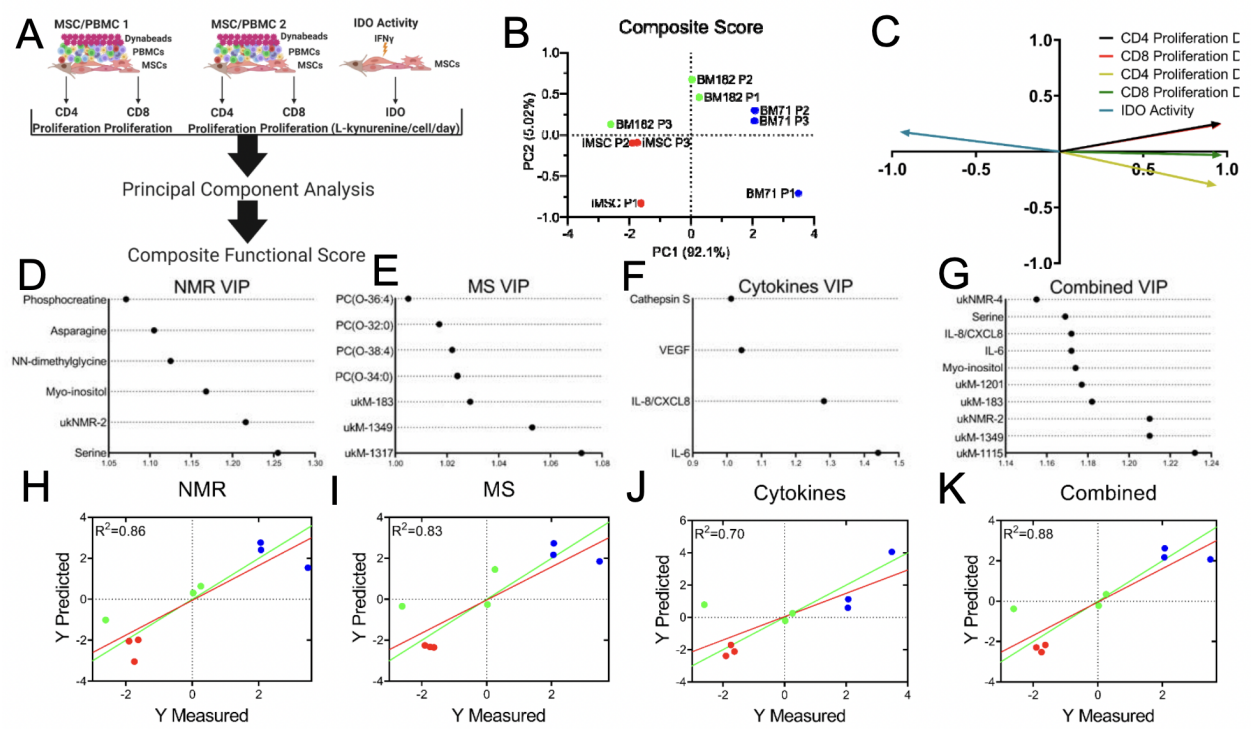


Figure 2.3.7: PLSR modeling of important features elucidates predictive markers. (A-C) $CD4^+$ and $CD8^+$ T cell proliferation from two donors and IDO activity (5 total functional outputs) were analyzed using PCA. PC1 accounted for 92.1% of the variance and was then taken and used as a composite functional score for further predictive modeling. VIP scores were assigned to the top (D) NMR, (E) MS, (F) cytokines, and (G) a combination of all features based on PLSR modeling (H-K). The VIP molecules chosen are considered as predictive markers.

2.4 Discussion

MSC therapies are currently being investigated due to their immunomodulatory properties, but translation to the clinic has been a challenge because of the heterogeneity of MSCs and a lack of CQAs that predict potency [4, 35].> Combinatorial, matrix-based approaches may help overcome these challenges by encompassing more than one functional metric since MSCs modulate the

immune system using different mechanisms [10]. In this study, we combined comprehensive metabolomics profiling with non-destructive cytokine profiling to determine several predictive markers for assessing MSC immunosuppressive capacity. Using a composite functional score, these markers were strongly correlated with MSC functional capacity. The approach outlined in this study takes large data sets that sample distinct cellular functions and identifies important features for MSC potency prediction.

Cellular metabolism is a key regulator of MSC fate and immunomodulatory potency [15]. Using an NMR based metabolomics approach, dynamic changes between cell lines and passages were observed . BM182 at P3 was functionally better than P1 and P2. Several intracellular metabolites were positively and negatively correlated with MSC functional capacity. The amino acids serine and asparagine were significant based on a PLSR model with VIP selection. Serine levels have been shown to control the self-renewal of epidermal stem cells (EpdSCs) as decreased extracellular serine activates de novo serine synthesis and promotes epidermal differentiation of EpdSCs. It has also been shown that blocking serine synthesis facilitates malignant progression [36]. Asparagine has been shown to regulate stem cell proliferation [37]. Additionally, depletion of asparagine in cell culture medium results in diminished cell growth [38]. NN-dimethylglycine, a derivative of the amino acid glycine, is also positively correlated with MSC function. It is also a byproduct of the metabolism of choline. To our knowledge, there are no studies that demonstrate a relationship between NN-dimethylglycine and stem cell biology and thus further investigation is warranted. Phosphocreatine has been shown to reduce reactive oxygen species and protect against diabetes-induced kidney injury [39, 40]. Myo-inositol is an important growth-promoting factor of mammalian cells, and possibly acts as an osmolyte [41, 42]. Myo-inositol constitutes a component of membrane phospholipids and mediates osmoregulation [43]. All the identified metabolites play essential roles in cell growth and facilitate cell proliferation although their role in MSC immunomodulatory capacity needs to be further investigated.

MS-based metabolic profiling revealed 7 metabolites that were associated with MSC immunomodulatory potency. Four of these were annotated as phosphatidylcholines (PCs), including PC(O-36:4), PC(O-38:4), and PC(O-32:0), and PC(O-34:0). PC constitutes a major portion of the cell membrane and play an important role in cellular reprogramming and signaling [44]. The intermediate of PC synthesis or hydrolysis, lysophosphatidylcholines (LPCs), have been reported as markers for discriminating different MSC sources and could be related to differences in MSC differentiation capacity and immunomodulatory properties [26]. The PCs in our study were negatively correlated with MSC immunomodulatory capacity. Lower levels of PCs were detected in iMSCs at all passages and BM 182 at passage 3, which were the MSC groups that demonstrated higher immunosuppressive capacity. Furthermore, we discovered that the 4 PCs group into 2 pairs (PC(O-38:4)/PC(O-36:4) and PC(O-34:0)/ PC(O-32:0)) sharing the same unsaturation degree and a difference of 2 carbons in their fatty acid chain composition, suggesting an underlying, yet unexplained, connection of those PC in their biosynthesis pathway. Increases in unsaturation levels of polyunsaturated PCs may alter membrane fluidity and hence can contribute to changes in MSC morphology, which has also been shown to predict MSC function [45–47]. However, to date, the biological role of PCs in stem cell metabolism at the fatty acid chain level is still poorly understood.

MSCs have also been shown to have beneficial effects by secreting proteins to modulate cell behavior in regenerative medicine and health applications [48–50]. For this study, cytokines that span multiple MSC functions of angiogenesis, tissue repair, and recruitment of immune cells were quantified [51]. The amount of each secreted cytokine/growth factor was investigated in a non-destructive manner to enable monitoring of temporal changes due to cell secretion, media changes during expansion and passaging, as well as cell uptake through autocrine or paracrine signaling. Cytokine profiling in MSCs has been described in the literature as a metric of MSC functionality, with certain secreted cytokines upregulated in MSCs, and other cytokines unchanged or reduced, and secreted in multiple cell-lines and donors [10, 52]. By investigating levels of cytokines in MSC culture media in conjunction with metabolites and T cell suppression, we have developed an assay

matrix for predicting MSC potency for immunomodulation. Besides serving as potential effectors of MSC immunomodulatory function, the cytokines profiled in this study have also been shown to directly impact MSC behavior [53, 54]. A potential function is priming of MSCs, where MSCs are conditioned with specific cytokines to increase their immunomodulatory properties [10]. Priming may be efficacious in instances where there is low MSC survival potential ex vivo or differences in sources and donor reduces effectiveness of MSC for use in regenerative and immunomodulatory applications. MSC priming using our in-process cytokines and culture conditions can be used to develop substrates or engineered tissue for regenerative medicine [55]. Also, non-destructive monitoring of secreted factors in spent media could potentially serve as predictive markers for large scale MSC manufacturing (in bioreactors, for example). MSC-secreted cytokines can regulate MSC function both in an autocrine and paracrine manner. For example, the binding of secreted IL-8 to its receptor CXCR1 or CXCR2 can activate intracellular PI3K, MAPK, Akt phosphorylation and initiate functions of cell survival, angiogenesis and cell migration [56]. In contrast, targeted ablation of secreted cytokines such as IL-6, either by gene silencing or inhibition, led to reduced MSC proliferation and reduced capacity of MSCs to suppress T cell proliferation [57]. Ultimately, studying this crosstalk between MSCs and secreted cytokine may be a relevant aspect of MSC expansion for manufacturing.

It is well documented that MSC functional heterogeneity is derived from differences in donor/tissue source, MSC doubling level, and manufacturing conditions [35, 58]. Being able to understand and predict these differences is a challenge that needs to be addressed in order to advance MSC therapies. Knowing what to measure can help screen for high potency MSCs and assess when these MSCs begin to lose potency due to senescence [46, 59, 60]. Metabolomics and cytokine profiling elucidated several correlated features that, when used alone or in combination, were indicative of MSC potency and were able to predict an increase in potency in the BM182 line. IL-6 and IL-8 are both inflammatory cytokines that recruit immune cells such as T cells, neutrophils, and macrophages. They have also been shown to inhibit T cell apoptosis and regulatory T cell

differentiation [61–63]. Higher levels of these cytokines were secreted by less potent lines (BM71). Although the role of PCs in MSCs on immunomodulatory capacity is largely unknown, studies have shown that oxidized phospholipids, such as PCs, play a role in preventing the activation of T cells and dendritic cells as well as mediate apoptosis [64–67]. Apoptosis of MSCs by cytotoxic T cells has been shown to play an important role in their immunomodulatory capacity [68]. Myo-inositol and serine's role in MSC immune suppression has not been investigated to this point, but soluble myo-inositol has been shown to be effective in treating autoimmune diseases such as thyroiditis and hypothyroidism [69]. Further investigation into the pathways involving these metabolites will improve our understanding of how MSCs modulate the immune system.

Effective prediction of MSC potency is an important factor in manufacturing high quality therapies. Predicting MSC functional capacity has been a major challenge because MSCs can exert their immunomodulatory effects through a number of different mechanisms and immune cells. Therefore, using a combination of cell metabolites and secreted cytokines can help better predict MSC potency and set specific markers to aid in process by design cell manufacturing systems [2, 4, 10, 70]. Our study used a combination of metabolites and cytokines in order to better predict MSC potency. These are easy to target and measure for manufacturing and understanding how these metabolites affect cell function can help refine and improve the manufacturing process. This approach of predictive marker discovery and understanding can also be translated into other cell therapies such as chimeric antigen receptor T-cell (CAR-T), iPSCs, neural stem cells, and even MSC-derived extracellular vesicles. Moving forward, interrogating pathways that involve these metabolites will be important for assay development and better understanding the relationship of MSC metabolism with immunomodulation. Non-destructive, in-process monitoring of MSC metabolism using conditioned medium will also enable on-demand, precise control of MSC manufacturing. Finally, this discovery platform can be used to establish MSC predictive markers for other therapeutic applications involving differentiation (e.g. osteo-, adipo-, and chondrogenesis), tissue engineering, cancer treatment, and angiogenesis.

2.5 Supplementary materials

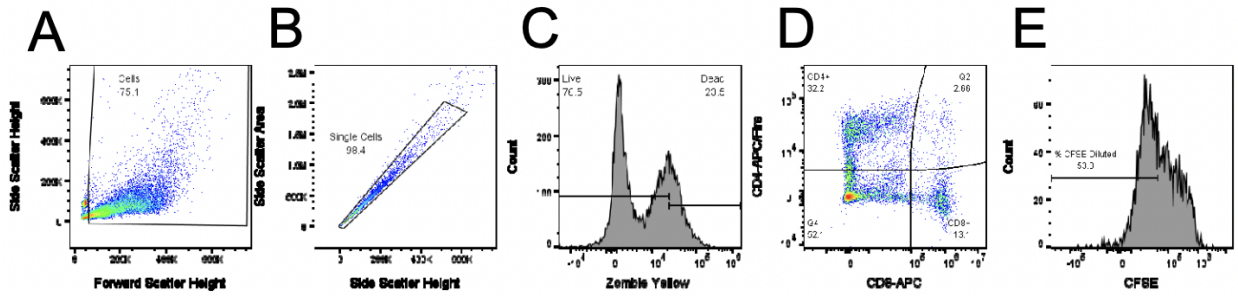


Figure 2.5.1: FlowJo gating strategy. Cellular debris and Dynabeads (A) were first removed followed by cell doublets (B). Using FMO controls, live cells (C) were then gated and used to determine $CD4^+$ and $CD8^+$ T cell populations (D). Negative control PBMCs (No stimulation, no MSCs) were then used for the CFSE dilution gate (E).

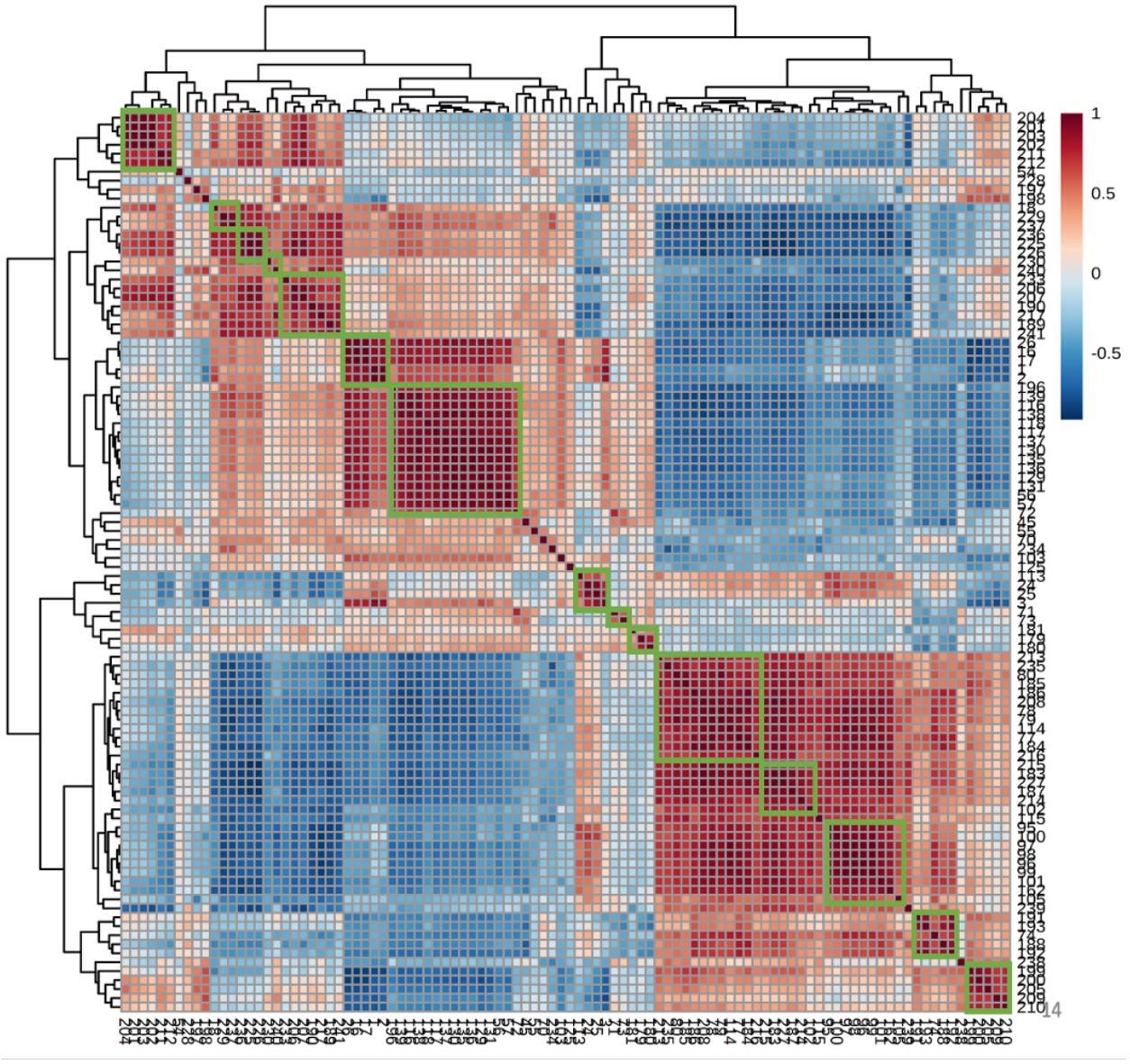


Figure 2.5.2: The correlation map of 100 unknown NMR features. We assume that correlation coefficient values greater or equal to 0.8 indicating those features are from same metabolite and grouped together as Indicated by green box in the figure. The features with correlation coefficient value less than 0.8, we considered them as individual metabolites. Total 100 unknown features were grouped to 29 unknown metabolites. The color bar indicating correlation coefficient values calculated using Z score.

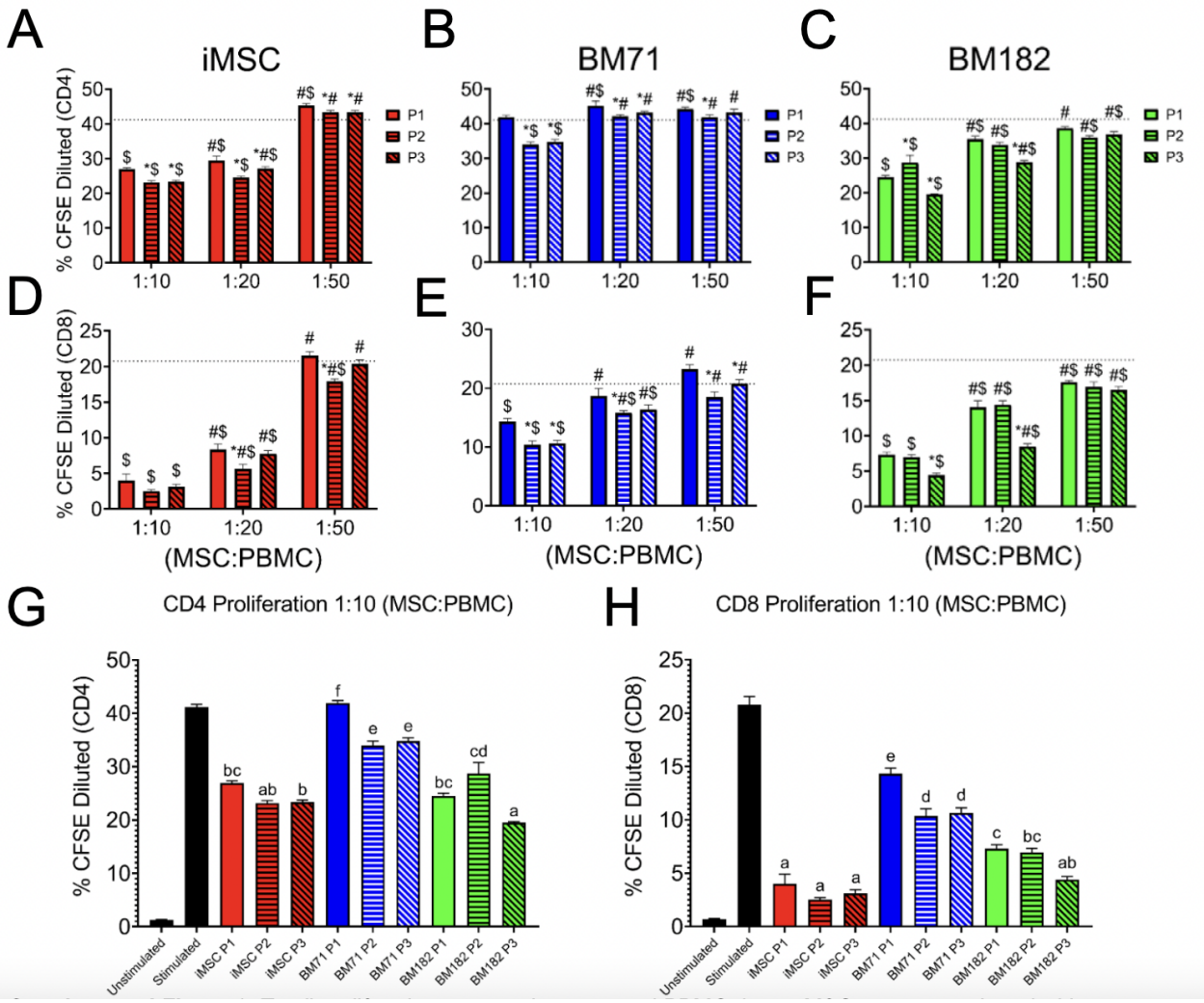


Figure 2.5.3: T cell proliferation assay using a second PBMC donor. MSCs were co-cultured with stimulated PBMCs at three different MSC:PBMC ratios (1:10, 1:20, 1:50). (A-C) $CD4^+$ and (D-F) $CD8^+$ T cell proliferation was assessed at each passage and ratio by % CFSE dilution. A 2-way ANOVA was used in order to determine if there was a significant difference from P1 within a ratio (*), a significant difference from the 1:10 ratio within a passage (#), or a significant difference from the stimulated control ($) with $p<0.05$ considered significant. (G) $CD4^+$ and (H) $CD8^+$ T cell proliferation comparing all cell-lines and passages at the 1:10 ratio. Groups not sharing the same lettering in G and H represents a statistical difference ($p<0.05$).

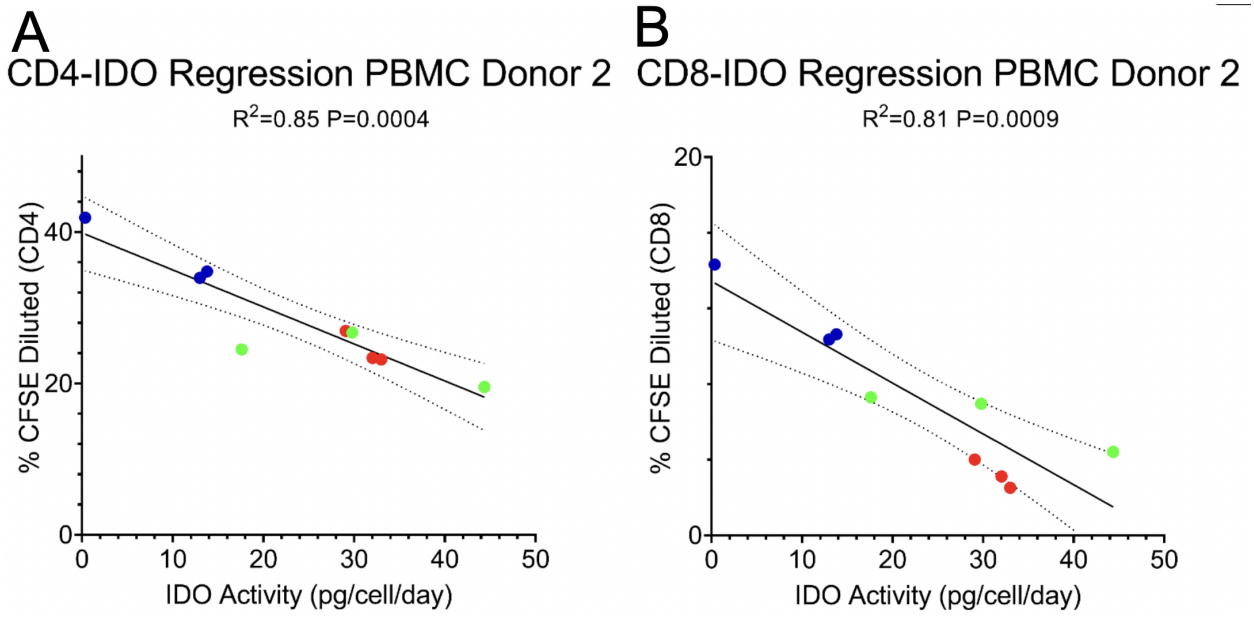


Figure 2.5.4: Regression of IDO activity and T cell proliferation with a second PBMC donor. Linear regression of the relationship between IDO activity and (A) $CD4^+$ and (B) $CD8^+$ T cell proliferation.

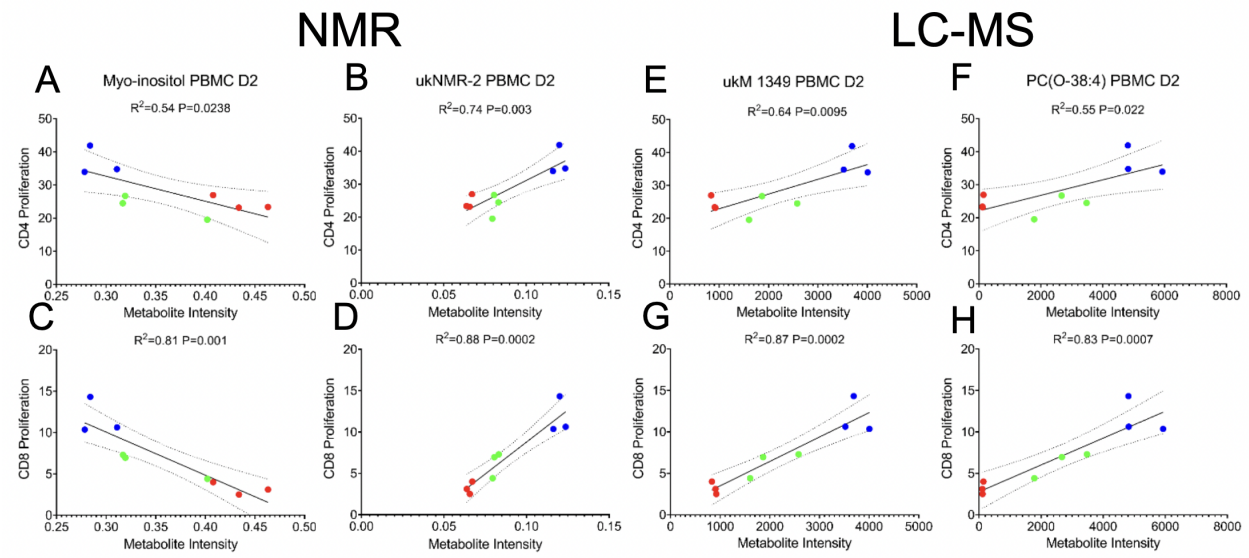


Figure 2.5.5: Linear regression analysis of the top correlated metabolites with PBMC donor 2. (A,C) Myo-inositol, (B,D) ukNMR-2, (E,G) ukM-1349, and (F,H) PC(O-38:4) all correlated with $CD4^+$ and $CD8^+$ T cell proliferation for both PBMC donors.

BIBLIOGRAPHY

[1] Marta E Castro-Manrreza and Juan J Montesinos. "Immunoregulation by mesenchymal stem cells: biological aspects and clinical applications". In: *Journal of immunology research* 2015 (2015).

[2] Jacques Galipeau et al. "International Society for Cellular Therapy perspective on immune functional assays for mesenchymal stromal cells as potency release criterion for advanced phase clinical trials". In: *Cytotherapy* 18.2 (2016), pp. 151–159.

[3] Timothy R Olsen et al. "Peak MSC—are we there yet?" In: *Frontiers in medicine* 5 (2018), p. 178.

[4] Kevin P Robb et al. "Mesenchymal stromal cell therapy: progress in manufacturing and assessments of potency". In: *Cytotherapy* 21.3 (2019), pp. 289–306.

[5] Charlotte De Wolf, Marja Van De Bovenkamp, and Marcel Hoefnagel. "Regulatory perspective on in vitro potency assays for human mesenchymal stromal cells used in immunotherapy". In: *Cytotherapy* 19.7 (2017), pp. 784–797.

[6] Qingdong Guan et al. "Inducible indoleamine 2, 3-dioxygenase 1 and programmed death ligand 1 expression as the potency marker for mesenchymal stromal cells". In: *Cytotherapy* 20.5 (2018), pp. 639–649.

[7] Maria Ester Bernardo and Willem E Fibbe. "Mesenchymal stromal cells: sensors and switchers of inflammation". In: *Cell stem cell* 13.4 (2013), pp. 392–402.

[8] Yufang Shi et al. "How mesenchymal stem cells interact with tissue immune responses". In: *Trends in immunology* 33.3 (2012), pp. 136–143.

[9] Armand Keating. "Mesenchymal stromal cells: new directions". In: *Cell stem cell* 10.6 (2012), pp. 709–716.

[10] Raghavan Chinnadurai et al. "Potency analysis of mesenchymal stromal cells using a combinatorial assay matrix approach". In: *Cell reports* 22.9 (2018), pp. 2504–2517.

[11] Takahiro Kamota et al. "Ischemic pre-conditioning enhances the mobilization and recruitment of bone marrow stem cells to protect against ischemia/reperfusion injury in the late phase". In: *Journal of the American College of Cardiology* 53.19 (2009), pp. 1814–1822.

[12] Muhammad R Afzal et al. “Preconditioning promotes survival and angiomyogenic potential of mesenchymal stem cells in the infarcted heart via nf-κ b signaling”. In: *Antioxidants & redox signaling* 12.6 (2010), pp. 693–702.

[13] Rui-Zhen Shi et al. “Angiotensin II induces vascular endothelial growth factor synthesis in mesenchymal stem cells”. In: *Experimental cell research* 315.1 (2009), pp. 10–15.

[14] Philipp Mattar and Karen Bieback. “Comparing the immunomodulatory properties of bone marrow, adipose tissue, and birth-associated tissue mesenchymal stromal cells”. In: *Frontiers in immunology* 6 (2015), p. 560.

[15] Yijun Liu and Teng Ma. “Metabolic regulation of mesenchymal stem cell in expansion and therapeutic application”. In: *Biotechnology progress* 31.2 (2015), pp. 468–481.

[16] Yijun Liu et al. “Commitment to aerobic glycolysis sustains immunosuppression of human mesenchymal stem cells”. In: *Stem cells translational medicine* 8.1 (2019), pp. 93–106.

[17] Xuegang Yuan, Timothy M Logan, and Teng Ma. “Metabolism in human mesenchymal stromal cells: a missing link between hMSC biomanufacturing and therapy?” In: *Frontiers in immunology* 10 (2019), p. 977.

[18] Kambiz Gilany et al. “Lipidomics of adipogenic differentiation of mesenchymal stem cells”. In: *Genomics, Proteomics, and Metabolomics*. Springer, 2019, pp. 123–140.

[19] Warwick B Dunn, Nigel JC Bailey, and Helen E Johnson. “Measuring the metabolome: current analytical technologies”. In: *Analyst* 130.5 (2005), pp. 606–625.

[20] Georgios A Theodoridis et al. “Liquid chromatography–mass spectrometry based global metabolite profiling: a review”. In: *Analytica chimica acta* 711 (2012), pp. 7–16.

[21] A Jiye et al. “Extraction and GC/MS analysis of the human blood plasma metabolome”. In: *Analytical chemistry* 77.24 (2005), pp. 8086–8094.

[22] Panagiotis A Vorkas et al. “Untargeted UPLC-MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease”. In: *Analytical chemistry* 87.8 (2015), pp. 4184–4193.

[23] Elizabeth J Want et al. “Global metabolic profiling procedures for urine using UPLC–MS”. In: *Nature protocols* 5.6 (2010), pp. 1005–1018.

[24] Fan Fei, Dawn ME Bowdish, and Brian E McCarry. “Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS”. In: *Analytical and bioanalytical chemistry* 406.15 (2014), pp. 3723–3733.

[25] Nur Ashikin Abdul-Hamid et al. “NMR metabolomics for evaluating passage number and harvesting effects on mammalian cell metabolome”. In: *Analytical biochemistry* 576 (2019), pp. 20–32.

[26] Seul Ji Lee et al. “Comparative study on metabolite level in tissue-specific human mesenchymal stem cells by an ultra-performance liquid chromatography quadrupole time of flight mass spectrometry”. In: *Analytica chimica acta* 1024 (2018), pp. 112–122.

[27] Anne E Carpenter et al. “CellProfiler: image analysis software for identifying and quantifying cell phenotypes”. In: *Genome biology* 7.10 (2006), pp. 1–11.

[28] Frank Delaglio et al. “NMRPipe: a multidimensional spectral processing system based on UNIX pipes”. In: *Journal of biomolecular NMR* 6.3 (1995), pp. 277–293.

[29] Keshav Kumar, Ralf Schweiggert, and Claus-Dieter Patz. “Introducing a novel procedure for peak alignment in one-dimensional 1 H-NMR spectroscopy: a prerequisite for chemometric analyses of wine samples”. In: *Analytical Methods* 12.28 (2020), pp. 3626–3636.

[30] Frank Dieterle et al. “Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics”. In: *Analytical chemistry* 78.13 (2006), pp. 4281–4290.

[31] Kerem Bingol et al. “Comprehensive metabolite identification strategy using multiple two-dimensional NMR spectra of a complex mixture implemented in the COLMARm web server”. In: *Analytical chemistry* 88.24 (2016), pp. 12411–12418.

[32] Jacquelyn M Walejko et al. “Global metabolomics of the placenta reveals distinct metabolic profiles between maternal and fetal placental tissues following delivery in non-labored women”. In: *Metabolites* 8.1 (2018), p. 10.

[33] G-P Shi et al. “Deficiency of the cysteine protease cathepsin S impairs microvessel growth”. In: *Circulation research* 92.5 (2003), pp. 493–500.

[34] Claire Ward et al. “Antibody targeting of cathepsin S inhibits angiogenesis and synergistically enhances anti-VEGF”. In: *PloS one* 5.9 (2010), e12543.

[35] Alison Wilson et al. “Multiplicity of mesenchymal stromal cells: finding the right route to therapy”. In: *Frontiers in immunology* 10 (2019), p. 1112.

[36] Sanjeethan C Baksh et al. “Extracellular serine controls epidermal stem cell fate and tumour initiation”. In: *Nature cell biology* 22.7 (2020), pp. 779–790.

[37] Ming Luo, Michael Brooks, and Max S Wicha. “Asparagine and glutamine: co-conspirators fueling metastasis”. In: *Cell metabolism* 27.5 (2018), pp. 947–949.

[38] Elizabeth Emma Palmer et al. “Asparagine synthetase deficiency causes reduced proliferation of cells under conditions of limited asparagine”. In: *Molecular genetics and metabolism* 116.3 (2015), pp. 178–186.

[39] Zheng Jing et al. “Phosphocreatine Promotes Osteoblastic Activities in H2O2-Induced MC3T3-E1 Cells by Regulating SIRT1/FOXO1/PGC-1α Signaling Pathway”. In: *Current Pharmaceutical Biotechnology* (2021).

[40] Abdullah Shopit et al. “Protection of diabetes-induced kidney injury by phosphocreatine via the regulation of ERK/Nrf2/HO-1 signaling pathway”. In: *Life sciences* 242 (2020), p. 117248.

[41] Dhani Raj Chhetri. “Myo-Inositol and its derivatives: Their emerging role in the treatment of human diseases”. In: *Frontiers in pharmacology* 10 (2019), p. 1172.

[42] Jacobus FA Jansen et al. “Stem cell profiling by nuclear magnetic resonance spectroscopy”. In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 56.3 (2006), pp. 666–670.

[43] A Lahiri Majumder and BB Biswas. *Biology of inositols and phosphoinositides*. Vol. 39. Springer Science & Business Media, 2006.

[44] Neale D Ridgway. “The role of phosphatidylcholine and choline metabolites to cell proliferation and survival”. In: *Critical reviews in biochemistry and molecular biology* 48.1 (2013), pp. 20–38.

[45] John K Meissen et al. “Induced pluripotent stem cells show metabolomic differences to embryonic stem cells in polyunsaturated phosphatidylcholines and primary metabolism”. In: (2012).

[46] Matthew W Klinker et al. “Morphological features of IFN-γ–stimulated mesenchymal stromal cells predict overall immunosuppressive capacity”. In: *Proceedings of the National Academy of Sciences* 114.13 (2017), E2598–E2607.

[47] Jihee Sohn et al. “Influence of cholesterol/caveolin-1/caveolae homeostasis on membrane properties and substrate adhesion characteristics of adult human mesenchymal stem cells”. In: *Stem cell research & therapy* 9.1 (2018), pp. 1–15.

[48] Mark F Pittenger et al. “Mesenchymal stem cell perspective: cell biology to clinical progress”. In: *NPJ Regenerative medicine* 4.1 (2019), pp. 1–15.

[49] Aleksandra Musiał-Wysocka, Marta Kot, and Marcin Majka. “The pros and cons of mesenchymal stem cell-based therapies”. In: *Cell transplantation* 28.7 (2019), pp. 801–812.

[50] Maria Beatriz Herrera et al. “Mesenchymal stem cells contribute to the renal repair of acute tubular epithelial injury”. In: *International journal of molecular medicine* 14.6 (2004), pp. 1035–1041.

[51] Rebekah M Samsonraj et al. “Concise review: multifaceted characterization of human mesenchymal stem cells for use in regenerative medicine”. In: *Stem cells translational medicine* 6.12 (2017), pp. 2173–2185.

[52] Chae Woon Park et al. “Cytokine secretion profiling of human mesenchymal stem cells by antibody array”. In: *International journal of stem cells* 2.1 (2009), p. 59.

[53] DC Lacey et al. “Proinflammatory cytokines inhibit osteogenic differentiation from stem cells: implications for bone repair during inflammation”. In: *Osteoarthritis and Cartilage* 17.6 (2009), pp. 735–742.

[54] Daniëlle G Leuning et al. “The cytokine secretion profile of mesenchymal stromal cells is determined by surface structure of the microenvironment”. In: *Scientific reports* 8.1 (2018), pp. 1–9.

[55] Nádia de Cássia Noronha et al. “Priming approaches to improve the efficacy of mesenchymal stromal cell-based therapies”. In: *Stem cell research & therapy* 10.1 (2019), pp. 1–21.

[56] Aijun Yang et al. “IL-8 enhances therapeutic effects of BMSCs on bone regeneration via CXCR2-mediated PI3k/Akt signaling pathway”. In: *Cellular Physiology and Biochemistry* 48.1 (2018), pp. 361–370.

[57] Akaitz Dorronsoro et al. “Intracellular role of IL-6 in mesenchymal stromal cell immunosuppression and proliferation”. In: *Scientific reports* 10.1 (2020), pp. 1–12.

[58] Susanne Kern et al. “Comparative analysis of mesenchymal stem cells from bone marrow, umbilical cord blood, or adipose tissue”. In: *Stem cells* 24.5 (2006), pp. 1294–1301.

[59] Madeleine C Killer et al. “Immunosuppressive capacity of mesenchymal stem cells correlates with metabolic activity and can be enhanced by valproic acid”. In: *Stem cell research & therapy* 8.1 (2017), pp. 1–8.

[60] Valentina Turinetto, Emanuela Vitale, and Claudia Giachino. “Senescence in human mesenchymal stem cells: functional changes and implications in stem cell-based therapy”. In: *International journal of molecular sciences* 17.7 (2016), p. 1164.

[61] Kevin Anton, Debabrata Banerjee, and John Glod. “Macrophage-associated mesenchymal stem cells assume an activated, migratory, pro-inflammatory phenotype with increased IL-6 and CXCL10 secretion”. In: *PloS one* 7.4 (2012), e35036.

[62] Stefan Rose-John. “IL-6 trans-signaling via the soluble IL-6 receptor: importance for the pro-inflammatory activities of IL-6”. In: *International journal of biological sciences* 8.9 (2012), p. 1237.

[63] Akihisa Harada et al. “Essential involvement of interleukin-8 (IL-8) in acute inflammation”. In: *Journal of leukocyte biology* 56.5 (1994), pp. 559–564.

[64] Maria Seyerl et al. “Oxidized phospholipids induce anergy in human peripheral blood T cells”. In: *European journal of immunology* 38.3 (2008), pp. 778–787.

[65] Vlad Serbulea, Dory DeWeese, and Norbert Leitinger. “The effect of oxidized phospholipids on phenotypic polarization and function of macrophages”. In: *Free Radical Biology and Medicine* 111 (2017), pp. 156–168.

[66] Stefan Blüml et al. “Oxidized phospholipids negatively regulate dendritic cell maturation induced by TLRs and CD40”. In: *The Journal of Immunology* 175.1 (2005), pp. 501–508.

[67] Niketa Sareen et al. “Early passaging of mesenchymal stem cells does not instigate significant modifications in their immunological behavior”. In: *Stem Cell Research & Therapy* 9.1 (2018), pp. 1–11.

[68] Antonio Galleu et al. “Apoptosis in mesenchymal stromal cells induces in vivo recipient-mediated immunomodulation”. In: *Science translational medicine* 9.416 (2017).

[69] Poupak Fallahi et al. “Myo-inositol in autoimmune thyroiditis, and hypothyroidism”. In: *Reviews in Endocrine and Metabolic Disorders* 19.4 (2018), pp. 349–354.

[70] Jacques Galipeau and Luc Sensébé. “Mesenchymal stromal cells: clinical challenges and therapeutic opportunities”. In: *Cell stem cell* 22.6 (2018), pp. 824–833.

CHAPTER 3

NON-DESTRUCTIVE, DYNAMIC PROFILING REVEALS METABOLITES THAT PREDICT MESENCHYMAL STROMAL CELL IMMUNOSUPPRESSION

[1]

[1] Xunan Shen, Ty S. Maughon, Ross A. Marklein, Steven L. Stice, Arthur S. Edison. To be submitted to Cytotherapy

Authors Contributions: This manuscript will be submitted to journal Cytotherapy. I am the first author of this paper. I participated in conceptual study design with Authur S. Edison and Steven L. Stice. Ty S. Maughon cultured the cells and collected the media samples daily. I performed NMR data acquisition and data pre-processing. I also developed the pipeline for data analysis. I wrote the manuscript with the help from Ty S. Maughon, Arthur S. Edison, Ross A. Marklein and Steven L. Stice.

Abstract

Background: Mesenchymal stromal/stem cells (MSCs) have shown promising results in clinical trials for their immunomodulatory function. However, MSC therapy isn't licensed by FDA, in part because of the heterogeneity of MSCs. The lack of predictive markers also makes it difficult to both manufacture and translate MSCs into clinic. Indoleamine 2,3-Dioxygenase (IDO) assay and T-cell suppression assays correlate with MSCs function. We previously showed that cellular metabolites can be used to predict IDO assay and T cell suppression results. Although these methods are promising, they are all destructive and time-consuming and therefore cannot easily translate to a cell manufacturing setting. A non-destructive, in-process method to evaluate cell immunomodulation effects would be extremely valuable.

Methods: Culture media from the growth of three different MSC cell-lines (two bone marrow, one iPSC) were sampled daily for NMR metabolomics analysis. MSC IDO activity and modulation of T cell proliferation were used as clinically-relevant potency assays. Linear regression was used to assess the media metabolic changes over time, and partial least squares regression (PLSR) was then used to obtain predictive media markers (PMMs) based on variable importance in projection (VIP) scores. In addition, pathway analysis was performed to relate media metabolites (MMs) and cell metabolites (CMs).

Results: For the first three-day culturing time period, PLSR of culture media regressed against a composite score resulted in R^2 values of 0.77. Several amnio acids were useful PMMs including leucine, threonine, and alanine. Correlation and pathway analyses related the consumption of valine and aspartate to the release of glycine and alanine during culture.

Discussion: The work described here used PLSR models to identify PMMs that can predict MSC function. The pathway analysis revealed that consuming of leucine and secreting of alanine directly correlate with MSC immunomodulation effects.

3.1 Introduction

Mesenchymal stromal cells are multipotent adult cells. Unlike other tissue-specific stem cells, MSCs can be isolated from various tissue types. In fact, evidence suggests that MSCs may be found in any vascularized tissue throughout the whole body [1]. MSCs have been widely used in clinical trials for their ability to modulate inflammation. MSC modulation of immune cells is regulated by cytokines, chemokines, cell surface molecules, as well as metabolic pathways [2–4]. MSCs suppress T-cell proliferation, cytokine secretion and cytotoxicity [5]. Activated MSCs can suppress $CD4^+$ and $CD8^+$ T cell proliferation [6]. Although over 700 MSC anti-inflammatory clinical trials have been conducted (data from clinicaltrials), the clinical use of MSCs is still limited, mainly because of their heterogeneity [7]. MSCs exhibit heterogeneity on multiple levels, such as the tissue source, donor-to-donor variability, and at a subpopulation level [8]. In addition, the lack of predictive markers also limits their use in clinic.

Finding predictive markers to establish MSC immunomodulation effects and reveal the underlying mechanisms become essential to translate MSC from research to clinic. The minimal criteria for defining MSCs described by the International Society for Cellular Therapy (ISCT) [9] is insufficient to predict MSCs potency [10]. To date, there is no consensus on any MSCs potency assay, in part because they have many different functions and mechanisms of action. No single test or assay can fully characterize MSCs and predict their clinical efficacy. A common approach to estimate MSCs anti-inflammatory capacity is an immune suppression assay. By activating a MSC co-culture with peripheral blood mononuclear cells (PBMCs) and measuring the proliferation rate of $CD4^+$and $CD8^+$ T cells, one can demonstrate the immune modulation capacity of the MSCs [11]. A lower T-cell proliferation rate indicates higher MSCs potency. Another important assay to

measure MSC potency is an indoleamine 2,3-dioxygenase (IDO) assay. IDO catalyzes tryptophan into different metabolites, resulting in tryptophan depletion [12]. Tryptophan is essential for T-cell proliferation; thus, the depletion of tryptophan can cause the suppression of T-cell proliferation. Higher IDO levels indicate higher MSCs potency [13]. Recently, we have shown that cellular metabolites, along with cytokines can be used as a surrogate of T-cell suppression assay and IDO assays to predict MSCs function [14]. By measuring key predictive metabolic markers and cytokine markers, we can predict a functional score generated from T-cell suppression assay and IDO assay. Although these assays can demonstrate immunomodulation capacity of MSCs, they are all destructive. In addition, they are somewhat complex to carry out and require the cells at to the end of culturing period. Significant recourses are required to measure MSC potency using these methods. Characterizing MSCs uses a non-destructive, in-process MSC characterization method that can provide an early indication of cell immunomodulation capacity is needed. The early prediction of MSC immunomodulation effects would significantly save time and money, ensure high quality product, increase safety and effectiveness.

MSC culture media components and their relative abundance alters cellular processes and characteristics. MSCs cultured with a high glucose media showed a higher cell count and earlier confluency in light microscopy compared to low glucose media [14, 15]. The expressions of CD10, CD90, CD105, CD140b CD146 and STRO-1 are highly influenced by different culture media [14]. Adipose MSCs (AMSCs) cultured with fetal bovine serum (FBS) can increase the secretion of anti-inflammatory factors including Interleukin 10 (IL-10), IDO and transforming growth factor beta (TGF-β) compared to cells cultured without FBS [16]. This shows that culture media also plays an important role in determining MSC anti-inflammatory capacity. By measuring the media during MSC culture, we may be able to find new insights into optimal culture conditions.

Metabolomics uses NMR (Nuclear Magnetic Resonance) or mass spectrometry to identify and quantify small molecules and relate these to phenotype [17]. Metabolomics is very promising in characterizing MSCs. Showalter et.al, investigated the hypoxia and serum deprivation of MSCs,

and found 21 distinct metabolites that are directly associated with MSCs immunoregulation [18]. Increased essential amino acids consuming in culture media as well as increased lactate production from cells has also been found in human MSC culture [19]. Activated MSCs can increase glycolysis, and inhibition of these metabolic changes reduced the production of key immunosuppressive cytokines [20]. Thus, metabolomics is a sensitive approach to characterize MSC immunosuppression capacity.

In our previous study, we reported the relationship between MSC function with cellular metabolites and secreted cytokines [21]. We found that both metabolites and cytokines can be used to predict IDO and T-cell suppression. Additional in process measurements would enhance and compliment these earlier findings. Here, we report the NMR analysis of media from the previous study and show that changes in the media components can also be used to predict MSC function. We performed a pathway analysis to investigate the relationships between cellular and media metabolites and found that changes in some media metabolites can predict patterns of cellular metabolites. For example, the consuming of aspartate from media shows the accumulation of asparagine in cells, meanwhile, the secretion of creatine to media indicates the increase of phosphocreatine in cells. The NMR measurements of media are simple and non-destructive and thus suitable for in-process analysis for MSC culturing in biomanufacturing.

3.2 Materials and methods

3.2.1 Media collection and cell immunodulation evaluation assays

MSCs were cultured as previously described [21]. Each cell-line has 10 replicates. Two bone marrow-derived MSC lines (RoosterBio, Frederick MD) lot #0071 (F, 18-30) and #0182 (F, 26) (which the manufacturer has both research and clinical-grade lots available), and one induced pluripotent stem cell derived MSC cell-line (Cellular Dynamics International, Madison WI) (Lot #0003, also prequalified) were used and referred here as BM71, BM182, and iMSC, respectively.

iMSC were cultured for 25 days, BM71 were cultured for 21 days and BM182 were cultured for 23 days. 100 μl media was sampled daily from each sample and stored at -80°C for further analysis.

3.2.2 Culture medium sample preparation for NMR

100 μl culture media was thawed on ice and then centrifuged at 14,000 x g for 15 min at 4°C. For each sample, 54 μl of media supernatant was transferred to a new Eppendorf tube. 30 μl of remaining media in each sample from the same cell-line was pooled together to generate six 54 μl internal quality control (QC) samples. 6 μl of 10/3 mM DSS-D6 (Cambridge Isotope Laboratory) in D2O (Cambridge Isotope Laboratory) was then added into each sample. In addition, six buffer blanks (6 μl of 10/3 mM DSS-D6 in 54 μl D2O) samples were added in each cell-line for quality assurance purpose. We had 10 replicates for each cell-line and time point. A total of 262 samples including 250 experimental samples, 6 pooled samples and 6 buffer blanks were generated for the iMSC group. A total of 222 samples including 210 experimental samples, 6 pooled samples and 6 buffer blanks were generated for BM71. A total of 242 samples including 230 experimental samples, 6 pooled samples and 6 buffer blanks were generated for BM182. All samples were randomized to reduce technical variance.

3.2.3 NMR data acquisition and pre-processing

Sixty μl of each sample was transferred to racks of 96 1.7-mm NMR tubes for data acquisition using a SamplePro tube robotic system (Bruker Biospin, Billerica, MA, USA). Samples were run on a Bruker NEO 800 MHz NMR spectrometer equipped with a 1.7-mm cryoprobe and Bruker SampleJet cooled to 5.6°C. One dimensional nuclear Overhauser enhancement spectroscopy with water suppression (noesypr1d) and J-resolve (jresgpprqf) spectroscopy were collected on all samples. Two-dimensional ^{1}H-^{1}3C Heteronuclear Single Quantum Coherence (HSQC) and ^{1}H-^{1}3C HSQC-Total Correlated Spectroscopy (HSQC-TOCSY) spectra on internal pooled samples were collected for metabolite identification.

1D and 2D spectra were processed using NMRPipe [22] and in-house MATLAB metabolomics toolbox (Github).

The NMR spectra were Fourier transformed, phased, baseline corrected and referenced using NMRpipe. The spectra were imported into MATLAB for further analysis. The water region from 4.69 ppm to 4.88 ppm was removed in each spectrum and replaced with value 0. The buffer blank and pooled samples were used to ensure data collection quality and removed in downstream analysis. 1D spectra were aligned using correlation optimized wrapping (COW) algorithm [23] and normalized using the probabilistic quotient normalization (PQN) algorithm [24].

3.2.4 Metabolite annotation and peak binning

The 2D NMR spectra were matched to a metabolite database using the COLMARm webserver (CCIC). The metabolites were assigned a confidence level ranging from 1 to 5 according to published criteria [25].

The non-overlapped peaks from 1D spectra were manually binned by taking the boundaries of individual peaks and calculating the area under the curve each for peak. After metabolite identification, peaks from same metabolite were binned together, and quantified according to the number of protons in each peak.

3.2.5 Statistical analysis

Variance-based feature selection strategy and Pearson correlation heatmaps were done using MATLAB R2020b. Simple linear regression was conducted using Python sklearn package. The parameters were used as following: (fit_intercept=True, normalize=False, copy_X=True, n_jobs=None, positive=False). The partial least square regression (PLSR) model was conducted using MATLAB R2020b PLS_Toolbox_89. The PLSR was used to regress metabolite changes against composite function score generated in our previous paper [21]. Leave one out cross validation was used in training process.

3.3 Results

3.3.1 Study design and pre-processing strategy

The culture time for each MSC depended on the cell confulence: iMSC BM71, and BM182 were cultured for 25, 21, and 23 days, respectively. Ten replicates were used for each cell line. Media from each sample was collected daily. Day 0 media was collected as a baseline control. The culture process required a media switch every 2-3 days, where fresh media replaced the spent media to replenish nutrients (Figure 3.3.1). NMR data were collected on 726 samples (including buffer blank and pooled controls), and a sample NMR spectrum is shown in Figure 3.6.1.

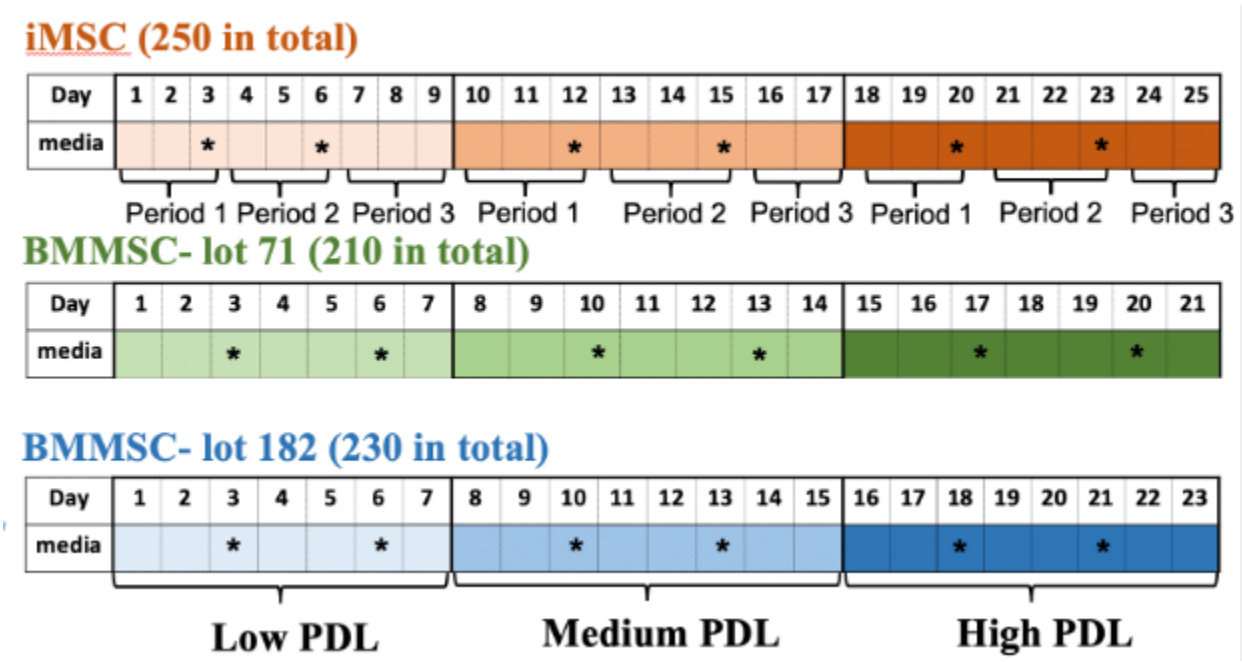


Figure 3.3.1: Culture media sampling strategy. Media were sampled daily for 21-25 days, depending on the cell-line. There were 10 replicates in each cell-line. The asterisk (*) indicates media switch, where fresh media replaced the spent media and the samples were taken before media switch at that day. Three time periods in each PDL are labelled and represent the individual time-course series for analysis.

3.3.2 Variance based feature selection and spectral binning

Each spectrum contains around 33,000 data points with many noises and insignificant peaks. Therefore, a feature reduction was performed to reduce the feature numbers to a more useful set. Variance based feature selection was used based on the assumption that features having high variance across different groups and time points would be the most useful [26]. Unlike other feature selection methods [27], it doesn't require predefined models.

The variance was calculated and ranked for each data point (Figures 3.3.2 and 3.6.3). The water region from 4.69 ppm to 4.88 ppm was removed so the variances of these regions were 0. The top

12% of the features with greatest variances were selected and the other 88% discarded. The 88% threshold was selected to keep most peaks while removing most of the noise and non-significant peaks in the spectra. This method, however, requires human input in determining threshold to filter out datapoints. On the other hand, the highly overlapped regions also show high variance and were kept after filtering. This method is depending on the quality of spectra alignment. However, by using this method here, we are able to filter out most of the noise and baseline of spectra and keep useful information for downstream analysis. The variance filtering reduced the data to 3942 data points. This strategy significantly reduced the complexity of analysis; however, the dataset still contained a very large number of features. Further feature reduction was necessary for downstream analysis.

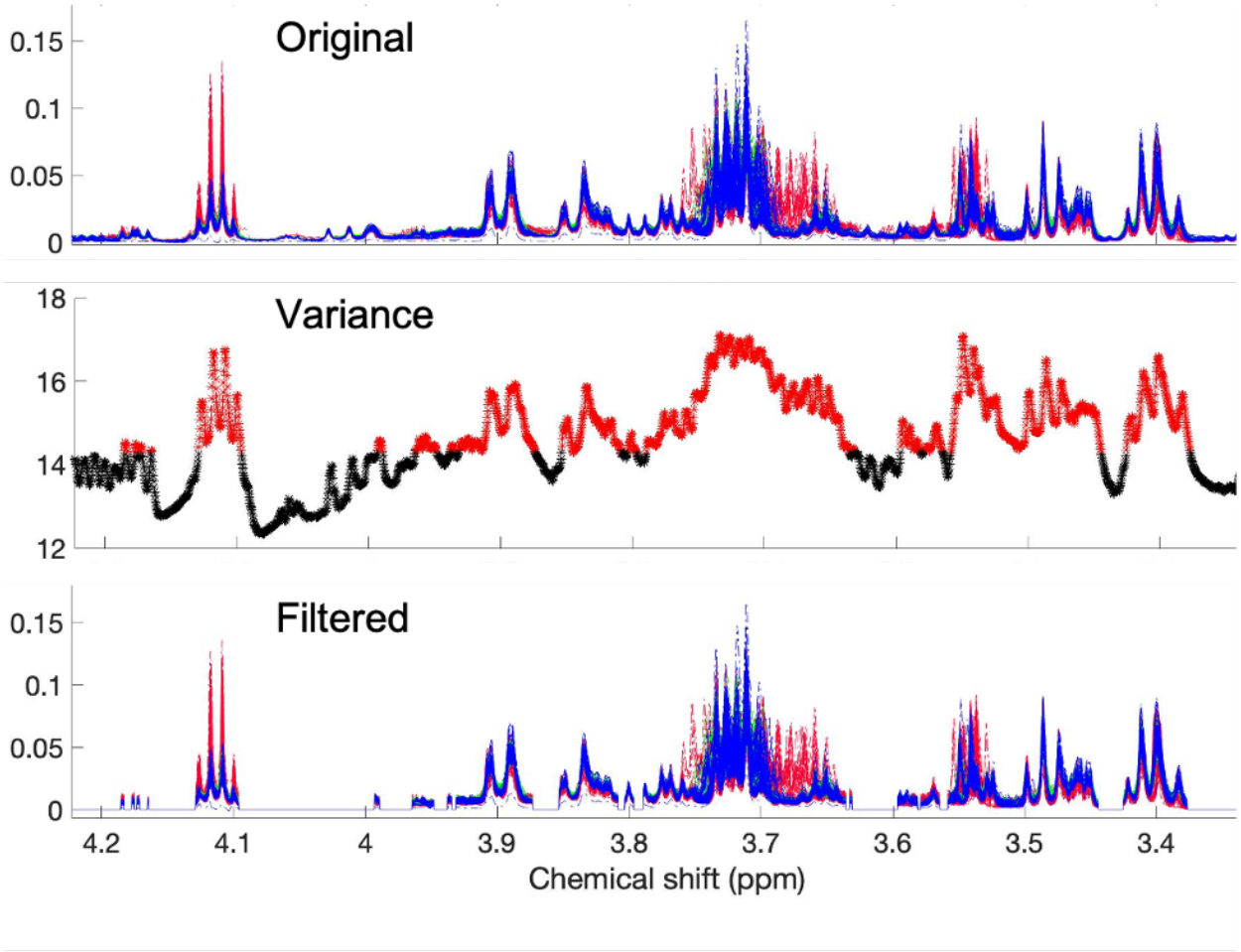


Figure 3.3.2: Variance-based NMR feature selection strategy using a zoomed region from 3.35 ppm to 4.2 ppm. All 690 spectrum of culture media were overlaid. The top panel is the original spectra plotted according to different cell types. red: iMSC, cyan: BM71, blue: BM182. The middle panel is the variance of the NMR resonances across the entire study. The variances were ranked and the top 12%, shown in red, were used. The black points indicate regions of lower variance, which were discarded. The bottom panel is the variance-filtered spectra that were used for downstream analysis.

3.3.3 Discovery of known features

Annotation and spectral binning were used here to further reduce the number of features. A total of 24 metabolites were identified through COLMARm with confidence scores greater than or equal to 3, as shown in Figure 3.6.1 and Table 3.1. Non-overlapped and identified peaks from same metabolite were binned together and quantified according to the number of protons in each peak.

Table 3.1: Identified metabolites with confidence scores

Metabolite	Confidence score	Change through time	Metabolite	Confidence score	Change through time
Leucine	4	↓	Glucose	4	↓
Isoleucine	4	↑	Glycine	4	↑
Valine	4	↑	Fructose	3	↑
Ethanol	4	↑	Lactate	4	↑
Threonine	4	↑	Tyrosine	3	↓
Alanine	4	↑	Phenylalanine	3	↓
Acetate	4	↑	Arginine	4	↓
Glutamine	4	↓	Aspartate	4	↓
Pyruvate	4	↓	Creatine	4	↑
Succinate	4	↑	Betaine	3	↓
Cytidine	3	↓	Proline	3	↓
Formate	3	↓	Myo*inositol*	4	↓

Confidence score 5: samples are spiked with standard materials; 4: 2D C-H and H-H matches database; 3: 2D C-H matches database; 2: Literature or external database matching; 1: putatively guessing.

3.3.4 Association of unknown features

In NMR spectra, non-overlapped peaks from same molecule will be perfectly correlated [28]. Peak overlap and noise can reduce this correlation [29]. Here we utilized this relation to further combine non-identified peaks. The unidentified and non-overlapped peaks were extracted and labeled as

unknown feature 1-26 (uk1-uk26) respectively, and the area under curve was calculated from each. The Pearson correlation was then calculated. As shown in Figure 3.3.3, the peaks in each green box have correlations coefficient greater or equal to 0.7 and thus were binned together as tentative metabolites. Remaining unknown features with correlation values less than 0.8 with all other features were treated as individual unknowns. After spectral binning and correlation analysis, a total of 39 metabolites (containing 24 known and 15 unknown metabolites) were extracted from original spectra.

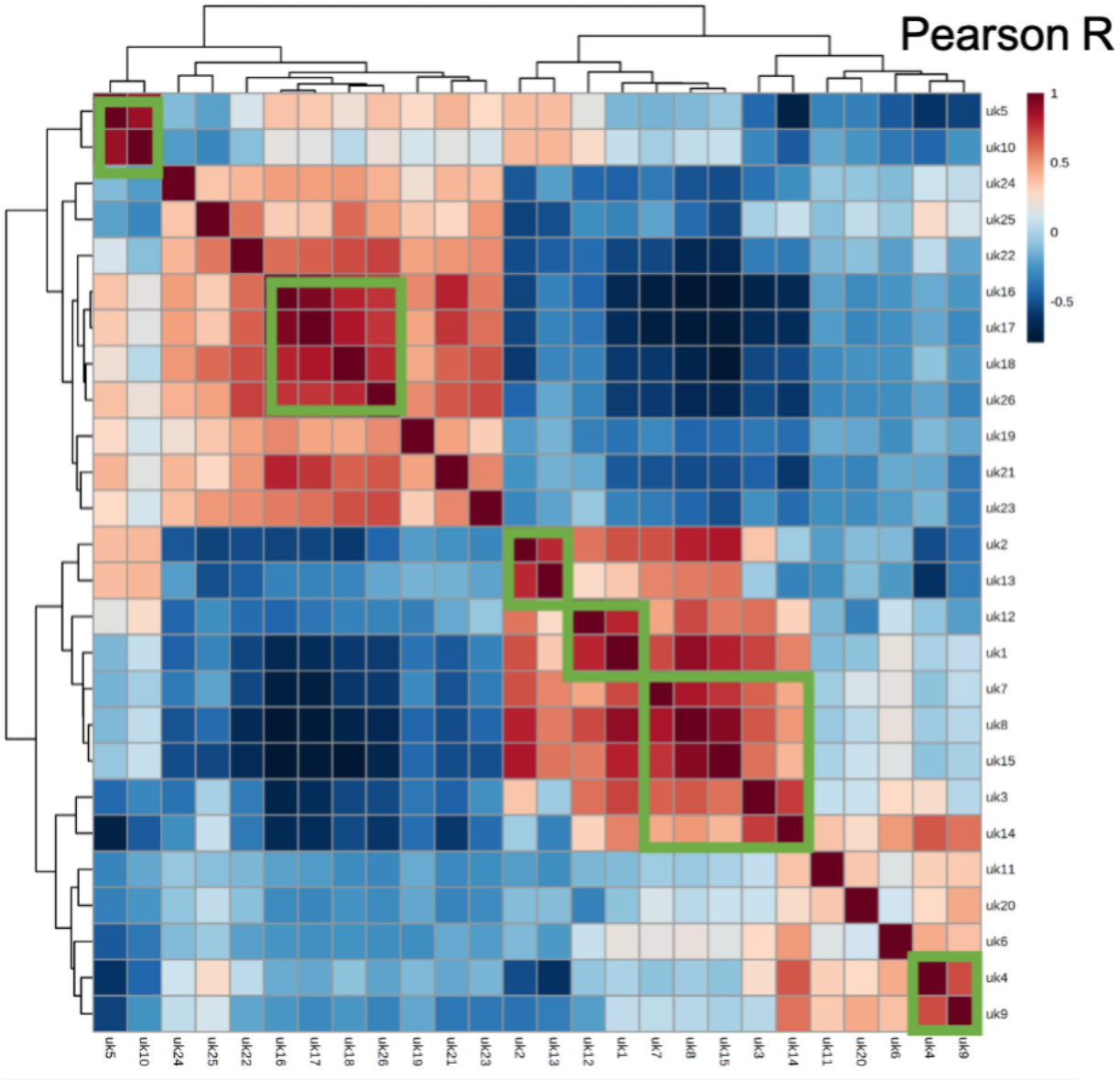


Figure 3.3.3: Unknown feature Pearson correlation and integration. The Pearson correlation was calculated using unknown features, which are designated as ukn (n=1-25). The green boxes indicate features that have correlation values greater or equal to 0.8, thus, have been binned together as tentative metabolites (designated as UKn). The rest of the features were taken as tentative metabolites by themselves.

3.3.5 Media metabolites can predict MSC immunomodulation function

As shown in Figure 3.3.1, this study results in an unbalanced dataset, because the culture time for each passage varied for different cell-lines (and at different passages within a cell-line) from 7-9 days. The correlation of each day metabolite level and composite score was calculated, and the adjusted p value was calculated by applying Benjamini/ Hochberg FDR correction. No metabolite was significant after FDR correction, indicating new data analysis strategy need to be applied to this dataset. In order to deal with the imbalance, culture days were divided to 3 time periods in each PDL, defined by the media switching days (Figure 3.3.1) We first determined metabolite change using linear regression to fit metabolite intensities against culture time. For the period that only contains one single data point, we imputed with day 0 media metabolites level, to make every time period contains at least 2 data points for downstream analysis. For each media switch period that defined a PDL, the metabolite intensities were averaged across replicates before fitting linear regression. All the fits with R2 less than 0.5 were discarded. The regression coefficients for each metabolite with R^2 greater or equal to 0.5 were then used for downstream analyses.

The metabolite regression coefficients were stacked together to form a rank 4 tensor with 34 metabolites x 3 time periods x 3 PDLs x 3 cell-lines (Figure 3.3.4). PLSR is a linear regression technique to find a projection of the predicted and observed variables to a new space. It is particularly useful when dealing with more variables than observations and multicollinearity datasets, where linear regression would fail. PLSR was used to fit the metabolite data tensor for each cell-line/passage (n=9 total experimental groups) against their respective functional composite score generated from our previous work. This composite score encompasses multiple measures of MSC immunomodulation (T-cell suppression and IDO activity) [21]. Variable importance in projection (VIP) scores are useful in ranking X space predictive variables that contributes most to the Y response variance. VIP scores were used to find predictive media markers (PMMs) that contributed most to model the functional composite scores.

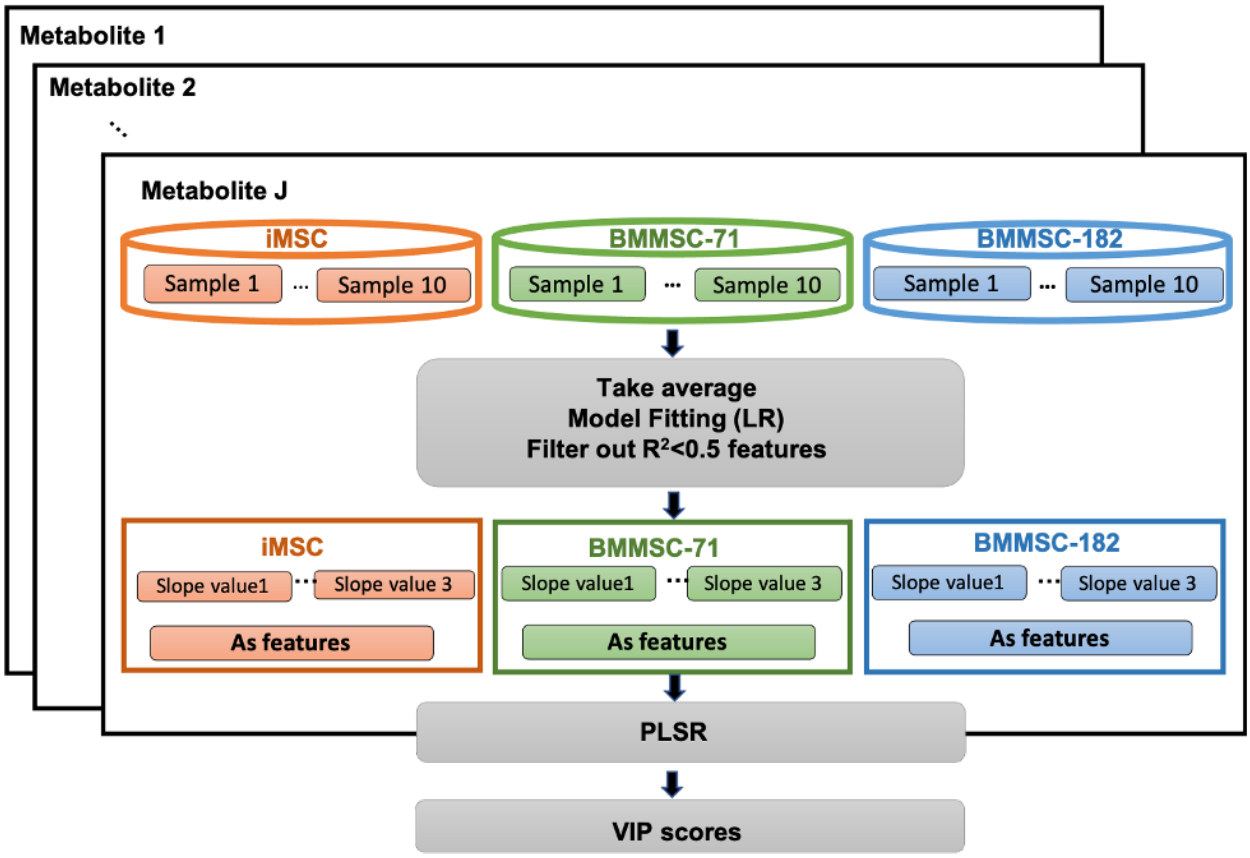


Figure 3.3.4: Time course media data analysis strategy. For each metabolite in each cell-line each passage, linear regression was used to fit averaged metabolite values against time. For each metabolite, linear regression coefficient value was used for downstream analysis. In each cell line, one slope values for each passage respectively. Total 9 slope values for each metabolite were generated. The metabolites with R^2 values less than 0.5 were discarded. Remaining values were used in a PLSR against functional composite scores. VIP scores were used to discover predictive media markers (PMMs).

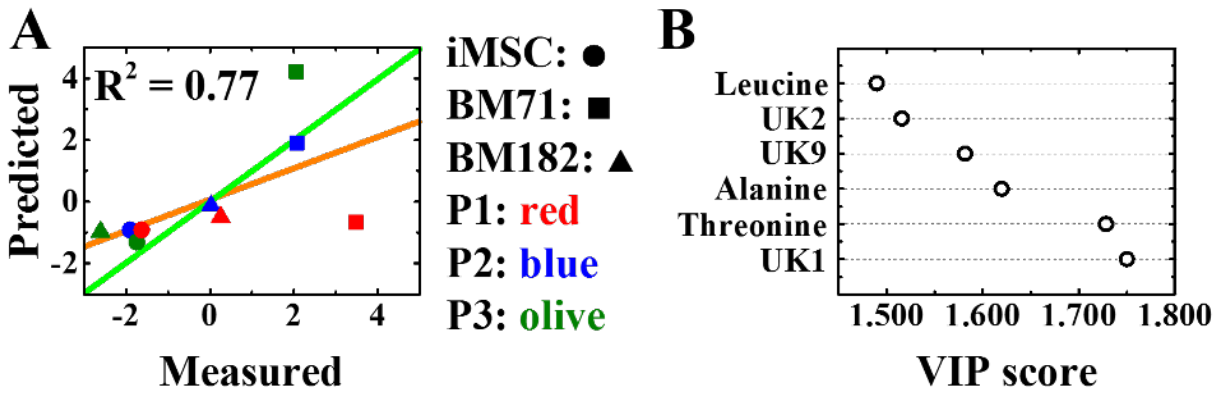


Figure 3.3.5: PLSR and VIP analysis for first three-day culture media. Circle: iMSC; Square: BM71; Triangle: BM182. Orange: p1; Blue: p2; Green: P3. The PLSR model plots show the NMR-predicted vs. measured phenotypic composite score in A. The R^2 values of fit are indicated. The red lines show the best fit and the green lines have a slope of 1. Each point represents a sample. (B) The variable importance plot (VIP) values for the most important six media metabolites that contributed to the model are shown.

Figure 3.3.5 shows PLSR results and the top 6 VIP scores for each culturing period for the iMSC, BM71 and BM182 cell-lines in each passage. The vip score for each metabolite was shown in Table 3.2 in supplementary materials. PMMs include leucine, threonine, alanine, and 3 unknown metabolites. Time period 2 and time period 3 PLSR results were shown in Figure 3.6.3. As the time period one yields around 0.77 R^2 value, indicating the in-process culture media as early as first three days can be used to predict MSC immunomodulation effects accurately. The metabolite change through time was indicated in Table 3.1, with up arrow indicting increase through time and down arrow indicating decrease through time.

3.3.6 Pathway analysis of cellular metabolites

Previously, we showed that predictive cell metabolites (PCMs) can be used to predict a composite immunomodulation functional score [21]. Since the PMMs reported here also predict the same function, we explored the relationship between cellular and media metabolites. The general approach for collecting both types of samples (in-process media and end of production cells) and analyzing metabolomic data is shown in Figure 3.3.1A. From our previous NMR data, we found that cellular phosphocreatine, asparagine, NN-dimethylglycine, myo-inositol, serine, and ukNMR-2 all were important in predicting functional assay results [21]; we called these predictive cellular metabolites (PCMs). We used PLSR to fit different time period media metabolite changes against these PCMs. The top 6 metabolites with highest VIP scores are shown in Figure 3.6.4. Alanine, valine, glycine, creatine and aspartate from the media were important in predicting multiple PCMs at different time periods.

We then used Biocyc to explore PCMs pathways [30]. The averaged intensities of PCMs for each cell-line at all passages were used for intracellular metabolic pathway analysis. The values were standardized using 0 centering and imported into Biocyc to retrieve all pathways that involve these PCMs. The upstream/downstream metabolites of these PCMs were linked together to get the pathway map shown in Figure 3.6.5. NN-dimethylglycine is a precursor to glycine, which is a substrate for serine biosynthesis. Glycine is also involved in the phosphocreatine biosynthesis. Alanine and serine are on the same biosynthetic pathway. Glycine, aspartate and pyruvate are link metabolites which involve in PCMs regulations.

3.3.7 Some media and cellular metabolites are correlated

We next did a Pearson correlation between media and cellular metabolites. We focused on time period 1 metabolic change of MMs including valine, alanine, glycine, creatine and aspartate were used for correlation analysis. These MMs were found predictive of multiple PCMs as shown in

Figure 3.6.4. The cell metabolites identified in Figure 3.6.5 were also used in correlation analysis, and Figure 3.3.6 shows the correlation coefficient values. Valine and aspartate in media negatively correlate with phosphocreatine, asparagine and glycine in cells but positively correlate with NN-dimethylglycine and serine in cells. In contrast, alanine, glycine and creatine in media positively correlate with phosphocreatine, asparagine and glycine in cells, and negatively correlate with NN-dimethylglycine and serine in cells. The aspartate is decreasing through time, and higher the decrease rate corresponding to higher asparagine level in cells. In contrast, the glycine is increasing through time, and higher glycine increase rate corresponding to higher NN-dimethylglycine level in end product of cells. These correlation relations were also used to label the metabolites change in cells and media respectively in Figure 3.3.7.

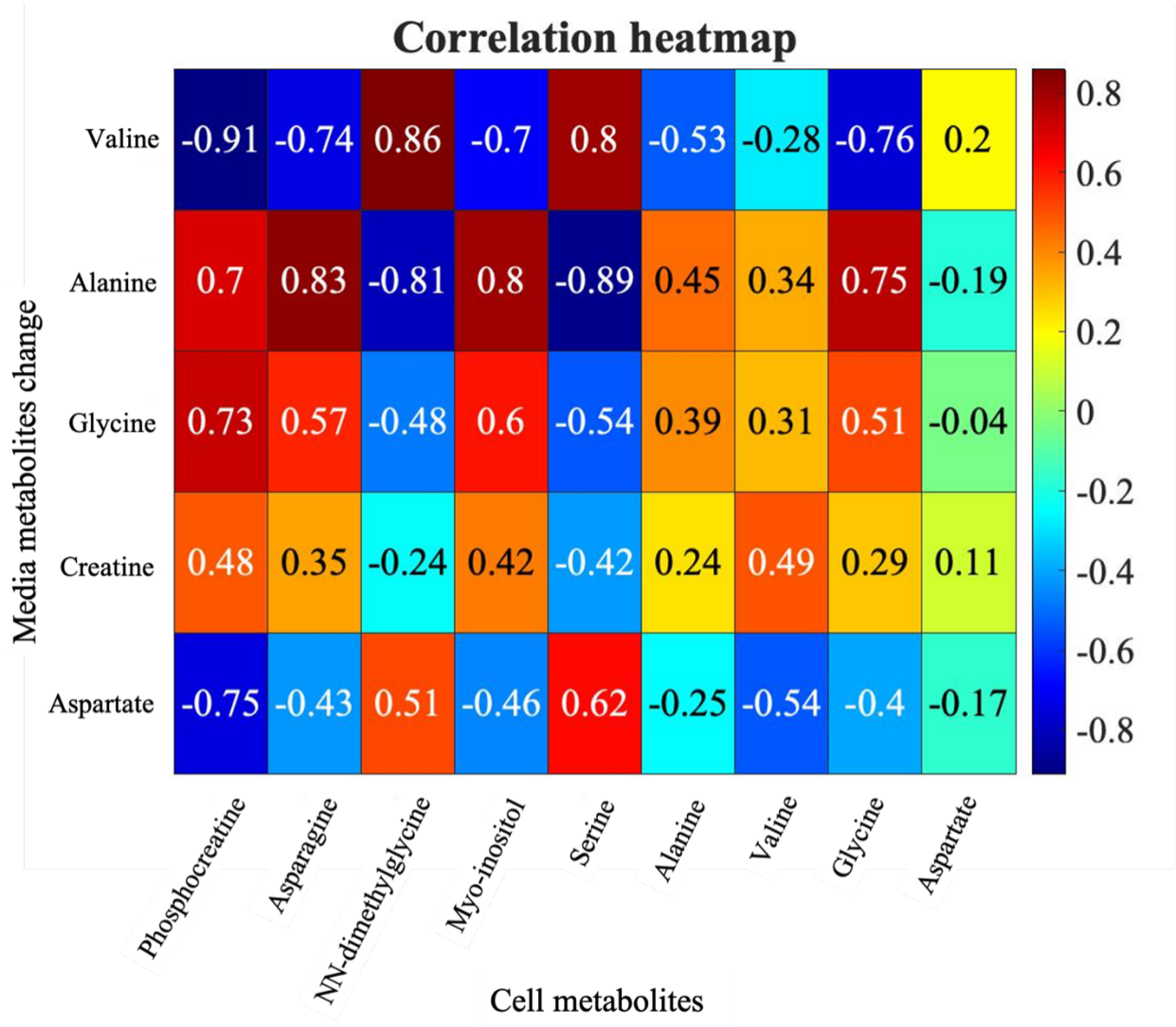


Figure 3.3.6: Correlation heatmap. Correlation of time period 1 metabolites change in media and metabolites relative intensities in cells. The media metabolites were selected according to Figure 3.6.4, where these media metabolites (MMs) were important in predicting cell metabolites. The cell metabolites were selected according to Figure 3.6.5, where these cell metabolites (CMs) were involved in main pathways we identified. Color bar and correlation coefficient values (r value) were indicated.

3.3.8 Union pathway analysis of CMs and MMs

Both CMs and MMs from Figure 3.3.6 were imported into biocyc for a union pathway analysis. All possible linkage of CMs and MMs were retrieved from the database and kept in main pathway. The pathways that did not include these CMs and MMs were removed to generate union pathway (Figure 3.3.7). The up/down regulation of metabolites in pathways were labelled using arrows according to Figure 3.3.6. The green arrows indicate metabolites from cells while the yellow arrows metabolites from media. The up/down regulation of metabolites in media corresponds to the increase/decrease of metabolite level in media through time. The up/down regulation in cells represents the relations between cellular metabolites and MSC T-cell suppression and IDO assay results.

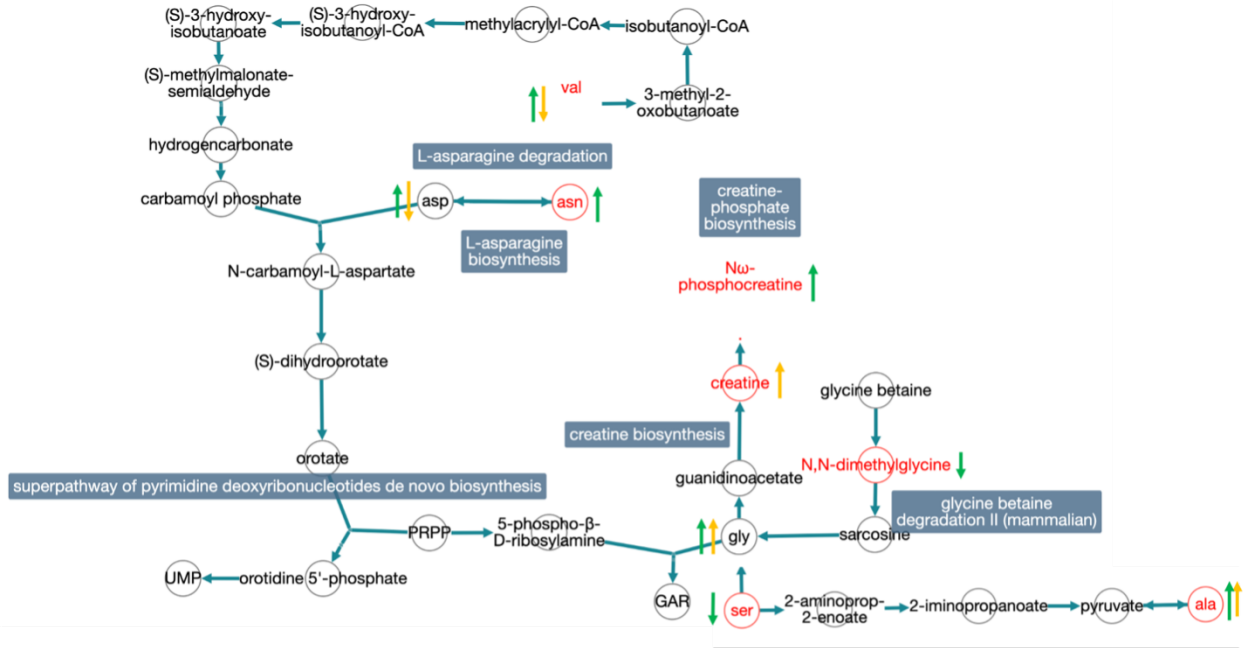


Figure 3.3.7: Pathway analysis of both cell pellets and media metabolites. Pathway analysis performed using biocyc. These pathways involve key metabolites in both cell pellets and media. The red metabolites are important metabolites we identified, and the black metabolites are the link metabolites. All pathways that are not involving target metabolites have been trimmed. The arrow direction indicates up/down regulation in cells/media based on correlation heatmap. The yellow arrows indicate regulation in media, the green arrows indicate regulation in cells.

3.4 Discussion

MSCs have been widely investigated for their well-documented immunomodulatory capacity, but lack of predictive biomarkers limits their use in clinic. We previously showed that cellular metabolites can accurately model IDO and T-cell suppression assays [21], We aim to find non-destructive methods that enable on-demand assessment of cell quality and to better understand how metabolism regulates MSC function.In this paper, we show that a simple, non-destructive in-process and high

throughput method of monitoring changes in culture media can predict MSC immunomodulatory function. In this study, we only used three time points in each time period, which limited our analysis of extracellular metabolites change to linear. In the future, real time monitoring is feasiable with benchtop NMR to get continuous metabolomics data from culture media and will make our prediction more accurate.

3.4.1 The effects of PMMs on MSCs

In this study, we find some PMMs that can predict MSCs IDO assay and T cell suppression assay results including leucine, threonine, and alanine. Leucine is one of the three essential branched chain amino acids (BCAAs) which is essential for cell proliferation [31–33]. In addition to cell proliferation, Leucine also has effects on immunity [34, 35]. Studies have shown that inadequate uptake of leucine leads to immune impairment [36–39]. Leucine metabolites hydroxymethylbutyrate (HMB) and ketoisocaproic acid (KIC) have also shown immunomodulatory effects. The treatment of HBM reduced TNF- concentrations in PBMC culture medium and impaired lymphocyte proliferation [40]. Similarly, KIC was found suppressing lymphocyte DNA synthesis in activated PBMCs [41]. Talita Sartori et al. have shown that leucine can improve MSCs proliferation. They also found leucine expand the immunosuppression capacity of MSCs by reducing the expression of the phosphorylated nuclear factor kappa-light-chain-enhancer of activated B cells (p-NFκB/NFκB ratio [42]. In our study, we found that leucine was decreased through time in media (Table 3.1), indicating the cells consuming leucine, the higher the decrease rate corresponding to the higher decrease of T cell proliferation and increase of IDO level in cells. Our findings are consistent with the literature [19] and indicate that MSCs consuming more leucine in media have higher immunomodulation effects. Previously study has shown that increase of alanine in culture media can decrease human induced pluripotent stem cells (hiPSCs) viability. When co-culture with differentiated cells, increase of alanine concentration in media can selectively eliminate undifferentiated hiPSCs [43]. Our data indicate the increase of alanine levels in media during culturing (Table 3.1). Our study also shows

threonine consumption over time (Table 3.1). However, the effects of threonine and alanine on human MSCs immunomodulation and the underlying mechanism still need to be investigated.

3.4.2 The relations between extracellular MMs and intracellular CMs in MSCs

Our study also investigated the relations between MMs and CMs in MSC. Several MMs are specifically predictive to CMs, including alanine, glycine, valine, creatine and aspartate. Correlation reveals some positive/negative relations between MMs and CMs. Pathway analyses were then performed to try to explain these relations.

In this union pathway map, we are not able to directly link valine and asparagine with other metabolites together, some other pathways that do not present in database may also involve in regulating these metabolites. Although we observe the relations between MMs and CMs, the underlying mechanisms are still not clear. However, we can still get some useful information from this pathway analysis.

From the union pathways we generated, valine and aspartate level were decreased in media over time, indicating that cells consume valine and aspartate during culture [19]. Valine is one of the BCAAs and as discussed before, BCAAs involve in MSCs immunosuppression pathways, increased level of BCAAs improve MSCs proliferation and expand the immunomodulatory effects. Our finding further confirms that increase of valine consumption in MSCs culture media is correlated with MSCs immunomodulation capacity The higher consumption rate corresponds to higher T cell suppression effects. Asparagine is synthesized from aspartate. Studies showed that human MSCs consumes aspartate in media [19]. Indeed, we observed a decrease of aspartate level in media through time, and the higher the decrease rate corresponds to higher asparagine level in cells. This evidence shows that cells consuming aspartate to generate asparagine. The increase of alanine in media suggests the MSCs secrete alanine into media. The higher level of glycine in cells directly links to the higher level of phosphocreatine in cells, as phosphocreatine is synthesized from glycine.

In this study, we find some PMMs that can predict MSC IDO assay and T cell suppression assay results, including leucine, threonine and alanine. Leucine is well-known a regulator of immunity in cells. However, no evidence shows the direct relations between threonine, alanine and cell immunomodulation effects The underlying mechanism of these PMMs effects need to be addressed in the future. We also performed pathway analysis to explain some of the underlying relationships between extracellular metabolites change and intracellular metabolites level. We found that cells consume valine and aspartate during culture, and the consumption of aspartate contributes to the accumulation of asparagine in cells. In addition, MSCs secrete alanine into media during culture. However, how these metabolites link to MSC immunomodulation function still unknown. Overall, this study shows MSCs immunosuppression capacity can be predicted by monitoring culture media. This method is non-destructive, high throughput and can be further applied to cell manufacturing.

3.5 Abbreviations

Mesenchymal stem/stromal cells (MSC); critical quality attributes (CQA); The International Society for Cell and Gene Therapy (ISCT); indoleamine 2,3-dioxygenase (IDO); nuclear magnetic resonance (NMR); partial least squares regression (PLSR); variable importance projection (VIP); predictive media markers (PMMs); predictive cell markers (PCMs); media markers (MMs); cell markers (CMs); one dimensional nuclear Overhauser enhancement spectroscopy with water suppression (1D-NOESY PR); correlation optimized wrapping (COW); probabilistic quotient normalization (PQN); heteronuclear single quantum coherence (HSQC); HSQC-total correlated spectroscopy (HSQC-TOCSY); unknown feature n (ukn); unknown NMR metabolite n (UK-n); asparaginase synthetase (ASNS); branched chain amino acids (BCAAs)

3.6 Supplementary materials

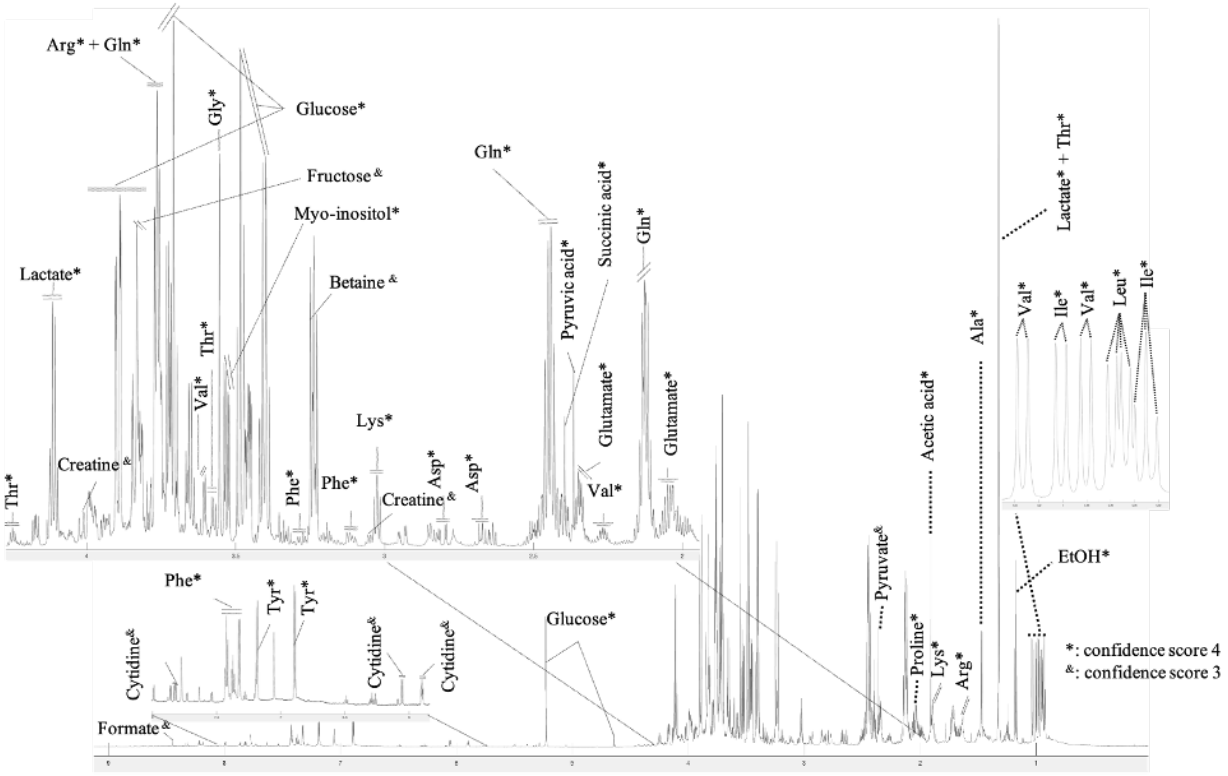


Figure 3.6.1: Media metabolites identification with confidence score. Confidence score 4: peak was matched to database on both 2D HSQC and HSQC-TOSCY, and 1D spectra matched to 1D database. Confidence score 3: peak was matched to database on either 2D HSQC or HSQC-TOSCY, and 1D spectra matched to 1D database.

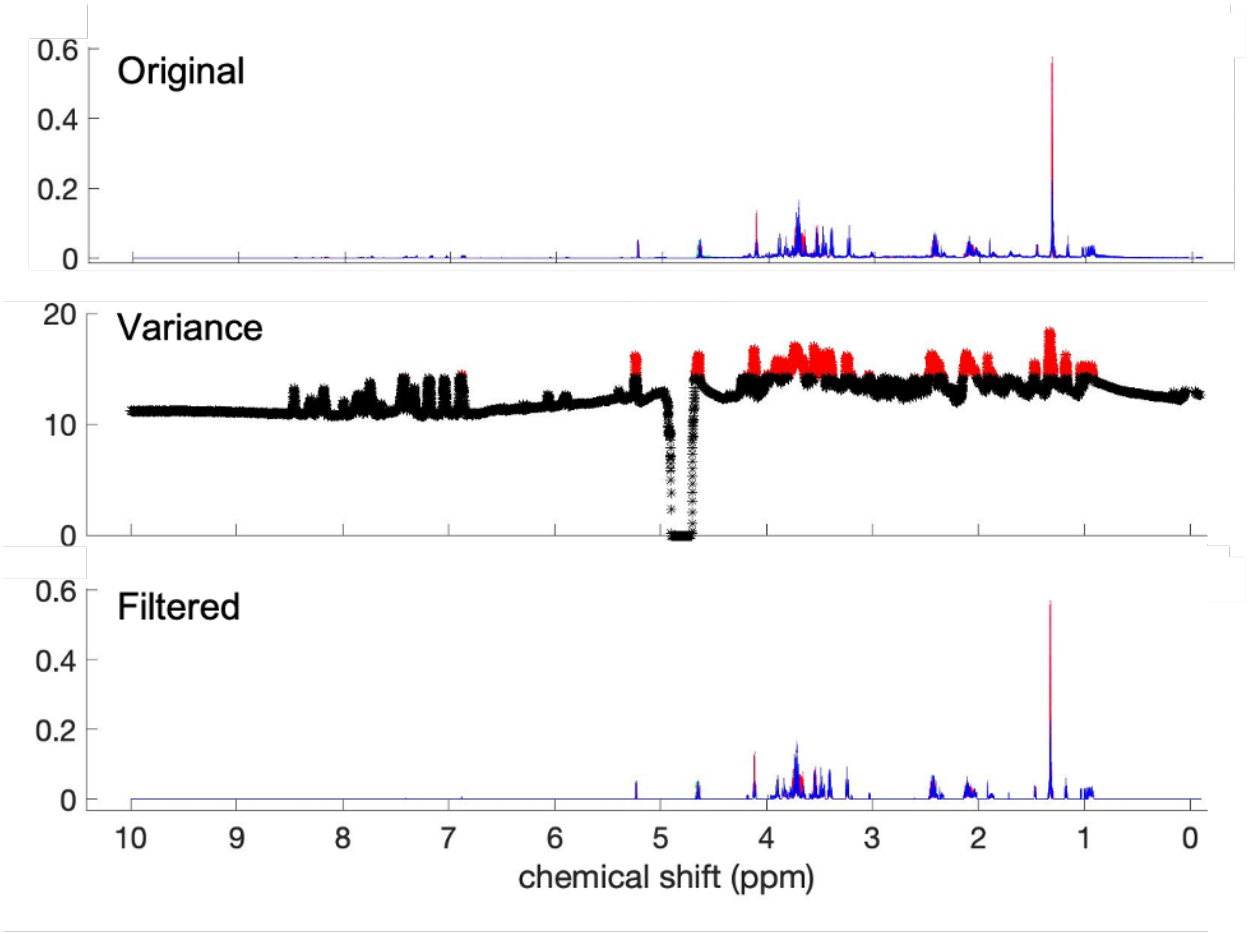


Figure 3.6.2: Variance based feature selection strategy. Top panel is the original spectra plotted according to different cell types. red: iMSC, cyan: BM71, blue: BM182. Middle panel is the variance calculation. The variances were ranked and top 12% were kept and the rest were filtered out. red indicates all the data points that passed the filtering. Black indicates the data points that have been discarded. Bottom panel is the filtered spectra.

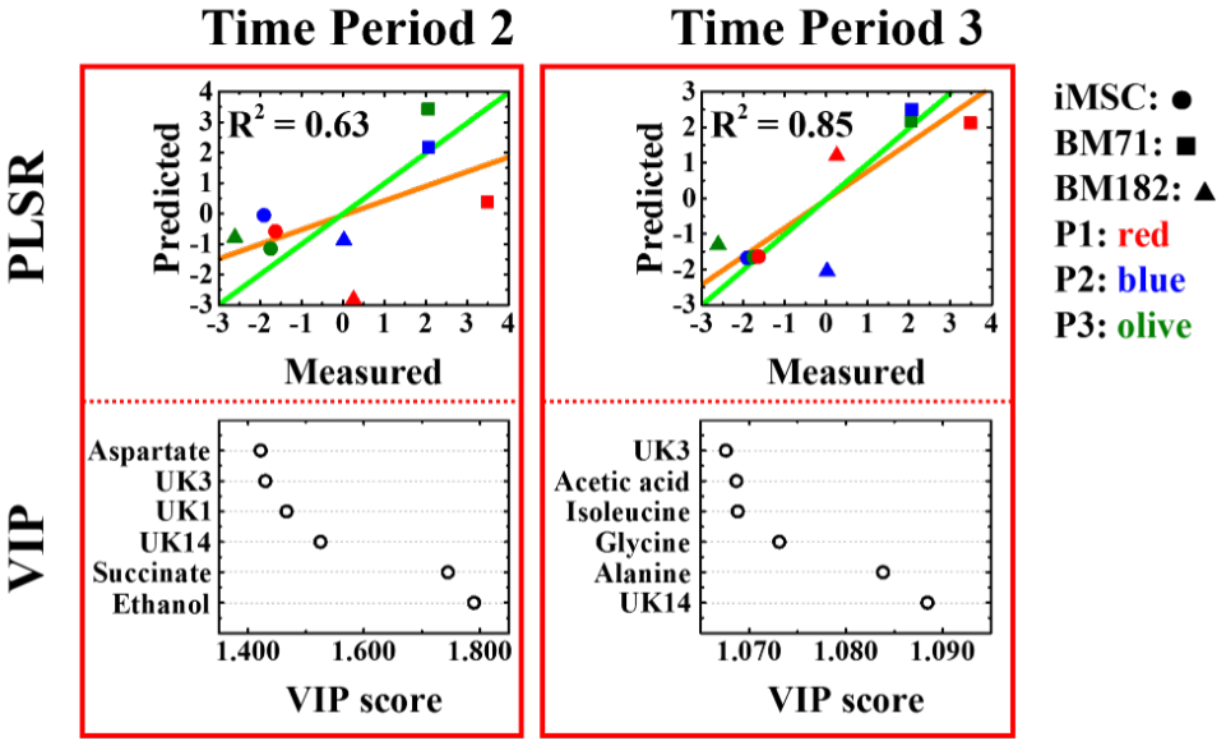
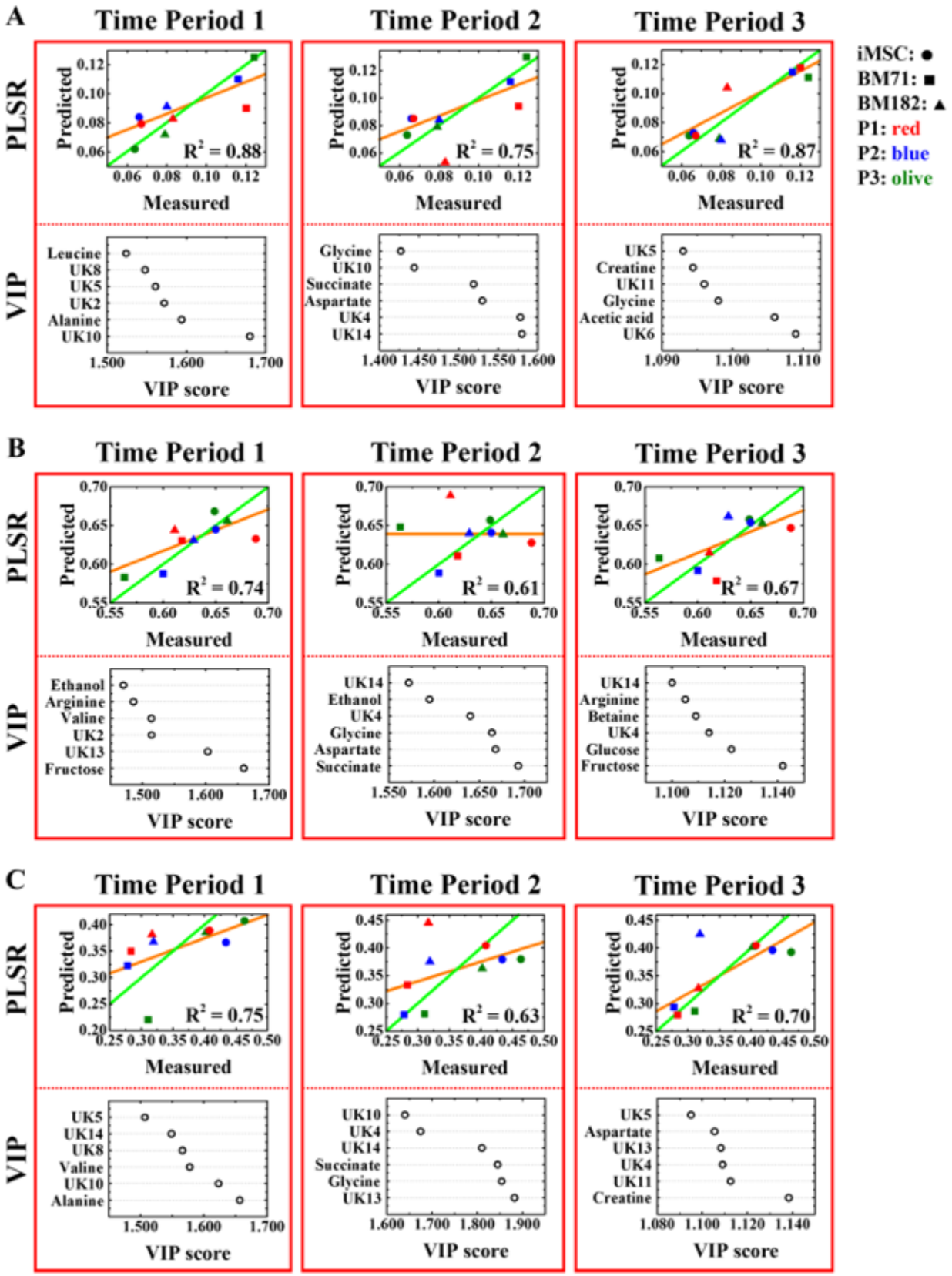


Figure 3.6.3: PLSR and VIP analysis for time period 2 and 3. Circle: iMSC; Square: BM71; Triangle: BM182. Orange: p1; Blue: p2; Green: P3. The PLSR model plots show the NMR-predicted vs. measured phenotypic composite score. The R^2 values of each fit are indicated. The red lines show the best fit and the green lines have a slope of 1. Each point represents a sample. The variable importance plot (VIP) values for the most important six media metabolites that contributed to the model are shown below each PLSR model.


A
Time Period 1
Time Period 2
Time Period 3
PLSR
VIP
Predicted
Measured
R² = 0.88
R² = 0.75
R² = 0.87
Leucine
UK8
UK5
UK2
Alanine
UK10
Glycine
UK10
Succinate
Aspartate
UK4
UK14
UK5
Creatine
UK11
Glycine
Acetic acid
UK6
VIP score
iMSC: ●
BM71: ■
BM182: ▲
P1: red
P2: blue
P3: olive
B
R² = 0.74
R² = 0.61
R² = 0.67
Ethanol
Arginine
Valine
UK2
UK13
Fructose
UK14
Ethanol
UK4
Glycine
Aspartate
Succinate
UK14
Arginine
Betaine
UK4
Glucose
Fructose
C
R² = 0.75
R² = 0.63
R² = 0.70
UK5
UK14
UK8
Valine
UK10
Alanine
UK10
UK4
UK14
Succinate
Glycine
UK13
UK5
Aspartate
UK13
UK4
UK11
Creatine

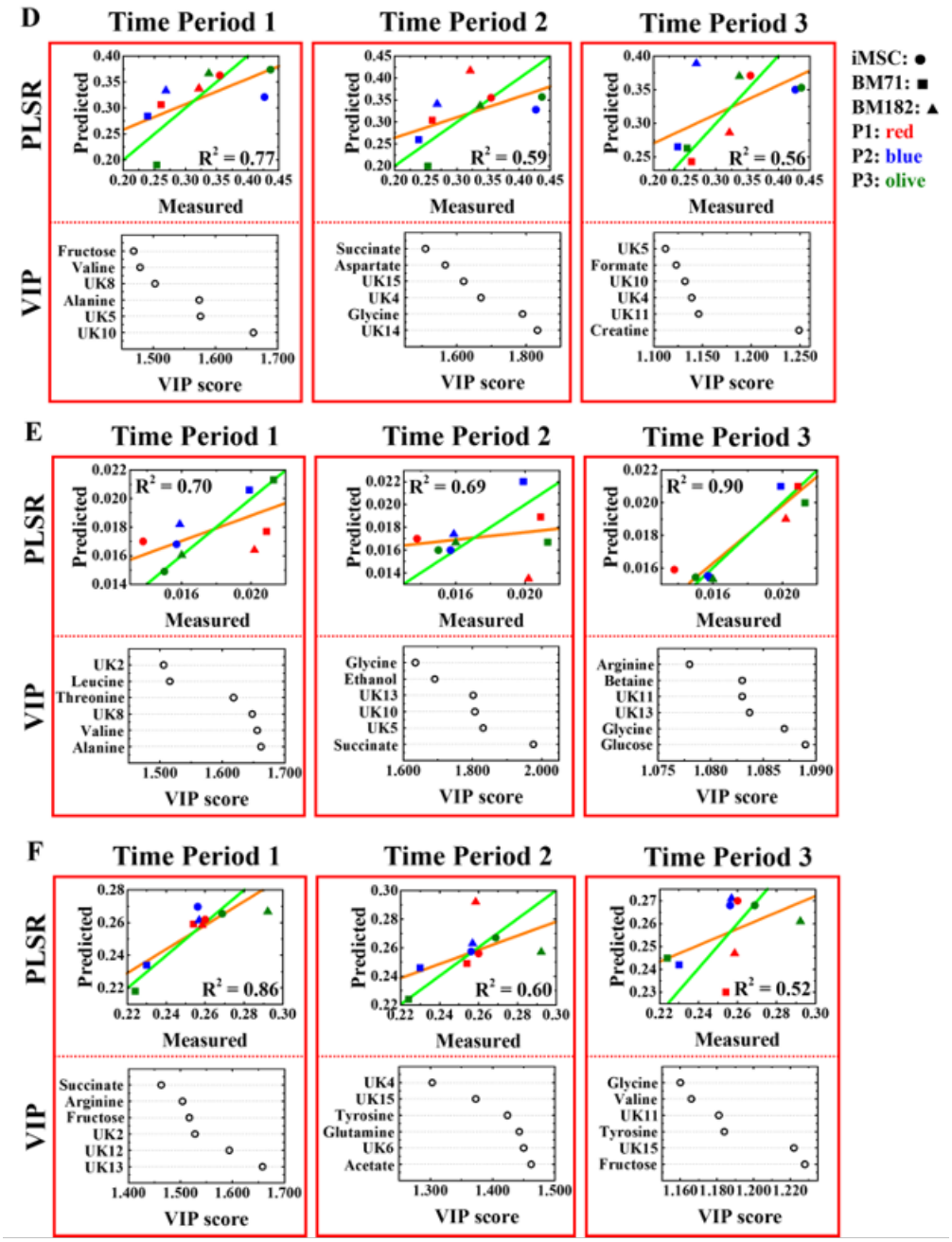


Figure 3.6.4: Time period media data regressed to metabolites in cell pellets. Circle: iMSC; Square: BM71; Triangle: BM182. Orange: p1; Blue: p2; Green: P3. The R^2 values of the fit are indicated. The red lines show the best fit and the green lines have a slope of 1. Each point represents a sample.From left to right indicates three time period.

From up to bottom showed serine, phosphocreatine, asparagine, myo-inositol, NN-dimethylglycine, and UKNMR2 PLSR model plots of the predicted vs measured values for each metabolite respectively, and corresponding variable importance plot (VIP) values of six most important metabolite change in media.

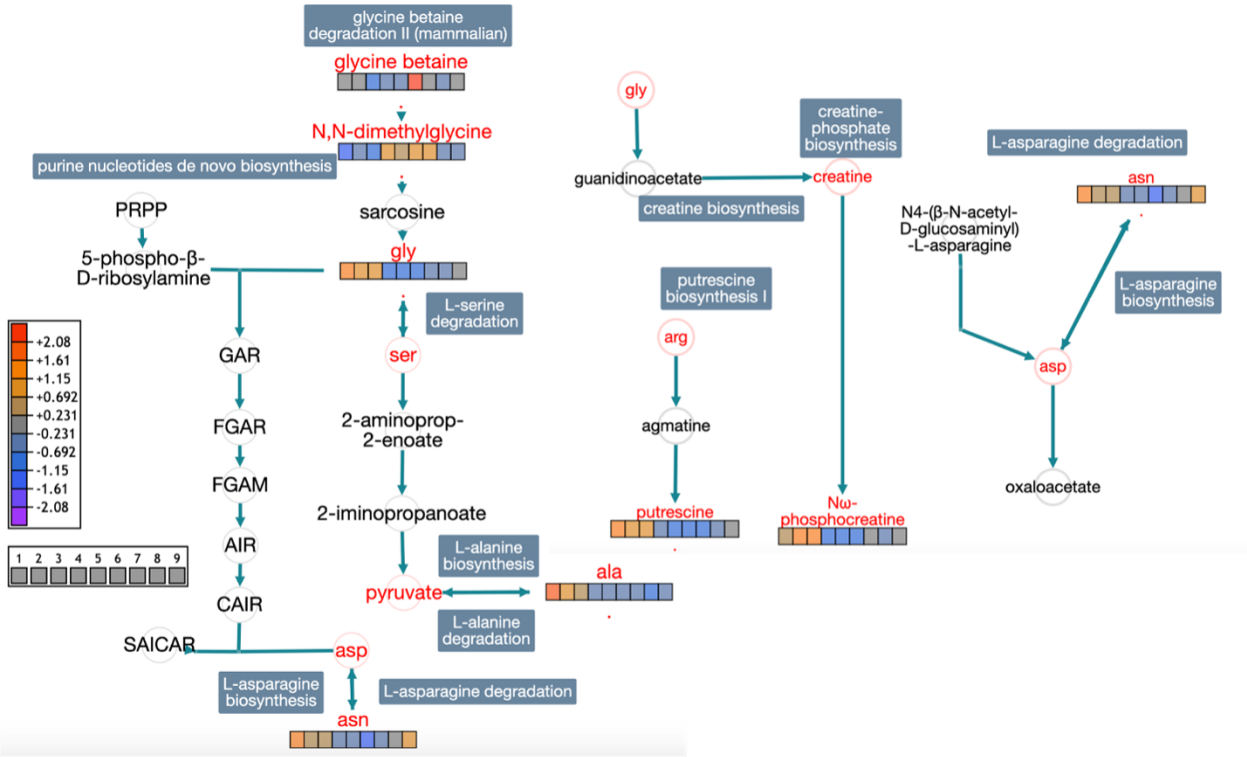


Figure 3.6.5: Pathway that involves predictive cell metabolites (PCMs). Pathway analysis performed using biocyc. The averaged metabolites values of each cell line in each population doubling level (PDL) were imported and standardized using 0 centering for generating pathway map. Color bar under each metabolite indicates the metabolite value after 0 centering. From left to right, they are: iMSC-low PDL, iMSC-medium PDL, iMSC-high PDL; BM71-low PDL, BM71-medium PDL, BM71-high PDL; BM182 -low PDL, BM182- medium PDL, BM182-high PDL. The red indicates the metabolites we identified in cell pellets and black indicates link metabolites.

Table 3.2: VIP scores of all metabolites in first time period

metabolite	VIP score
uk1	1.75009178
threonine	1.72833506
alanine	1.61968445
uk9	1.5814669
uk2	1.51561933
leucine	1.48950142
valine	1.48286105
arginine	1.38966506
glucose	1.36902451
fructose	1.35307034
lactate	1.34457938
myo$_i$*nositol*	1.32965603
succinate	1.23302812
asparate	1.13530122
ethanol	1.05322467
glycine	1.00834252
creatine	0.97187434
acetic$_a$*cid*	0.86386268
uk4	0.83102048
uk14	0.78144143
tyrosine	0.76789738
betaine	0.65692652
uk3	0.64944866

uk13	0.61157025
glutamine	0.60015384
formate	0.5754882
uk6	0.55941284
uk7	0.51159124
uk15	0.46344727
uk11	0.45663475
phenylalanine	0.36734515
pyruvate	0.3009289
uk12	0.25590462
uk8	0.24345107
proline	0.23801424
uk5	0.16931079
isoleucine	0.13039402
cytidine	0.11090003
uk10	0.01926329

BIBLIOGRAPHY

[1] Feng-Juan Lv et al. “Concise review: the surface markers and identity of human mesenchymal stem cells”. In: *Stem cells* 32.6 (2014), pp. 1408–1419.

[2] Yufang Shi et al. “Immunoregulatory mechanisms of mesenchymal stem and stromal cells in inflammatory diseases”. In: *Nature Reviews Nephrology* 14.8 (2018), pp. 493–507.

[3] Mudasir Bashir Gugjoo, Gutulla Taru Sharma, et al. “Equine mesenchymal stem cells: properties, sources, characterization, and potential therapeutic applications”. In: *Journal of equine veterinary science* 72 (2019), pp. 16–27.

[4] Mudasir Bashir Gugjoo, Amarpal Amarpal, and Gutulla Taru Sharma. “Mesenchymal stem cell basic research and applications in dog medicine”. In: *Journal of cellular physiology* 234.10 (2019), pp. 16779–16811.

[5] Rosa Yañez et al. “Adipose tissue-derived mesenchymal stem cells have in vivo immunosuppressive properties applicable for the control of the graft-versus-host disease”. In: *Stem cells* 24.11 (2006), pp. 2582–2591.

[6] Nora G Singer and Arnold I Caplan. “Mesenchymal stem cells: mechanisms of inflammation”. In: *Annual Review of Pathology: Mechanisms of Disease* 6 (2011), pp. 457–478.

[7] Donald G Phinney. “Functional heterogeneity of mesenchymal stem cells: implications for cell therapy”. In: *Journal of cellular biochemistry* 113.9 (2012), pp. 2806–2812.

[8] CM McLeod and RL Mauck. “On the origin and impact of mesenchymal stem cell heterogeneity: new insights and emerging tools for single cell analysis”. In: *European cells & materials* 34 (2017), p. 217.

[9] Jacques Galipeau et al. “International Society for Cellular Therapy perspective on immune functional assays for mesenchymal stromal cells as potency release criterion for advanced phase clinical trials”. In: *Cytotherapy* 18.2 (2016), pp. 151–159.

[10] MLBK Dominici et al. “Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement”. In: *Cytotherapy* 8.4 (2006), pp. 315–317.

[11] S Schachtele, C Clouser, and J Aho. "Markers and Methods to Verify Mesenchymal Stem Cell Identity". In: *Potency, and Quality.[(accessed on 29 January 2020)]* ().

[12] Li-Tzu Wang et al. "Differentiation of mesenchymal stem cells from human induced pluripotent stem cells results in downregulation of c-Myc and DNA replication pathways with immunomodulation toward CD4 and CD8 cells". In: *Stem Cells* 36.6 (2018), pp. 903–914.

[13] Roland Meisel et al. "Human bone marrow stromal cells inhibit allogeneic T-cell responses by indoleamine 2, 3-dioxygenase–mediated tryptophan degradation". In: *Blood* 103.12 (2004), pp. 4619–4621.

[14] Ty S Maughon et al. "Multi-omics characterization of mesenchymal stem/stromal cells for the identification of putative critical quality attributes". In: *bioRxiv* (2021).

[15] Sebastien Hagmann et al. "Different culture media affect growth characteristics, surface marker distribution and chondrogenic differentiation of human bone marrow-derived mesenchymal stromal cells". In: *BMC musculoskeletal disorders* 14.1 (2013), pp. 1–11.

[16] Corina Vater, Philip Kasten, and Maik Stiehler. "Culture media for the differentiation of mesenchymal stromal cells". In: *Acta biomaterialia* 7.2 (2011), pp. 463–477.

[17] Amandda Évelin Silva-Carvalho, Francisco Assis Rocha Neves, and Felipe Saldanha-Araujo. "The immunosuppressive mechanisms of mesenchymal stem cells are differentially regulated by platelet poor plasma and fetal bovine serum supplemented media". In: *International Immunopharmacology* 79 (2020), p. 106172.

[18] Markus M Rinschen et al. "Identification of bioactive metabolites using activity metabolomics". In: *Nature Reviews Molecular Cell Biology* 20.6 (2019), pp. 353–367.

[19] Megan R Showalter et al. "Primed mesenchymal stem cells package exosomes with metabolites associated with immunomodulation". In: *Biochemical and biophysical research communications* 512.4 (2019), pp. 729–735.

[20] Gustavo A Higuera et al. "Patterns of amino acid metabolism by proliferating human mesenchymal stem cells". In: *Tissue engineering Part A* 18.5-6 (2012), pp. 654–664.

[21] Wei Jiang and Jianyong Xu. "Immune modulation by mesenchymal stem cells". In: *Cell proliferation* 53.1 (2020), e12712.

[22] Xuegang Yuan, Timothy M Logan, and Teng Ma. "Metabolism in human mesenchymal stromal cells: a missing link between hMSC biomanufacturing and therapy?" In: *Frontiers in immunology* 10 (2019), p. 977.

[23] Frank Delaglio et al. "NMRPipe: a multidimensional spectral processing system based on UNIX pipes". In: *Journal of biomolecular NMR* 6.3 (1995), pp. 277–293.

[24] Keshav Kumar, Ralf Schweiggert, and Claus-Dieter Patz. “Introducing a novel procedure for peak alignment in one-dimensional 1 H-NMR spectroscopy: a prerequisite for chemometric analyses of wine samples”. In: *Analytical Methods* 12.28 (2020), pp. 3626–3636.

[25] Frank Dieterle et al. “Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics”. In: *Analytical chemistry* 78.13 (2006), pp. 4281–4290.

[26] Jacquelyn M Walejko et al. “Global metabolomics of the placenta reveals distinct metabolic profiles between maternal and fetal placental tissues following delivery in non-labored women”. In: *Metabolites* 8.1 (2018), p. 10.

[27] Wei Zheng et al. “Unsupervised feature selection by self-paced learning regularization”. In: *Pattern Recognition Letters* 132 (2020), pp. 4–11.

[28] Olivier Cloarec et al. “Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets”. In: *Analytical chemistry* 77.5 (2005), pp. 1282–1289.

[29] Andrés Charris-Molina et al. “Tackling the Peak Overlap Issue in NMR Metabolomics Studies: 1D Projected Correlation Traces from Statistical Correlation Analysis on Nontilted 2D 1H NMR J-Resolved Spectra”. In: *Journal of proteome research* 18.5 (2019), pp. 2241–2253.

[30] Ron Caspi et al. “The MetaCyc database of metabolic pathways and enzymes-a 2019 update”. In: *Nucleic acids research* 48.D1 (2020), pp. D445–D453.

[31] Peng Li et al. “Amino acids and immune function”. In: *British Journal of Nutrition* 98.2 (2007), pp. 237–252.

[32] Phillip C Calder. “Branched-chain amino acids and immunity”. In: *The Journal of nutrition* 136.1 (2006), 288S–293S.

[33] JC Chuang, CL Yu, and SR Wang. “Modulation of human lymphocyte proliferation by amino acids”. In: *Clinical & Experimental Immunology* 81.1 (1990), pp. 173–176.

[34] Andrea Bonvini et al. “Immunomodulatory role of branched-chain amino acids”. In: *Nutrition reviews* 76.11 (2018), pp. 840–856.

[35] Vinicius Fernandes Cruzat, Maurıcio Krause, and Philip Newsholme. “Amino acid supplementation and impact on immune function in the context of exercise”. In: *Journal of the international Society of Sports Nutrition* 11.1 (2014), pp. 1–13.

[36] Yu-Ping Deng et al. “Dietary leucine improves flesh quality and alters mRNA expressions of Nrf2-mediated antioxidant enzymes in the muscle of grass carp (Ctenopharyngodon idella)”. In: *Aquaculture* 452 (2016), pp. 380–387.

[37] Jae Hoon Lee et al. "Anti-inflammatory and anti-genotoxic activity of branched chain amino acids (BCAA) in lipopolysaccharide (LPS) stimulated RAW 264.7 macrophages". In: *Food science and biotechnology* 26.5 (2017), pp. 1371–1377.

[38] Roberta De Simone et al. "Branched-chain amino acids influence the immune properties of microglial cells and their responsiveness to pro-inflammatory signals". In: *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1832.5 (2013), pp. 650–659.

[39] Ting Liu et al. "NF-κB signaling in inflammation". In: *Signal transduction and targeted therapy* 2.1 (2017), pp. 1–9.

[40] Talita Sartori et al. "Branched chain amino acids improve mesenchymal stem cell proliferation, reducing nuclear factor kappa B expression and modulating some inflammatory properties". In: *Nutrition* 78 (2020), p. 110935.

[41] Michail E Klontzas et al. "Metabolomics analysis of the osteogenic differentiation of umbilical cord blood mesenchymal stem cells reveals differential sensitivity to osteogenic agents". In: *Stem cells and development* 26.10 (2017), pp. 723–733.

[42] Peter B Alexander, Jian Wang, and Steven L McKnight. "Targeted killing of a mammalian cell based upon its specialized metabolic state". In: *Proceedings of the National Academy of Sciences* 108.38 (2011), pp. 15828–15833.

[43] Takunori Nagashima et al. "Selective elimination of human induced pluripotent stem cells using medium with high concentration of L-alanine". In: *Scientific reports* 8.1 (2018), pp. 1–9.

CHAPTER 4

PREDICTION AND APPLICATION OF CULTURE MEDIA MEASUREMENT IN MSC MANUFACTURING

In chapter 2, I showed that cellular metabolites can be used to predict T cell suppression and IDO activity. In chapter 3, I moved one step further and showed that extracellular metabolites can also be used to predict composite functional score generated by T cell suppression and IDO activity by using only the first 3 day culture media. In this chapter, I will talk about the application of real time media monitoring in bioreactors. I will also show that batch effects can be corrected if we have reference materials across different studies.

4.1 The application of real time monitoring in cell manufacturing

4.1.1 Real time monitoring of culture media in bioreactor

Low amounts of MSCs in all tissues and insufficient isolated MSC quantity limit MSC clinical use. In order to gain desired therapeutic outcome, a dosage of $2x10^6$ cells/kg body weight is commonly given to the patients [1, 2]. The large scale of MSC expansion becomes a bottleneck of MSC manufacturing. MSCs can be expanded in vitro in laboratories using cell culture plates and flasks, however, this strategy is not feasible for cell manufacturing as the cell number needed is much higher [3]. Many types of commercial bioreactors are available for large-scale MSC expansion. The capacity ranges from 1.3L to 50L depending on bioreactor type [4, 5]. The cell expansion ratio is about 1.85-fold to 42.67-fold depending on the bioreactor and culture protocol used [6]. Despite different types and expansion ratio of bioreactors, there is no consensus on the choice of bioreactor, and it is essential to develop methods to monitor MSC growth and potency in bioreactors in real time.

Bioreactors can be used to provide clinically relevant cell numbers, while maintaining low cell passage numbers. These cultures can increase cell yields, direct cell differentiation, and increase cell secreted products including cytokines, metabolites and proteins. Application of real time media monitoring and validation of key metabolites during bioreactor mediated MSC culturing period are important for cell manufacturing. We collaborated with investigators at the school of engineering at University of Georgia to test the metabolites from spent culture media cultured with MSCs using different bioreactors.

The first generation bioreactor was designed and validated by Mantay M., et, al [7]. A second-generation system for MSC culturing under dynamic forces was designed by Dr.Cheryl Gomillion's group at University of Georgia. Compared to the first generation bioreactor, the second generation

can support both 2D culture via a monilayer insert and 3D culture via a microcarrier insert. It has the continuous perfusion flow through the system for cell culturing and achieved significant dynamic applied force. The extracellular metabolites were tested using NMR at the end of culture. This section is based on joint work with Daniel Christian Shah and Dr.Cheryl Gomillion in the school of engineering at the University of Georgia. Daniel Christian Shah cultured the MSCs and provided the media samples for analysis. I collected NMR metabolomics data. We performed the data analysis together. I participated in the study design with Dr.Cheryl Gomillion, Dr.Arthur S. Edison and Daniel Christian Shah.

Bone marrow derived MSCs, specifically, D1, Mus musculus cells (D1 CRL-124TM, ATCC® Manassas, VA) were used in this study. The cells were cultured in T-75 flask, 1^{st} generation and 2^{nd} generation bioreactor respectively. Researches have found that MSCs with low passage shows higher potency in immunomodulation function [8]. Thus, the low passages were used in this study. There are three replicates in each group. The cell yields and averaged cell yield normalized by growth area are presented in Figure [4] and Table [5] respectively. The 1^{st} generation bioreactor yields most cells in single unit area, and significantly increased compared to T-75 flask. There is no statistical differences in cell yields between first generation and second generation bioreactors.

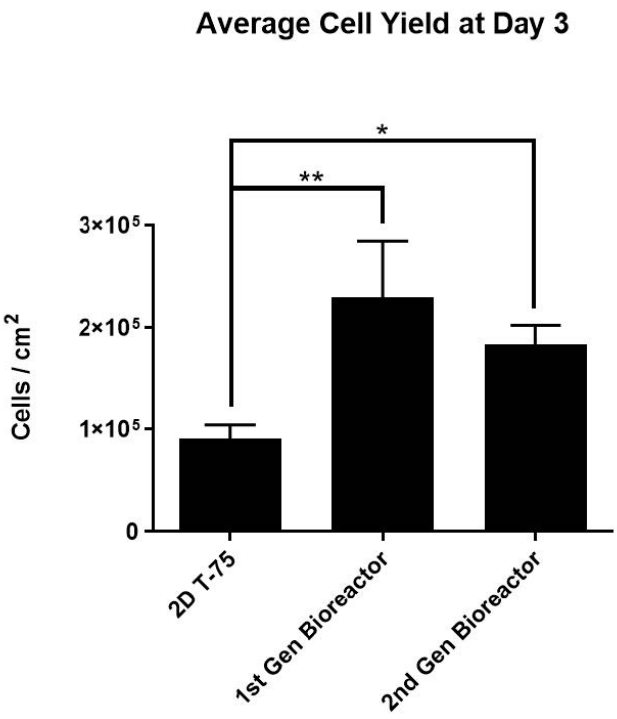


Figure 4.1.1: Cell yields after culture. The averaged cell yields were normalized by growth area. Statistical difference is presented with asterisks (* for $p \leq 0.05$ and ** for $p \leq 0.01$).

Table 4.1: Averaged cell yield normalized by growth area

(cells/cm^2)	2D T-75	1^{st} generation bioreactor	2^{nd} generation bioreactor
Trial 1	103,106.67	231,373.72	191,982.49
Trial 2	92,833.33	284,243.51	197,330.15
Trial 3	75,666.67	174,574.75	162,265.99
Average	90,535.56	230,063.99	183,859.54
Standard deviation	13,863.56	54,846.11	18,890.75

On the last day of culturing, the spent culture media as well as control media were collected from each system for NMR metabolomics analysis. Fifty-three individual peaks were binned from full NMR spectra and ANOVA was used to test the differences between different groups, and the results are shown in Figure 4.1.1 and 4.1.1. Cells consume most of the amino acids and secrete alanine into

media during culture. There are significant differences comparing T-75 flasks with first and second generation bioreactors. The second generation bioreactor consumed less amino acids and secreted less analine into culture media. These findings are consistent with previous chapter, where we found cells consuming valine and aspartate and secreting alanine during culture. We didn't perform T cell suppression and IDO assays in this study, which prevents correlations between metabolite levels with cell potency impossible. However, this type of study showed that it is feasible to monitor culture media during MSC culturing in real time using bioreactors. Furthermore, The closed loop of bioreactor system can integrate with benchtop NMR to apply to industrial setting.

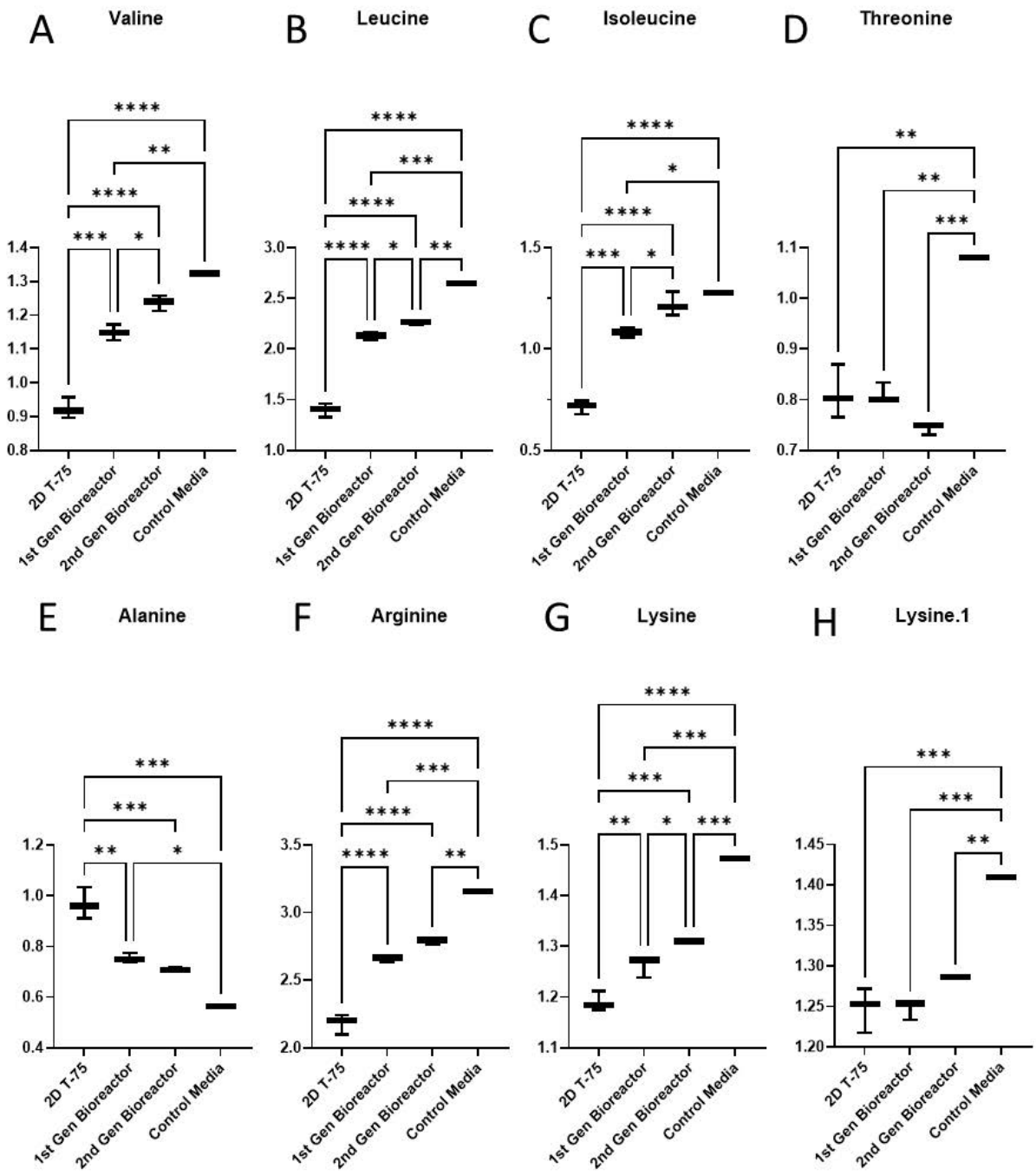


Figure 4.1.2: Amino acids and comparisons between different groups. Statistical difference is presented with asterisks (* for $p \leq 0.05$, ** for $p \leq 0.01$, *** for $p \leq 0.001$, and **** for $p \leq 0.0001$).

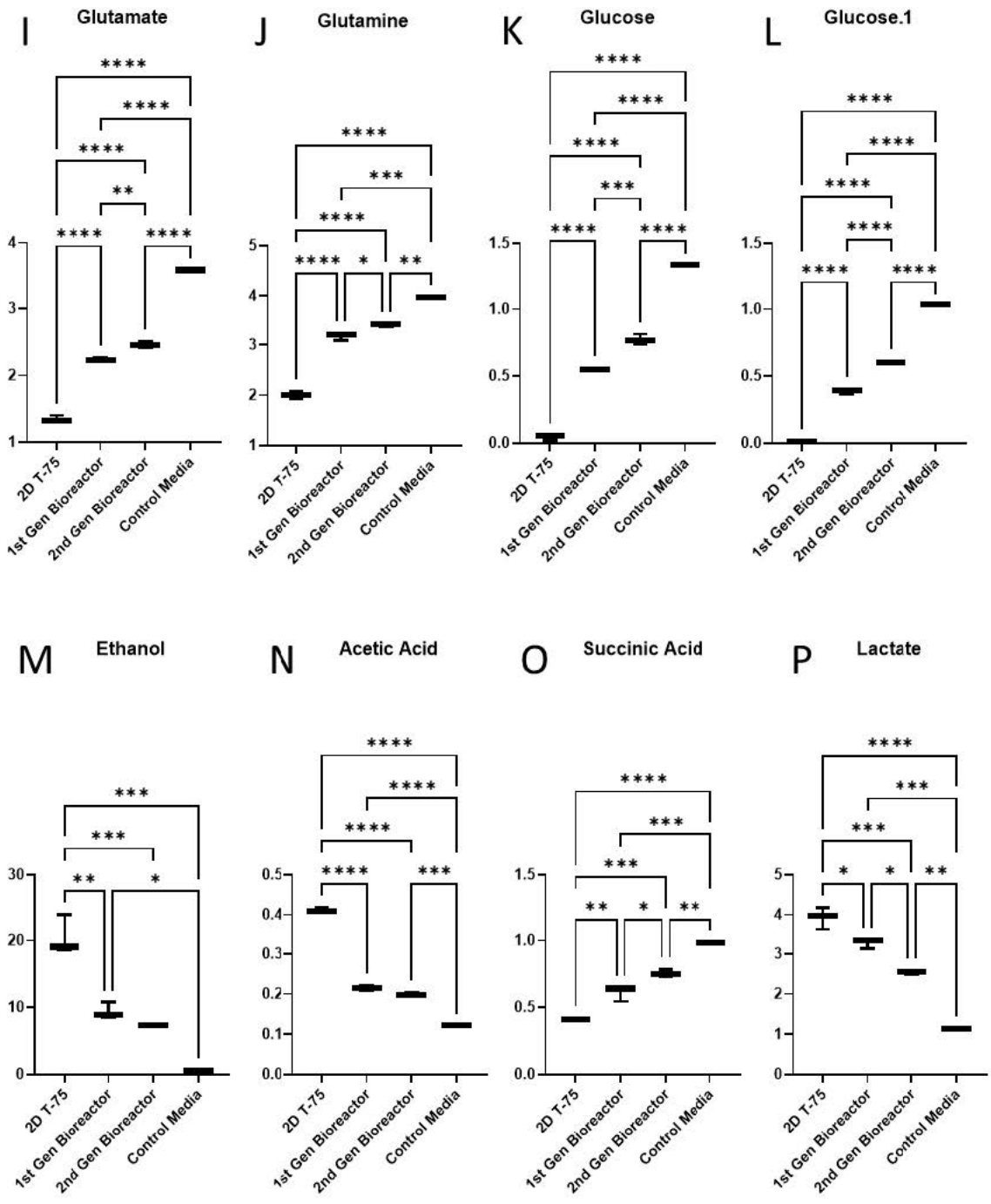


Figure 4.1.3: Metabolites and comparisons between different groups. Statistical difference is presented with asterisks (* for $p \leq 0.05$, ** for $p \leq 0.01$, *** for $p \leq 0.001$, and **** for $p \leq 0.0001$).

4.1.2 Application of real time monitoring with bench top NMR

In Chapter 3, we discretely sampled media each day and only obtained two or three samples in each time period. The sampling strategy limited our analysis to linear regressions, and nonlinearity was not considered. However, metabolites changes in media are not always linear [9]. Moreover, we analyzed the media samples using an expensive 800 MHz NMR, which is not practical in a biomanufacturing setting. However, lower magnetic field benchtop NMR instruments are now available, some with flow capabilities. In principle, one of these instruments could be added to a closed loop flow from a bioreactor and data could be collected continuously. We and others have shown the benefits of continuous monitoring of metabolism in monitoring the growth of microorganisms over time [10, 11]. Computational tools exist that can analyze continuous time-course NMR data [12], and these could be applied to continuous NMR monitoring of MSCs in biomanufacturing.

However, a bench top NMR operating at 80 MHz is much lower resolution and sensitivity than a research instrument operating at 800 MHz. To test the feasibility of using a benchtop system to monitor MSCs in real time, we simulated an NMR media spectrum at 80 MHz and 800 MHz and compared these to one of our experimental NMR spectra. We used a computational resource called GISSMO (guided ideographic spin system model optimization) for this simulation. The limitation is that we can only simulate identified metabolites, so for the unknown features from out data, we added compounds from GISSMO that would approximate the same overlap.

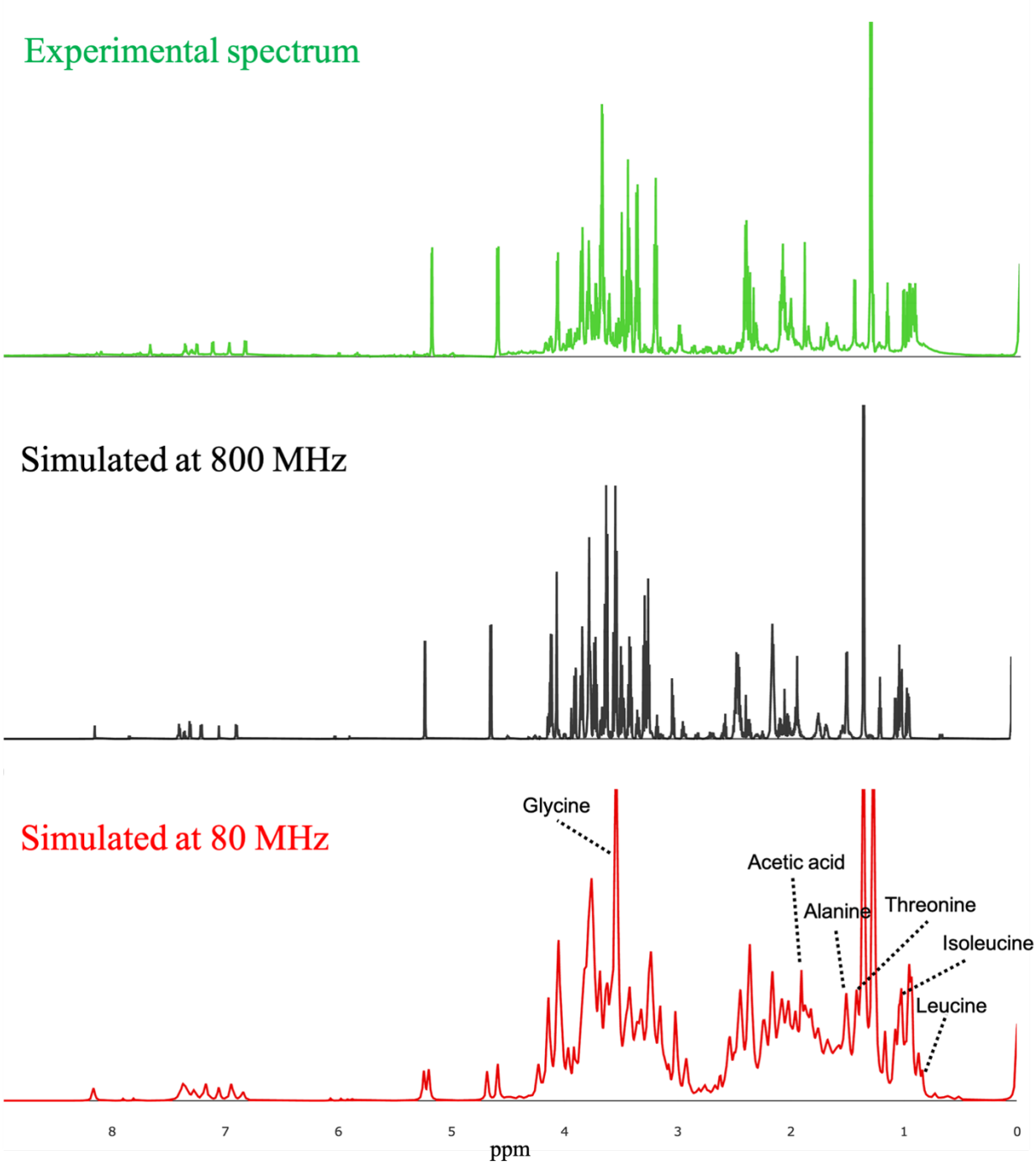


Figure 4.1.4: GISSMO simulation of experimental spectrum. The top panel is the experimental spectrum.

The middle panel is the spectrum simulated at 800 MHz containing all known metabolites in experiments and some approximated compounds. The bottom panel is the simulated spectrum at 80 MHz.

In GISSMO simulation process, we added 5 compounds to approximate the overlap and non-identified peaks in experimental samples. They are: 2-Hydroxyisocaproic acid, N-Acetyl-L-glutamic acid, Tyramine, and L-Anserine. The spectrum simulated at 800 MHz showed high similarity compared to experimental spectrum. Then the spectrum was simulated at 80MHz to approximate the benchtop NMR signals. The annotation of PMMs is labeled. The GISSMO simulation results suggest that our method can be applied to cell biomanufacturing setting with benchtop NMR and bioreactors. This real time monitoring will make prediction of MSC immunomodulation effects more accurate and efficient.

4.2 Prediction of functional composite score

This work is a collaboration with multiple centers across different institutions, including Marcus Center at Georgia Institution of Technology, Regenerative Center at the University of Georgia and Complex Carbohydrate Research Center at the University of Georgia. In this work, I participated in study design with Ty S. Maughon, Annie C. Bowles-Welch, Ross A. Marklein, Steven L. Stice, and Arthur S. Edison. I participated in sample preparation with Ying Xiang, data collection with Courtney Manning and Maxwell Colonna, and data processing as well as the data analysis. This is an ongoing project and I will show some preliminary results below in this chapter.

Culture media plays an essential role in MSC expansion and differentiation. After isolating from donors, MSCs are expanded in vitro for clinical use. One of many challenges is to determine the appropriate cell culture media. Multiple commercial MSC culture media are available for MSC expansion, cryopreservation and differentiation, including FBS (Fetal Bovine Serum)-contained media, serum-free media (SFM) and Xeno-free media (XFM). The most commonly used research MSC culture media contains FBS. FBS provides factors that are important for MSC proliferation,

including growth factors, adhesive proteins and hormones [13–15]. However, this type of media is only suitable for research instead of clinical application because it raises the risks of transferring animal factors to human [16]. SFM doesn't contain serum, instead, it contains a broad range of hormones, growth factors, proteins and ployamines derived from bovine or human sources [17]. XFM, on the other hand, is formulated using synthetic, recombinant or human-derived purified substances [18–20].

Even within these media types, different formulations show different culture outcomes. Ovine MSC (oMSC) has shown different morphology, cell proliferation rate and surface antigen expression in culture media with different FBS concentrations [21]. Evidence suggests adipose tissue-derived MSCs (ASCs) show different growth kinetics and proliferation rates with different media formulation [22]. One study showed that MSCs cultured with Dulbecco's Modified Eagle's Medium (D-MEM) and Mesenchymal Stem Cell Basal Medium (MSCBM; Cambrex®) express different endothelial and hematopoietic lineage markers [23]. The various media components and formulations make finding predictive markers generalized for different conditions difficult.

In order to address these issues, we aim to predict MSC functional composite score in different cell lines with different culture media using the model we developed previously, and later, these predictions can be validated by real data collected in the lab to evaluate our model.

4.2.1 The minimal batch effects across different studies

In order to compare results from different studies, I first need to show that there is minimal batch effects across different studies. We took the advantages of the high reproducibility of NMR and tested if this hypothesis stands by comparing identical samples with 2 very different NMR probes on different days.

Bone marrow derived MSCs, specifically, RB139 (Male, 25, RoosterBio) and RB147 (Male, 25, RoosterBio) were used to test this hypothesis. The cells were cultured in two different media types, they are laboratory generated αMEM and RoosterBio generated MSC culture media. Spent

culture media were collected at day 3. The NMR metabolomics data were collected with a 5-mm cyroprobe and a 5-mm room temperature probe, both operating at 800MHz using these culture media respectively. The PCA plot can clearly separate Roosterbio media from αMEM as shown in Figure 4.2.1.

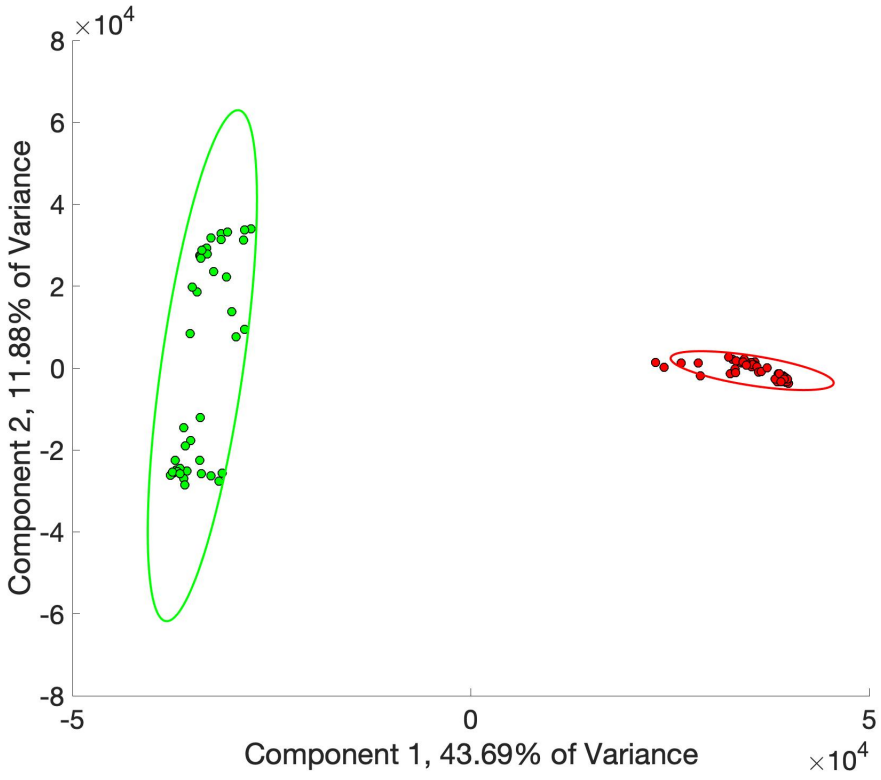


Figure 4.2.1: PCA plot of two media types. The green dots indicate αMEM media and the red dots indicate Roosterbio media.

The data were normalized to reference material DSS. Twenty five metabolites identified from the work presented in Chapter 3 were binned from whole spectra and the area under curve was used to calculate metabolite relative intensities respectively. Linear regression was then used to regress cryoprobe collected metabolite intensity against room temperature collected metabolite intensity.

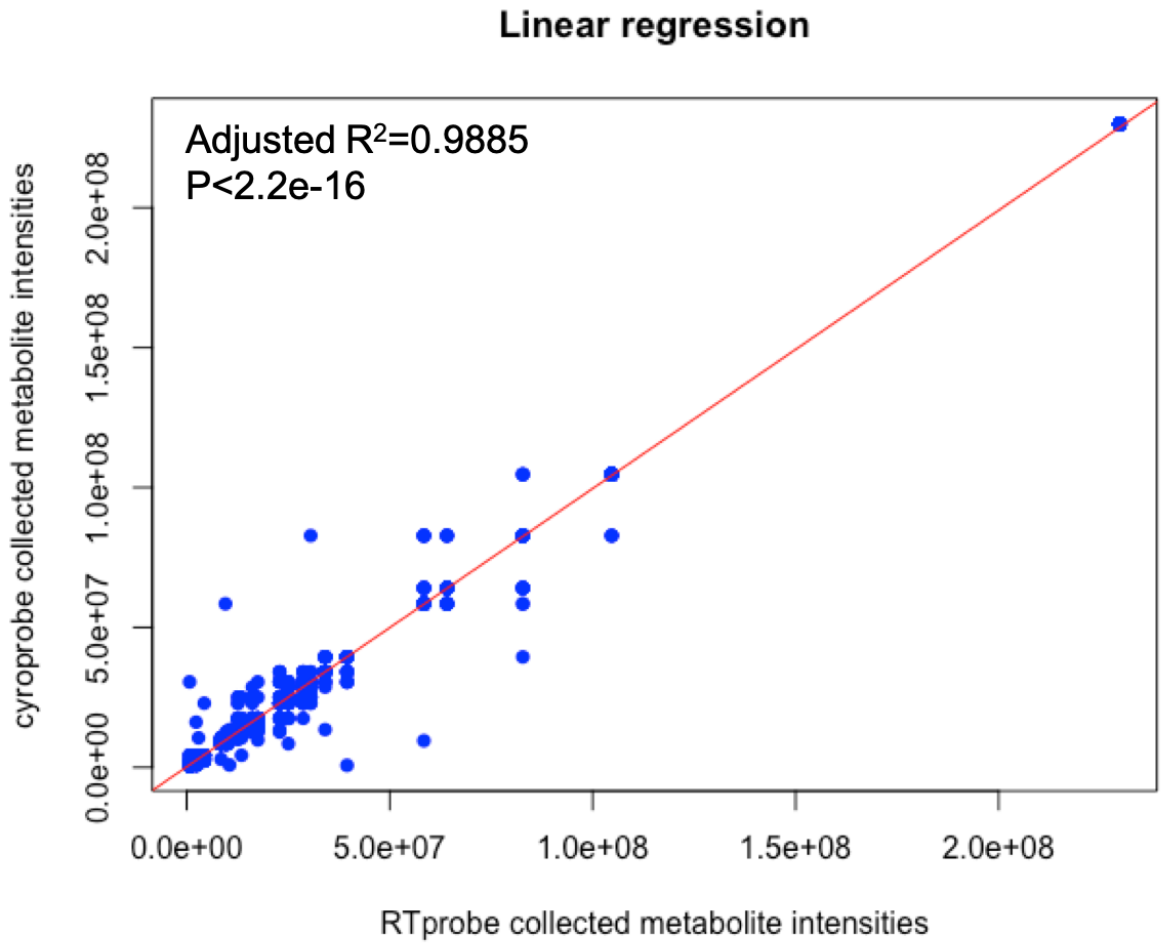


Figure 4.2.2: Linear regression of cyroprobe collected data against room temperature collected data.

The linear regression shows there are many points overlapped at the low intensity range. The coefficient value is 0.99 and intercept is 0.29. The high value of R^2 and low value of p value indicates this model has a good performance and indeed represent the data we generated. An easy conclusion can be draw from this model, that is the batch effects can be corrected across different studies if we have reference.

4.2.2 Future directions

In previous chapter, we developed a model to predict T cell suppression assay and IDO activity using the first three days of spent culture media. Presumably, with more data points added to this model, the model will evolve gradually and become more accurate. In the future, we aim to predict composite functional score using previous model. This is a collaborate project with Marcus Center at Georgia Institute of Technology, and the data has been collected for analysis.

Five different bone marrow derived MSCs were used. They are: RB139 (Male, 25, RoosterBio), RB174 (Male, 25, RoosterBio), RB177 (Male, 22, RoosterBio), RB179 (Male, 21, RoosterBio), RB183 (Feamle, 26, RoosterBio). Each cell line has 10 replicates for expansion, along with one control plate, which contains only media without cells in the plate. The media were sampled daily for continuous 9 days until the cells got 80% of the confluence. Sampled media were then used for extracellular metabolomics study using NMR, the process is shown in Figure 4.2.2.

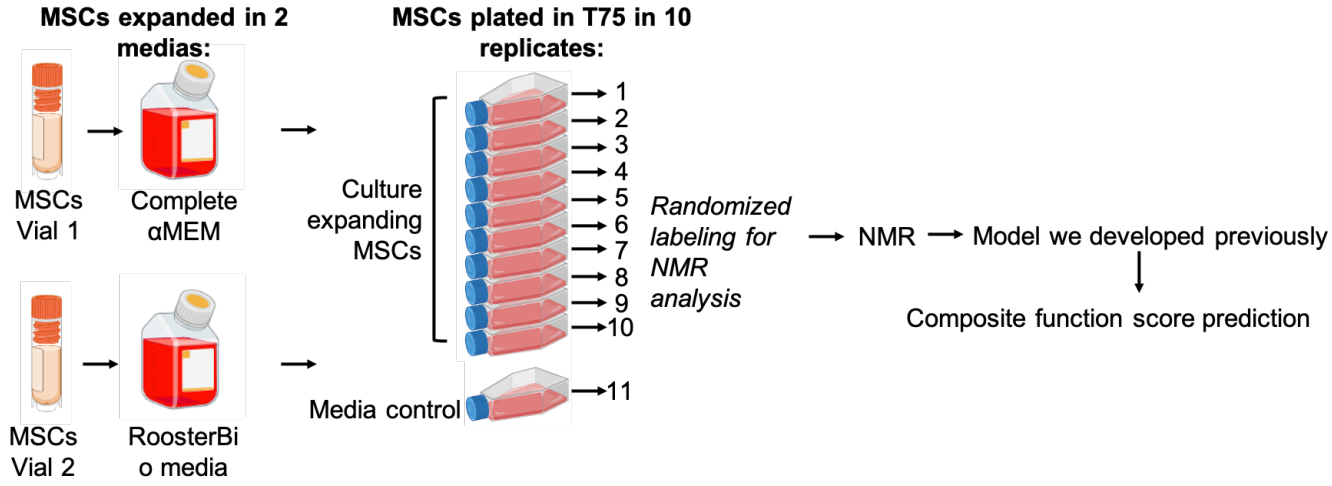


Figure 4.2.3: Study design to predict composite function score

The first three day culture media can be used to calculate the metabolite change, and the composite functioal score can be predicted using the model we developed previously. These scores can further

be validated by collecting real data in the lab. The validation of composite function scores can evaluate our model's accuracy. By adding more data into the model, our model can evolve and become more accurate. Different data will be collected and used to improve the model we generated in previous chapter as well as validate the predictive markers we identified. Table 4.2 shows the data we collected/are about to collect.

Table 4.2: Data collected/are about to collect. *: data already collected, -: data will be collected, /: data will not be collected.

MSC	Intracellular metabolomics	Extracellular metabolomics	T cell assay	IDO activity	Single cell RNA-seq
iMSC	*	-	*	*	/
RB71	*	-	*	*	/
RB182	*	-	*	*	/
RB175	*	-	-	-	/
RB183	*	-	-	-	*
RB174	*	-	-	-	*
RB179	*	-	-	-	*
RB177	*	-	-	-	*
RB139	*	-	-	-	*

The design of the experiment will lead us to validate the predictive markers we identified before as well as reveal some of the underlying mechanisms. For five of the cell lines (RB183, RB174, RB179, RB177 and RB139), some of the data were collected previously, including single cell RNA-seq, lipidomics/metabolomics. The daily extracellular metabolites as well as T cell suppression assay and IDO activity assay will also be measured. The single cell RNA seq will allow us to interrogate the cell subpopulation heterogeneity. This complete dataset would allow us to validate

predictive markers we identified, as well as identify some of the unknown predictive markers. The single cell RNA-seq can provide direct visualization of MSC heterogeneity at a single-cell level and uncovers subpopulations with distinct signature genes and signaling pathways [24]. By combining single cell RNA-seq with metabolomics, we will be able to identify some of the new metabolic pathways. The other dataset includes six cell lines, they are iMSC, RB71, RB182, RB175, RB183 and RB174. This list was build upon the previous study, where we only used iMSC, RB71 and RB182. We expanded the list and further collect the daily extracellular metabolites, T cell assay and IDO assay. With this set of data, we will also be able to compare to previous dataset and further validate our predictive markers.

BIBLIOGRAPHY

[1] Usha Nekanti et al. "Optimization and scale-up of Wharton's jelly-derived mesenchymal stem cells for clinical applications". In: *Stem cell research* 5.3 (2010), pp. 244–254.

[2] Sâmia R Caruso et al. "Growth and functional harvesting of human mesenchymal stromal cells cultured on a microcarrier-based system". In: *Biotechnology progress* 30.4 (2014), pp. 889–895.

[3] Muhammad Najib Fathi Bin Hassan et al. "Large-scale expansion of human mesenchymal stem cells". In: *Stem Cells International* 2020 (2020).

[4] Patrıcia Aparecida Tozetti et al. "Expansion strategies for human mesenchymal stromal cells culture under xeno-free conditions". In: *Biotechnology progress* 33.5 (2017), pp. 1358–1367.

[5] Marcos FQ Sousa et al. "Production of oncolytic adenovirus and human mesenchymal stem cells in a single-use, vertical-wheel bioreactor system: impact of bioreactor design on performance of microcarrier-based cell culture processes". In: *Biotechnology progress* 31.6 (2015), pp. 1600–1612.

[6] Muhammad Najib Fathi Bin Hassan et al. "Large-scale expansion of human mesenchymal stem cells". In: *Stem Cells International* 2020 (2020).

[7] Morgan Elizabeth Mantay. "Design and validation of a low-cost bioreactor system for therapeutic mesenchymal stromal cell expansion". PhD thesis. University of Georgia, 2019.

[8] Mandana Mohyeddin Bonab et al. "Aging of mesenchymal stem cell in vitro". In: *BMC cell biology* 7.1 (2006), pp. 1–7.

[9] Volker Behrends et al. "Time-resolved metabolic footprinting for nonlinear modeling of bacterial substrate utilization". In: *Applied and Environmental Microbiology* 75.8 (2009), pp. 2453–2463.

[10] Michael T Judge et al. "Continuous in vivo metabolism by NMR". In: *Frontiers in molecular biosciences* 6 (2019), p. 26.

[11] Arthur S Edison et al. "NMR: unique strengths that enhance modern metabolomics research". In: *Analytical Chemistry* 93.1 (2020), pp. 478–499.

[12] Yue Wu et al. “RTExtract: time-series NMR spectra quantification based on 3D surface ridge tracking”. In: *Bioinformatics* 36.20 (2020), pp. 5068–5075.

[13] Mark F Pittenger et al. “Mesenchymal stem cell perspective: cell biology to clinical progress”. In: *NPJ Regenerative medicine* 4.1 (2019), pp. 1–15.

[14] Tae-Hoon Shin et al. “Mesenchymal stem cell therapy for inflammatory skin diseases: clinical potential and mode of action”. In: *International journal of molecular sciences* 18.2 (2017), p. 244.

[15] Li-Tzu Wang et al. “Human mesenchymal stem cells (MSCs) for treatment towards immune-and inflammation-mediated diseases: review of current clinical trials”. In: *Journal of biomedical science* 23.1 (2016), pp. 1–13.

[16] Ana CC Paula et al. “Human adipose tissue-derived stem cells cultured in xeno-free culture condition enhance c-MYC expression increasing proliferation but bypassing spontaneous cell transformation”. In: *Stem cell research & therapy* 6.1 (2015), pp. 1–19.

[17] S Gottipamula et al. “Serum-free media for the production of human mesenchymal stromal cells: a review”. In: *Cell proliferation* 46.6 (2013), pp. 608–627.

[18] Mark F Pittenger et al. “Mesenchymal stem cell perspective: cell biology to clinical progress”. In: *NPJ Regenerative medicine* 4.1 (2019), pp. 1–15.

[19] Tae-Hoon Shin et al. “Mesenchymal stem cell therapy for inflammatory skin diseases: clinical potential and mode of action”. In: *International journal of molecular sciences* 18.2 (2017), p. 244.

[20] M Cimino et al. “Xeno-free strategies for safe human mesenchymal stem/stromal cell expansion: supplements and coatings”. In: *Stem cells international* 2017 (2017).

[21] Carina Adamzyk et al. “Different culture media affect proliferation, surface epitope expression, and differentiation of ovine MSC”. In: *Stem cells international* 2013 (2013).

[22] Justyna Czapla et al. “The effect of culture media on large-scale expansion and characteristic of adipose tissue-derived mesenchymal stromal cells”. In: *Stem cell research & therapy* 10.1 (2019), pp. 1–11.

[23] Sayaka Nakamura et al. “Culture medium study of human mesenchymal stem cells for practical use of tissue engineering and regenerative medicine”. In: *Bio-medical materials and engineering* 18.3 (2008), pp. 129–136.

[24] Wenduo Gu et al. “Single-cell RNA-sequencing and metabolomics analyses reveal the contribution of perivascular adipose tissue stem cells to vascular remodeling”. In: *Arteriosclerosis, thrombosis, and vascular biology* 39.10 (2019), pp. 2049–2066.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Achievements of goals

In this dissertation, I have made original contributions towards predicting MSC anti-inflammatory effects using metabolites by applying traditional statistics and machine learning methods. The studies presented here specifically provide, for the first time, the cell metabolites can be used as surrogates of T cell suppression and IDO assays to predict MSC anti-inflammatory effects. I identified several important predictive markers including NN-dimethylglycine,myo-inositol, phsophocreatine, and asparagine. In addition, machine learning methods including random forest, KNN, logistic regression and SVM were also used to validate the metabolites we found predictive using PLSR model (data presented in Appendix). I also showed that the first three days culture media can be used to predict MSC anti-inflammatory effects. I have identified these markers including threonine, leucine and alanine. In chapter 4, I talked about the application of real time culture media monitoring in cell manufacturing. In addition, I predicted the composite function score using new cell line first three day culture media metabolomics profile.

5.2 Limitation of projects

In this thesis, we performed analysis based on metabolic change rate in three days culture period. Our model for metabolic change rate is linear regression, however, most of the Longitudinal metabolic change is not linear [1]. The noncontinuous study design limited our analysis to linear. In the future, the integration of benchtop NMR with bioreactors can make real time monitoring feasible. in this case, the metabolic change can be estimated more accurate with nonlinear functions. The other limitation of these study is that we only used nine data points for modeling and it is very easy to have overfitting issue. In order to eliminate the overfitting problem, we could add more data points into the modeling to make the prediction more accurate as well.

In this thesis, we used open source available databases to annotate metabolites, including BMRB, HMDB and COLMARm (COLMARm). These databases contains tons of metabolites structure in NMR, and the number is still increasing. However, there are still metabolites that not able to be identified using these databases. In our study, we found several unknown structures that are important in predicting MSC function capacity. These unknowns are not involved in the databases. To identify those unknowns, multiple steps need to be taken. Isabel Garcia-Perez et al. published a paper talked about the detailed steps to identify unknowns in NMR metabolomics study [2]. Basically, this requires the combination of statistical tools including STOCSY (statistical total correlation spectroscopy), STORM (subset optimization by reference matching) and RED-STORM (resolution-enhanced STORM), along with different techniques including 1D and 2D NMR, LC-NMR-MS and Liquid chromatography (LC)-solid-phase extraction (SPE)-nuclear magnetic resonance (NMR)-mass spectrometry (MS) coupling (LC-SPE-NMR-MS). The combination of these methods would allow one to elucidate the structure of unknown metabolites. In this study, we didn't identify those unknowns but this is certainly a good future direction to work on. By identifying those metabolites, we may be able to find some new pathways that involves in the regulating of cellular metabolites in MSCs.

5.3 Future directions

The observations and findings presented here provide the guidance to MSC cell manufacturing. I highlight some avenues for future investigation in light of the presented novel studies.

5.3.1 Further validation of predictive markers

In this dissertation, we thoroughly investigated the predictive markers in mostly bone-marrow derived MSCs from different donors. Other tissue derived MSCs have not been studied here. Although bone marrow derived MSCs are most commonly used in clinical trials, other alternatives are currently under investigations, including adipose tissue derived MSCs and skeletal muscles derived MSCs [3, 4]. Studies have shown that different tissue derived MSCs have different immunosuppression effects and express different markers [5, 6]. Adipose derived MSCs show a higher level of secretion of cytokines including interleukin-6 and transforming growth factor-β1, and this correlates with higher metabolic activity [7]. The expression of CD146 marker is decreased in adipose tissue derived MSCs and skeletal muscles derived MSCs but maintained in bone marrow MSCs (BM-MSCs) [8]. This evidence indicates MSCs from different tissues express different biological functions and show different immunosuppression effects. Thus, generalizing predictive markers across different tissue types is essential for next step. By building models to predict MSC functionality using mixed tissue type-derived MSCs, one can find the generalized predictive markers. In addition, culture MSCs in different media types and compare the immunomodulation functions could find the optimal media for MSCs culture.

Although the predictive markers can be used to predict MSC immunosuppression capacity, whether altering these markers will cause MSC immunomodulation effects change is still not addressed. Therefore, altering the key components in culture media and measuring MSC immunosuppression effects is also an important direction that needs to be investigated.

5.3.2 The investigations of MSC exosomes

Exosomes are nanosized (30-150nm) extracellular vesicles (EVs) that are involved in intercellular communication system [9]. The scientific field initially considered exosomes as cell waste. Recently, they found exosomes are functional vesicles that carry a complex cargo of proteins [10], lipids [11], and nucleic aids [12, 13], and are able to deliver these materials to target cells, thus play an important role in cell-cell communication. Exosome surface markers and carried cargoes have been studied as potential biomarkers to diagnose different diseases, including ovarian cancer, preeclamptic pregnancies and metabolic diseases [14–16]. Exosomes have also been explored as nanodelivery systems to transfer small interfering RNAs across brain-blood barrier [17–19].

MSC-secreted exosomes have shown contribute to its differentiation functions [20, 21], thus have been widely used in regenerative biomedicine, including treating cardiovascular disease [22], kidney injury [23], and fibrotic liver disease [24]. MSC derived exosomes also have essential roles in anti-inflammatory functions. Teng et,al. found MSC derived exosomes restrain the inflammation response by inhibiting T-cell proliferation in vitro [25]. Another study showed that an MSC exosomes treated group expressed less inflammatory cells compared to a control group. MSC derived exosomes also attenuate the pro-inflammatory macrophages [26]. Comparing to MSC therapy, exosomes show potentials and advantages in regnerative biomedicine. They show increased potency [27], greater consistency [28] and lower cost [29]. Thus, the MSC derived exosomes are worth investigating to find alternatives for cell therapy.

In previous chapters, we used 14,000g to centrifuge samples and get the supernatant for analysis. This force allows us to discard the cell pellets and cell debris. However, this method can not separate exosomes from media. Thus our media analysis contained exosomes materials.

To date, most studies of MSC exosomes have focused mainly on the materials carried incluing RNA and proteins. Showalter et, al. first showed that primary MSC derived exosomes contain numerous primary metabolites involving in amino acids and carbohydrates metabolic networks and nucleodises biosynthetic pathways. Limited research investigated the relations between MSC de-

rived exosomes contained metabolites and MSC anti-inflammatory effects. Therefore, investigating MSC derived exosomes and identifying key metabolites in exosomes that can predict their function are crucial for MSC biomedicine field.

5.3.3 The investigations of underlying mechanisms

In this dissertation, we worked on the metabolic pathways to reveal some of the mechanisms of how MSC cultured media metabolites affects MSC final cellular metabolites level. However, there are missing linkages between key metabolites we identified, for example, valine and phosphocreatine. Although our analysis showed that they have negative correlation (-0.91), there is no database available pathways that link these two metabolites together. This indicates some unknown pathways may involve in regulating metabolites in MSCs. Other techniques can be used to explore this part including genomics, transcriptomics, and proteomics. Systems biology method is the method integrate data from diverse biological datasets including genes, RNAs, proteins, metabolites and other factors. Such an approach makes understanding phenotype mechanism easier [30]. An integrated systems biology approach can help to identify regulatory networks and metabolic pathways [31]. By exploring new metabolic pathways and linking DNA, RNA with metabolites and proteins, some underlying mechanisms of culture media metabolites regulating cellular metabolites may be revealed.

BIBLIOGRAPHY

[1] Carmelinda Ruggiero et al. "High basal metabolic rate is a risk factor for mortality: the Baltimore Longitudinal Study of Aging". In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 63.7 (2008), pp. 698–706.

[2] Isabel Garcia-Perez et al. "Identifying unknown metabolites using NMR-based metabolic profiling techniques". In: *Nature Protocols* 15.8 (2020), pp. 2538–2567.

[3] Jeong Chan Ra et al. "Safety of intravenous infusion of human adipose tissue-derived mesenchymal stem cells in animals and humans". In: *Stem cells and development* 20.8 (2011), pp. 1297–1308.

[4] B Fang et al. "Favorable response to human adipose tissue-derived mesenchymal stem cells in steroid-refractory acute graft-versus-host disease". In: *Transplantation proceedings*. Vol. 39. 10. Elsevier. 2007, pp. 3358–3362.

[5] Li Hu et al. "Side-by-side comparison of the biological characteristics of human umbilical cord and adipose tissue-derived mesenchymal stem cells". In: *BioMed research international* 2013 (2013).

[6] Christian Behm Andrukhov, Alice Blufstein, and Xiaohui Rausch-Fan. "Immunomodulatory properties of dental tissue-derived mesenchymal stem cells: implication in disease and tissue regeneration". In: *World journal of stem cells* 11.9 (2019), p. 604.

[7] Sara M Melief et al. "Adipose tissue-derived multipotent stromal cells have a higher immunomodulatory capacity than their bone marrow-derived counterparts". In: *Stem cells translational medicine* 2.6 (2013), pp. 455–463.

[8] Urszula Kozlowska et al. "Similarities and differences between mesenchymal stem/progenitor cells derived from various human tissues". In: *World journal of stem cells* 11.6 (2019), p. 347.

[9] Tek N Lamichhane et al. "Emerging roles for extracellular vesicles in tissue engineering and regenerative medicine". In: *Tissue Engineering Part B: Reviews* 21.1 (2015), pp. 45–54.

[10] Richard J Simpson et al. "Exosomes: proteomic insights and diagnostic potential". In: *Expert review of proteomics* 6.3 (2009), pp. 267–283.

[11] Michel Vidal et al. “Asymmetric distribution of phospholipids in the membrane of vesicles released during in vitro maturation of guinea pig reticulocytes: evidence precluding a role for “aminophospholipid translocase””. In: *Journal of cellular physiology* 140.3 (1989), pp. 455–462.

[12] Hadi Valadi et al. “Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells”. In: *Nature cell biology* 9.6 (2007), pp. 654–659.

[13] Anders Waldenström et al. “Cardiomyocyte microvesicles contain DNA/RNA and convey biological messages to target cells”. In: *PloS one* 7.4 (2012), e34653.

[14] Kalpana Deepa Priya Dorayappan et al. “The biological significance and clinical applications of exosomes in ovarian cancer”. In: *Gynecologic oncology* 142.1 (2016), pp. 199–205.

[15] Carlos A Escudero et al. “Role of extracellular vesicles and microRNAs on dysfunctional angiogenesis during preeclamptic pregnancies”. In: *Frontiers in physiology* 7 (2016), p. 98.

[16] Günter Müller. “Microvesicles/exosomes as potential novel biomarkers of metabolic diseases”. In: *Diabetes, metabolic syndrome and obesity: targets and therapy* 5 (2012), p. 247.

[17] Dongmei Sun et al. “A novel nanoparticle drug delivery system: the anti-inflammatory activity of curcumin is enhanced when encapsulated in exosomes”. In: *Molecular therapy* 18.9 (2010), pp. 1606–1614.

[18] Lydia Alvarez-Erviti et al. “Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes”. In: *Nature biotechnology* 29.4 (2011), pp. 341–345.

[19] Anuradha Kalani, Alka Tyagi, and Neetu Tyagi. “Exosomes: mediators of neurodegeneration, neuroprotection and therapeutics”. In: *Molecular neurobiology* 49.1 (2014), pp. 590–600.

[20] Fábio G Teixeira et al. “Mesenchymal stem cells secretome: a new paradigm for central nervous system regeneration?” In: *Cellular and Molecular Life Sciences* 70.20 (2013), pp. 3871–3882.

[21] Ahmed M Katsha et al. “Paracrine factors of multipotent stromal cells ameliorate lung injury in an elastase-induced emphysema model”. In: *Molecular Therapy* 19.1 (2011), pp. 196–203.

[22] Ruenn Chai Lai et al. “Exosome secreted by MSC reduces myocardial ischemia/reperfusion injury”. In: *Stem cell research* 4.3 (2010), pp. 214–222.

[23] Zhen-zhen Jiang et al. “Exosomes secreted by human urine-derived stem cells could prevent kidney complications from type I diabetes in rats”. In: *Stem cell research & therapy* 7.1 (2016), pp. 1–13.

[24] Tingfen Li et al. “Exosomes derived from human umbilical cord mesenchymal stem cells alleviate liver fibrosis”. In: *Stem cells and development* 22.6 (2013), pp. 845–854.

[25] Xiaomei Teng et al. “Mesenchymal stem cell-derived exosomes improve the microenvironment of infarcted myocardium contributing to angiogenesis and anti-inflammation”. In: *Cellular Physiology and Biochemistry* 37.6 (2015), pp. 2415–2424.

[26] Xuan Sun et al. “Intravenous mesenchymal stem cell-derived exosomes ameliorate myocardial inflammation in the dilated cardiomyopathy”. In: *Biochemical and biophysical research communications* 503.4 (2018), pp. 2611–2618.

[27] Reka Agnes Haraszti et al. “Exosomes produced from 3D cultures of MSCs by tangential flow filtration show higher yield and improved activity”. In: *Molecular Therapy* 26.12 (2018), pp. 2838–2847.

[28] Ruenn Chai Lai et al. “Exosome secreted by MSC reduces myocardial ischemia/reperfusion injury”. In: *Stem cell research* 4.3 (2010), pp. 214–222.

[29] Zhan-Jun Ma et al. “Mesenchymal stem cell-derived exosomes: Toward cell-free therapeutic strategies in regenerative medicine”. In: *World journal of stem cells* 12.8 (2020), p. 814.

[30] James T Robinson et al. “Integrative genomics viewer”. In: *Nature biotechnology* 29.1 (2011), pp. 24–26.

[31] Amith S Maroli et al. “Omics in weed science: A perspective from genomics, transcriptomics, and metabolomics approaches”. In: *Weed Science* 66.6 (2018), pp. 681–695.

APPENDIX A

MACHINE LEARNING METHODS TO VALIDATE PREDICTIVE MARKERS

A.1 Machine learning methods

The K-means clustering was used to cluster IDO assay and T cell suppression assay results into 2 groups- low potency group and high potency group. Based on the clustering result, the original data was labelled as high potency group and low potency group.

Logistic regression (LG), random forest (RF), K nearest neighbor (KNN) and support vector machine (SVM) were used to classify samples. Seventy percent of the samples were used as training data and 30% of the samples were left as testing data. Hyperparameter tuning was used in training process. Python scikit-learn pacakges were used in training and testing process.

Parameters setting using hyperparameter tuning and grid searching:

Table A.1: Logistic regression parameter tuning

Parameter	Value
C_value	100
penalty	12
solver	liblinear

Table A.2: Random forest parameter tuning

Parameter	Value
n_estimators	100
min_samples_split	2
min_samples_leaf	1
max_features	sqrt
max_depth	20
bootstrap	True

Table A.3: K nearest neighbor parameter tuning

Parameter	Value
leaf_size	1
p	1
n_neighbors	1

Table A.4: SVM parameter tuning

Parameter	Value
C_value	100
gamma	0.1
kernel	rbf

A.2 Model evaluation metrics

The following formulas were used to evaluate model performance. Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative:

Accuracy measures the overall percentage of correct prediction:

$$\text{Accuracy } = \frac{TP+TN}{TP+TN+FP+FN} \tag{A.1}$$

Precision measures the percentage of correct prediction out of the total high potency samples:

$$\text{Precision } = \frac{TP}{TP+FP} \tag{A.2}$$

Recall measures the percentage of correct prediction out of total low potency samples:

$$\text{Recall } = \frac{TP}{TP+FN} \tag{A.3}$$

A.3 Results

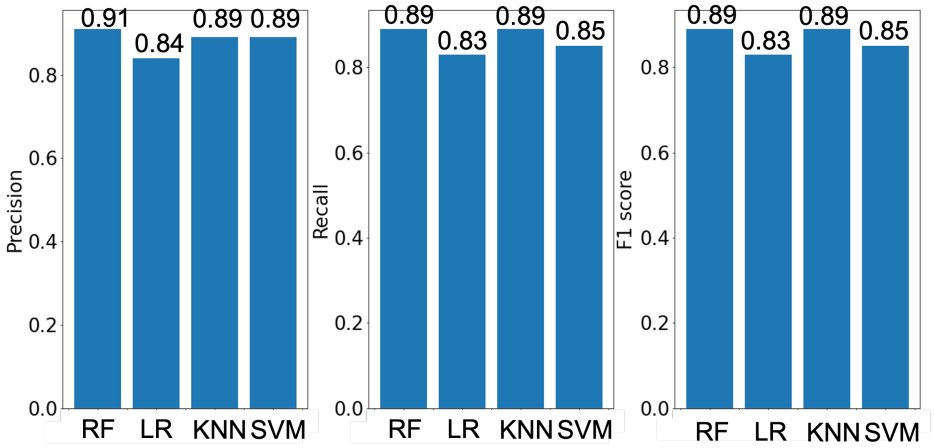


Figure A.3.1: Algorithm results comparison

By comparing the results, Random forest performed best in these 4 algorithms. The importance of features were plotted in Figure A.3. ukN-2, NN-dimethylglycine and myo-inositol are still top metabolites that distinguish high potency group from low potency group, which further validated our results.

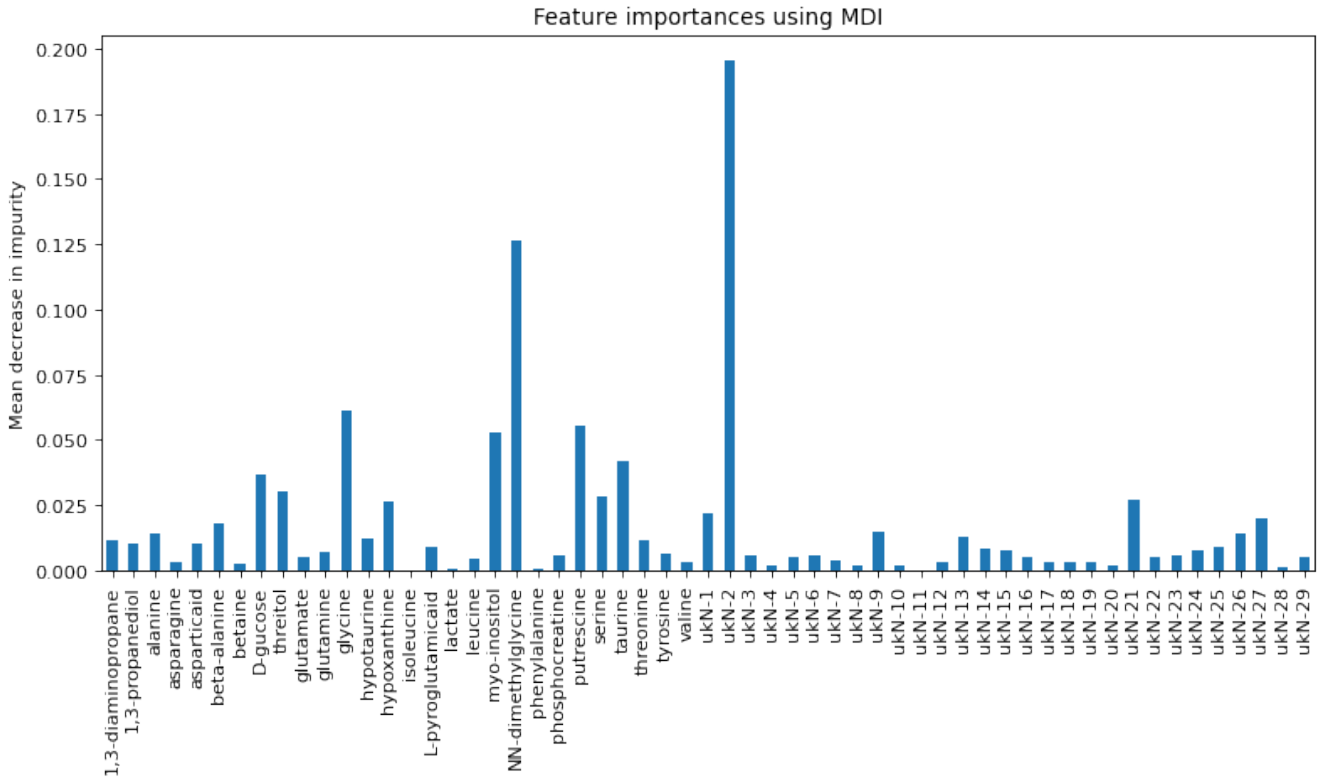


Figure A.3.2: Random forest feature importance

APPENDIX B

SUPPLEMENTARY DATA FOR CHAPTER 2

Table B.1: NMR averaged regression against PBMC donor1

Metabolite CD4	**P value**	**Metabolite CD8**	**P value**
myo-inositol	0.0005	myo-inositol	0.0005
ukN-2	0.0008	ukN-2	0.002
glycine	0.0016	glycine	0.0024
asparagine	0.0021	asparagine	0.0026
NN-dimethylglycine	0.0028	NN-dimethylglycine	0.0027
phosphocreatine	0.0043	phosphocreatine	0.0068
putrescine	0.0061	putrescine	0.0077
ukN-26	0.0083	ukN-26	0.0106
threonine	0.0102	threonine	0.0129
taurine	0.0124	taurine	0.0133
ukN-21	0.0133	ukN-21	0.0159
hypoxanthine	0.0162	hypoxanthine	0.0177

Table B.1 continued from previous page

serine	0.0169	serine	0.0193
hypotaurine	0.0205	hypotaurine	0.023
ukN-22	0.0219	ukN-22	0.0259
beta-alanine	0.0243	beta-alanine	0.0293
alanine	0.0304	alanine	0.0389
ukN-25	0.035	ukN-25	0.0434
ukN-4	0.0389	ukN-4	0.0484
glutamine	0.039	glutamine	0.0496
ukN-28	0.0397	ukN-28	0.0516
ukN-5	0.0422	ukN-5	0.0524
glutamate	0.044	ukN-10	0.055
ukN-10	0.0442	1,3-diaminopropane	0.0554
1,3-diaminopropane	0.0455	glutamate	0.0572
L-pyroglutamicaid	0.048	L-pyroglutamicaid	0.0597
ukN-29	0.0638	ukN-29	0.0706
ukN-23	0.0699	ukN-1	0.0736
ukN-16	0.0748	ukN-23	0.078
ukN-1	0.0751	phenylalanine	0.0929
ukN-27	0.0821	ukN-16	0.094
phenylalanine	0.0895	ukN-27	0.1035
ukN-19	0.0916	valine	0.1206
lactate	0.1163	lactate	0.1265
ukN-14	0.1316	ukN-14	0.1298
valine	0.1328	ukN-19	0.1357
ukN-9	0.1499	leucine	0.1464

Table B.1 continued from previous page

ukN-12	0.1519	tyrosine	0.1779
leucine	0.156	ukN-9	0.1824
ukN-6	0.1612	ukN-12	0.1857
ukN-17	0.1686	ukN-6	0.1993
ukN-24	0.1738	ukN-17	0.2018
tyrosine	0.1816	ukN-24	0.2047
ukN-15	0.2037	ukN-15	0.2246
ukN-7	0.2288	ukN-7	0.2586
ukN-8	0.245	ukN-8	0.2823
isoleucine	0.4121	isoleucine	0.3892
asparticaid	0.4647	asparticaid	0.5197
D-glucose	0.5637	ukN-3	0.5474
ukN-18	0.5793	ukN-18	0.5728
threitol	0.5886	betaine	0.616
betaine	0.5911	D-glucose	0.6228
ukN-13	0.6154	threitol	0.6509
ukN-3	0.6171	ukN-13	0.7216
1,3-propanediol	0.7056	1,3-propanediol	0.7567
ukN-20	0.9493	ukN-20	0.9817
ukN-11	0.9659	ukN-11	0.9898

Table B.2: NMR averaged regression against PBMC donor2

Metabolite CD4	**P Value**	**R Squared**	**Metabolite CD8**	**P value**	**R squared**
ukN-1	0.0017	0.7775	ukN-2	0.0002	0.8823
ukN-2	0.003	0.7387	myo-inositol	0.001	0.8086
myo-inositol	0.0238	0.5415	glycine	0.0019	0.769
ukN-14	0.0288	0.518	ukN-26	0.0019	0.7684
phosphocreatine	0.0359	0.4896	NN-dimethylglycine	0.003	0.7384
NN-dimethylglycine	0.0394	0.4769	phosphocreatine	0.0033	0.7311
serine	0.0434	0.4639	taurine	0.0066	0.6757
valine	0.0546	0.4315	ukN-21	0.0076	0.6619
glycine	0.0612	0.4148	hypoxanthine	0.0106	0.6303
ukN-26	0.0621	0.4126	ukN-10	0.0148	0.5962
asparagine	0.0963	0.3449	hypotaurine	0.0152	0.5932
ukN-21	0.099	0.3405	putrescine	0.0173	0.5788
taurine	0.1136	0.3182	ukN-22	0.0209	0.5568
ukN-10	0.1646	0.256	ukN-23	0.0222	0.5497
hypoxanthine	0.1662	0.2544	beta-alanine	0.0225	0.5485
leucine	0.1692	0.2513	ukN-29	0.0244	0.5387
ukN-23	0.1697	0.2508	threonine	0.0264	0.529
ukN-29	0.1925	0.229	ukN-1	0.0282	0.5207
putrescine	0.1935	0.2281	asparagine	0.029	0.517
phenylalanine	0.1941	0.2276	ukN-5	0.0353	0.4918
hypotaurine	0.1952	0.2266	ukN-28	0.0379	0.4824
ukN-22	0.1993	0.223	L-pyroglutamicaid	0.0472	0.4523
tyrosine	0.203	0.2198	ukN-25	0.0482	0.4493

Table B.2 continued from previous page

threonine	0.2188	0.2068	alanine	0.0535	0.4343
ukN-19	0.229	0.1988	glutamine	0.0542	0.4326
ukN-11	0.2329	0.1958	1,3-diaminopropane	0.0565	0.4264
beta-alanine	0.2331	0.1957	glutamate	0.0579	0.423
isoleucine	0.2808	0.1632	ukN-4	0.0662	0.4032
ukN-28	0.2819	0.1625	ukN-12	0.0739	0.3865
ukN-5	0.2953	0.1545	ukN-16	0.08	0.3741
glutamate	0.3073	0.1476	ukN-27	0.0831	0.3684
ukN-3	0.3233	0.1389	ukN-14	0.0916	0.3529
glutamine	0.3276	0.1366	serine	0.103	0.3343
ukN-12	0.3297	0.1355	ukN-6	0.1639	0.2567
alanine	0.3416	0.1294	ukN-9	0.1693	0.2512
L-pyroglutamicaid	0.346	0.1273	ukN-8	0.1873	0.2337
ukN-25	0.358	0.1215	ukN-7	0.2074	0.2161
ukN-4	0.3705	0.1157	phenylalanine	0.2148	0.21
1,3-diaminopropane	0.3778	0.1124	ukN-24	0.2166	0.2085
ukN-16	0.3894	0.1073	ukN-15	0.2187	0.2068
ukN-27	0.4364	0.08869	lactate	0.2372	0.1927
lactate	0.5603	0.05068	valine	0.2482	0.1847
ukN-6	0.597	0.04196	ukN-17	0.2693	0.1705
ukN-9	0.6002	0.04125	ukN-19	0.2903	0.1575
ukN-8	0.6771	0.02625	leucine	0.3623	0.1194
ukN-15	0.7016	0.02227	asparticaid	0.3624	0.1194
betaine	0.7034	0.02199	tyrosine	0.3715	0.1152
ukN-24	0.7233	0.01904	ukN-18	0.4587	0.08075

Table B.2 continued from previous page

ukN-17	0.7266	0.01858	D-glucose	0.4985	0.06784
ukN-7	0.7817	0.0117	threitol	0.5039	0.0662
1,3-propanediol	0.7913	0.01068	1,3-propanediol	0.5208	0.06127
threitol	0.8277	0.00724	ukN-13	0.536	0.05705
ukN-18	0.831	0.00696	betaine	0.6191	0.0372
D-glucose	0.8347	0.00666	isoleucine	0.6632	0.02866
ukN-20	0.8399	0.006242	ukN-11	0.6817	0.02547
ukN-13	0.9032	0.002267	ukN-20	0.7014	0.02229
asparticaid	0.9916	0.00001691	ukN-3	0.8212	0.007806

Table B.3: Top 50 MS metabolites regression against PBMC donor 1 CD4+ T cells

Metabolite	**P Value**	**R Squared**
1115	0.0002	0.8758
1201	0.0002	0.8703
1349	0.0002	0.8802
954	0.0003	0.8623
1209	0.0004	0.8455
183	0.0005	0.8393
1134	0.0005	0.8388
97	0.0008	0.8166
101	0.0008	0.8183
1190	0.0008	0.8164
1107	0.001	0.8073
1352	0.001	0.8069

93	0.0011	0.8041
121	0.0012	0.7948
89	0.0015	0.7854
913	0.0015	0.7837
1317	0.0016	0.7781
1187	0.0017	0.7758
1205	0.0018	0.7732
1186	0.0019	0.769
1098	0.002	0.7657
173	0.0021	0.7623
934	0.0021	0.7638
1363	0.0024	0.7531
332	0.0025	0.75
708	0.0025	0.7505
88	0.0027	0.7453
274	0.0027	0.7462
693	0.0027	0.7464
1282	0.0027	0.7466
923	0.0028	0.7443
931	0.0028	0.7432
1316	0.0028	0.7441
1357	0.0029	0.7402
927	0.0031	0.735
1213	0.0031	0.7351
1088	0.0032	0.7338
1304	0.0032	0.7323

933	0.0033	0.7306
924	0.0035	0.7274
937	0.0035	0.7268
91	0.0037	0.7227
247	0.0037	0.7225
1181	0.0037	0.7222
252	0.004	0.7166
1131	0.004	0.7159
1099	0.0041	0.7144
1100	0.0041	0.7154
1200	0.0041	0.7137
288	0.0042	0.7124

Table B.4: Top 50 MS metabolites regression against PBMC donor 1 CD8+ T cells

Metabolite	**P Value**	**R Squared**
1349	0.0004	0.8475
1115	0.0005	0.8407
954	0.0007	0.8273
1201	0.0007	0.8263
1209	0.0009	0.8116
1134	0.001	0.8092
1352	0.0012	0.7971
183	0.0014	0.7888
97	0.0017	0.7748

101	0.0017	0.7764
1107	0.0017	0.7753
1190	0.0018	0.7727
1317	0.0018	0.7711
93	0.0022	0.7605
121	0.0025	0.7519
913	0.0026	0.7494
89	0.0028	0.7426
1316	0.0029	0.7403
1205	0.003	0.7372
1187	0.0032	0.7328
1186	0.0033	0.7303
332	0.0034	0.7292
1098	0.0035	0.7259
1304	0.0035	0.7272
1363	0.0035	0.7276
934	0.0039	0.7183
931	0.0041	0.7141
1213	0.0043	0.7113
937	0.0045	0.7076
1282	0.0045	0.7075
88	0.0046	0.7052
173	0.0047	0.704
927	0.0047	0.7046
274	0.0049	0.6998
829	0.005	0.6981

923	0.005	0.6991
708	0.0052	0.6948
693	0.0053	0.6937
1181	0.0055	0.6912
1026	0.0056	0.6888
924	0.0059	0.6846
1088	0.0059	0.6851
91	0.006	0.6837
247	0.0061	0.682
1357	0.0061	0.6823
201	0.0063	0.6789
106	0.0064	0.6778
288	0.0064	0.6773
1212	0.0064	0.6772
252	0.0065	0.6763

Table B.5: Top 50 MS metabolites regression against PBMC donor 2 CD4+ T cells

Metabolite	**P Value**	**R Squared**
732	0.0002	0.8704
733	0.0003	0.8689
1304	0.0005	0.8425
1321	0.0005	0.8453
765	0.0008	0.8156
1339	0.0009	0.8148

766	0.001	0.8084
773	0.0012	0.7949
1229	0.0012	0.7984
1316	0.0013	0.7917
1230	0.0015	0.7858
1317	0.0016	0.7802
954	0.0017	0.7748
693	0.002	0.7644
708	0.0022	0.7582
395	0.0023	0.7576
637	0.0025	0.7514
201	0.0032	0.734
689	0.0034	0.7291
611	0.0041	0.7143
173	0.0042	0.713
483	0.0059	0.6842
991	0.0059	0.6854
925	0.006	0.6832
767	0.0064	0.6773
863	0.0066	0.6756
928	0.0067	0.6738
599	0.0077	0.661
469	0.0079	0.6594
788	0.0079	0.6593
466	0.008	0.6575
971	0.008	0.6578

1361	0.0082	0.6555
985	0.0083	0.6541
590	0.0088	0.6491
1014	0.0088	0.6482
546	0.0093	0.6432
567	0.0093	0.6429
570	0.0093	0.6429
126	0.0094	0.6423
1349	0.0095	0.6408
873	0.01	0.6366
984	0.0102	0.6343
869	0.0103	0.6332
1363	0.0116	0.621
183	0.0121	0.6171
1201	0.0167	0.5827
266	0.0201	0.5619
248	0.0208	0.5577
1209	0.0215	0.5539

Table B.6: Top 50 MS metabolites regression against PBMC donor 2 CD8+ T cells

Metabolite	**P Value**	**R Squared**
201	0.0001	0.9354
765	0.0001	0.9045
766	0.0001	0.9038

954	0.0001	0.9352
1229	0.0001	0.9067
1230	0.0001	0.8992
1304	0.0001	0.9358
1316	0.0001	0.9449
1317	0.0001	0.9521
1339	0.0001	0.9066
183	0.0002	0.8704
1349	0.0002	0.8734
732	0.0004	0.8509
1209	0.0004	0.8486
1363	0.0005	0.842
733	0.0006	0.8326
913	0.0006	0.8339
1134	0.0006	0.8352
1201	0.0006	0.8358
483	0.0007	0.8239
1115	0.0007	0.8283
97	0.0008	0.8172
101	0.0008	0.8185
788	0.0008	0.8186
934	0.0008	0.8162
1352	0.0009	0.8099
121	0.001	0.8066
708	0.001	0.8075
93	0.0011	0.8034

693	0.0011	0.7996
933	0.0012	0.7983
1321	0.0012	0.7973
89	0.0013	0.794
173	0.0013	0.793
778	0.0014	0.7878
1107	0.0014	0.7892
247	0.0015	0.7849
252	0.0015	0.7847
430	0.0016	0.7808
779	0.0017	0.7769
1190	0.0017	0.7751
88	0.0019	0.7684
637	0.0019	0.7694
1213	0.0019	0.7683
773	0.002	0.7665
1212	0.002	0.7667
427	0.0021	0.7616
1098	0.0023	0.7556
1181	0.0023	0.7563
1187	0.0023	0.7554

APPENDIX C

SUPPLEMENTARY DATA FOR CHAPTER 3

Table C.1: Pearson Correlation Table of Unknown Features

	uk5	uk10	uk24	uk25	uk22	uk16	uk17	uk18	uk26	uk19	uk21	uk23	uk2
uk5	1	0.87565	-0.11255	-0.20269	0.12681	0.35759	0.3393	0.22783	0.36807	0.27563	0.40572	0.28187	0.3788
uk10	0.87565	1	-0.21824	-0.2999	-0.097936	0.17857	0.16065	0.037254	0.21394	0.12052	0.17328	0.12235	0.39458
uk24	-0.11255	-0.21824	1	0.35005	0.40462	0.47939	0.47161	0.49434	0.41286	0.23583	0.39851	0.37155	-0.48018
uk25	-0.20269	-0.2999	0.35005	1	0.57871	0.33224	0.34794	0.60981	0.46352	0.34907	0.29208	0.49507	-0.5475
uk22	0.12681	-0.097936	0.40462	0.57871	1	0.60333	0.63798	0.69065	0.71781	0.46755	0.5016	0.52394	-0.5176
uk16	0.35759	0.17857	0.47939	0.33224	0.60333	1	0.94172	0.79568	0.75491	0.53306	0.79502	0.56039	-0.54117
uk17	0.3393	0.16065	0.47161	0.34794	0.63798	0.94172	1	0.82866	0.74472	0.45739	0.74785	0.59702	-0.53416
uk18	0.22783	0.037254	0.49434	0.60981	0.69065	0.79568	0.82866	1	0.78021	0.44069	0.6296	0.67391	-0.58058
uk26	0.36807	0.21394	0.41286	0.46352	0.71781	0.75491	0.74472	0.78021	1	0.52478	0.65863	0.69442	-0.43375
uk19	0.27563	0.12052	0.23583	0.34907	0.46755	0.53306	0.45739	0.44069	0.52478	1	0.46453	0.32484	-0.2184
uk21	0.40572	0.17328	0.39851	0.29208	0.5016	0.79502	0.74785	0.6296	0.65863	0.46453	1	0.53692	-0.25758
uk23	0.28187	0.12235	0.37155	0.49507	0.52394	0.56039	0.59702	0.67391	0.69442	0.32484	0.53692	1	-0.29564
uk2	0.3788	0.39458	-0.48018	-0.5475	-0.5176	-0.54117	-0.53416	-0.58058	-0.43375	-0.2184	-0.25758	-0.29564	1
uk13	0.38286	0.40298	-0.21635	-0.50491	-0.44966	-0.31619	-0.31297	-0.3023	-0.17713	-0.13971	-0.14266	-0.19176	0.78027
uk12	0.17258	0.27107	-0.42264	-0.27038	-0.40233	-0.42906	-0.36857	-0.31774	-0.30915	-0.32333	-0.17426	-0.05746	0.58294
uk1	-0.12577	0.064455	-0.4437	-0.31179	-0.53433	-0.66671	-0.63516	-0.58011	-0.56901	-0.37543	-0.47078	-0.3172	0.67195
uk7	-0.14008	-0.023242	-0.35405	-0.19338	-0.53058	-0.73238	-0.72976	-0.58487	-0.57645	-0.2913	-0.491	-0.34917	0.67564
uk8	-0.11368	0.056431	-0.49231	-0.40867	-0.65417	-0.79159	-0.76317	-0.69232	-0.66282	-0.43327	-0.51671	-0.42595	0.79737
uk15	-0.062391	0.075104	-0.51189	-0.52805	-0.64416	-0.78564	-0.767	-0.78045	-0.68192	-0.41748	-0.51548	-0.50943	0.82727
uk3	-0.41851	-0.2949	-0.37429	-0.014698	-0.34904	-0.68925	-0.63636	-0.52927	-0.52003	-0.36011	-0.44683	-0.26603	0.3545
uk14	-0.71406	-0.46892	-0.27038	0.07663	-0.35395	-0.65849	-0.63224	-0.51915	-0.59431	-0.3935	-0.5854	-0.40232	-0.031441
uk11	-0.30965	-0.17977	-0.05846	-0.094854	-0.12301	-0.19985	-0.22135	-0.27594	-0.29573	-0.1631	-0.28867	-0.26942	-0.21391
uk20	-0.32918	-0.24	-0.07357	0.059343	-0.088361	-0.28529	-0.30053	-0.25077	-0.2841	-0.17887	-0.31548	-0.27512	-0.10005
uk6	-0.47503	-0.35902	-0.10954	-0.044749	-0.20559	-0.24235	-0.2648	-0.25615	-0.27114	-0.26784	-0.17477	-0.22206	-0.11591
uk4	-0.59959	-0.42586	0.09242	0.26818	0.04656	-0.16727	-0.18065	-0.081005	-0.1927	-0.11359	-0.17969	-0.13214	-0.51767
uk9	-0.54759	-0.253	0.057609	0.12031	-0.18668	-0.2355	-0.29273	-0.23783	-0.31524	-0.1767	-0.3579	-0.35994	-0.38228

Table C.2: Pearson Correlation Table of Unknown Features

	uk13	uk12	uk1	uk7	uk8	uk15	uk3	uk14	uk11	uk20	uk6	uk4	uk9
uk5	0.38286	0.17258	-0.12577	-0.14008	-0.11368	-0.062391	-0.41851	-0.71406	-0.30965	-0.32918	-0.47503	-0.59959	-0.54759
uk10	0.40298	0.27107	0.064455	-0.023242	0.056431	0.075104	-0.2949	-0.46892	-0.17977	-0.24	-0.35902	-0.42586	-0.253
uk24	-0.21635	-0.42264	-0.4437	-0.35405	-0.49231	-0.51189	-0.37429	-0.27038	-0.05846	-0.07357	-0.10954	0.09242	0.057609
uk25	-0.50491	-0.27038	-0.31179	-0.19338	-0.40867	-0.52805	-0.014698	0.07663	-0.094854	0.059343	-0.044749	0.26818	0.12031
uk22	-0.44966	-0.40233	-0.53433	-0.53058	-0.65417	-0.64416	-0.34904	-0.35395	-0.12301	-0.088361	-0.20559	0.04656	-0.18668
uk16	-0.31619	-0.42906	-0.66671	-0.73238	-0.79159	-0.78564	-0.68925	-0.65849	-0.19985	-0.28529	-0.24235	-0.16727	-0.2355
uk17	-0.31297	-0.36857	-0.63516	-0.72976	-0.76317	-0.767	-0.63636	-0.63224	-0.22135	-0.30053	-0.2648	-0.18065	-0.29273
uk18	-0.3023	-0.31774	-0.58011	-0.58487	-0.69232	-0.78045	-0.52927	-0.51915	-0.27594	-0.25077	-0.25615	-0.081005	-0.23783
uk26	-0.17713	-0.30915	-0.56901	-0.57645	-0.66282	-0.68192	-0.52003	-0.59431	-0.29573	-0.2841	-0.27114	-0.1927	-0.31524
uk19	-0.13971	-0.32333	-0.37543	-0.2913	-0.43327	-0.41748	-0.36011	-0.3935	-0.1631	-0.17887	-0.26784	-0.11359	-0.1767
uk21	-0.14266	-0.17426	-0.47078	-0.491	-0.51671	-0.51548	-0.44683	-0.5854	-0.28867	-0.31548	-0.17477	-0.17969	-0.3579
uk23	-0.19176	-0.05746	-0.3172	-0.34917	-0.42595	-0.50943	-0.26603	-0.40232	-0.26942	-0.27512	-0.22206	-0.13214	-0.35994
uk2	0.78027	0.58294	0.67195	0.67564	0.79737	0.82727	0.3545	-0.031441	-0.21391	-0.10005	-0.11591	-0.51767	-0.38228
uk13	1	0.27757	0.3425	0.53889	0.57217	0.58038	-0.038863	-0.31845	-0.26935	-0.10081	-0.23052	-0.60483	-0.3348
uk12	0.27757	1	0.7835	0.45856	0.69207	0.566	0.59528	0.31279	-0.12604	-0.31795	0.078072	-0.074403	-0.21413
uk1	0.3425	0.7835	1	0.69703	0.89997	0.80031	0.71109	0.54809	-0.10874	-0.084104	0.19131	-0.0064376	0.0591
uk7	0.53889	0.45856	0.69703	1	0.83218	0.75345	0.62989	0.44487	-0.028919	0.13357	0.18022	-0.082592	0.048264
uk8	0.57217	0.69207	0.89997	0.83218	1	0.92291	0.66243	0.49164	-0.055407	0.037367	0.20203	-0.036922	0.031356
uk15	0.58038	0.566	0.80031	0.75345	0.92291	1	0.59449	0.4047	-0.0052131	0.088123	0.15997	-0.082545	-0.01302
uk3	-0.038863	0.59528	0.71109	0.62989	0.66243	0.59449	1	0.73323	0.064389	0.084882	0.28238	0.26241	0.02972
uk14	-0.31845	0.31279	0.54809	0.44487	0.49164	0.4047	0.73323	1	0.35636	0.26933	0.4863	0.65802	0.58923
uk11	-0.26935	-0.12604	-0.10874	-0.028919	-0.055407	-0.0052131	0.064389	0.35636	1	0.3455	0.15577	0.31704	0.34068
uk20	-0.10081	-0.31795	-0.084104	0.13357	0.037367	0.088123	0.084882	0.26933	0.3455	1	0.10932	0.28371	0.44186
uk6	-0.23052	0.078072	0.19131	0.18022	0.20203	0.15997	0.28238	0.4863	0.15577	0.10932	1	0.43879	0.36987
uk4	-0.60483	-0.074403	-0.0064376	-0.082592	-0.036922	-0.082545	0.26241	0.65802	0.31704	0.28371	0.43879	1	0.68903
uk9	-0.3348	-0.21413	0.0591	0.048264	0.031356	-0.01302	0.02972	0.58923	0.34068	0.44186	0.36987	0.68903	1

Table C.3: Important Metabolites Coefficient Value From Linear Regression

	leucine	threonine	alanine	asparate	succinate	ethanol	acetic_acid	isoleucine	glycine
iMSC_P1	-2500000000	2500000000	1.65E+10	-1.05E+10	5.55E+10	4.9E+10	3500000000	8500000000	6.5E+10
iMSC_P2	-4000000000	2500000000	1.4E+10	-8E+09	3E+10	7.35E+10	2.4E+10	3000000000	1.3E+10
iMSC_P3	-1.35E+10	5500000000	2000000000	-1.3E+10	4.5E+10	7.3E+10	5500000000	1.35E+10	5.5E+10
BM71_P1	-1.5E+10	4E+10	2.6E+10	-8.5E+09	6E+10	6.6E+10	1.6E+10	6500000000	1.4E+10
BM71_P2	-1E+10	1.4E+10	3.15E+10	-2.3E+10	9.55E+10	2.43E+11	2.5E+10	3000000000	1.81E+11
BM71_P3	-2.5E+10	3E+10	3.05E+10	-3.1E+10	9.2E+10	2.06E+11	3.25E+10	8000000000	1.99E+11
BM182_P1	-2000000000	2500000000	2E+10	-1E+10	5.4E+10	1.19E+11	1.65E+10	4000000000	1.06E+11
BM182_P2	-1000000000	1.5E+10	1.25E+10	-1.05E+10	4.8E+10	7.15E+10	3500000000	3500000000	1.14E+11
BM182_P3	-500000000	4000000000	1.3E+10	-500000000	3.7E+10	8.6E+10	1.6E+10	-500000000	6000000000

Table C.4: Correlation of metabolites at each time point and composite score

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	FDR adjusted_pvalue	significant
glycine_Day2	Composite score	0.7973	9	0.283	0.9555	0.0101	0.72774	FALSE
phenylalanine_day3	Composite score	0.757	9	0.1869	0.9457	0.0182	0.72774	FALSE
uk13_Day2	Composite score	-0.7266	9	-0.9381	-0.1207	0.0266	0.72774	FALSE
uk9_Day2	Composite score	0.7254	9	0.1182	0.9377	0.027	0.72774	FALSE
glycine_Day 1	Composite score	0.7128	9	0.0924	0.9345	0.0311	0.72774	FALSE
proline_Day 1	Composite score	0.6954	9	0.0582	0.93	0.0375	0.73125	FALSE
threonine_Day 1	Composite score	0.6776	9	0.0245	0.9253	0.0449	0.750471429	FALSE

Table C.4 continued from previous page

betaine_day3	Composite score	0.6526	9	-0.0203	0.9186	0.0567	0.8292375	FALSE
tyrosine_Day2	Composite score	0.6037	9	-0.1009	0.905	0.0852	0.91809759	FALSE
formate_day3	Composite score	0.5981	9	-0.1095	0.9034	0.0889	0.91809759	FALSE
ethanol_Day 1	Composite score	0.5901	9	-0.1218	0.9011	0.0944	0.91809759	FALSE
tyrosine_day3	Composite score	0.5784	9	-0.1393	0.8977	0.1028	0.91809759	FALSE
creatine_Day 1	Composite score	0.5702	9	-0.1512	0.8953	0.1089	0.91809759	FALSE
uk2_Day2	Composite score	-0.5633	9	-0.8932	0.1611	0.1143	0.91809759	FALSE
uk11_Day2	Composite score	0.5466	9	-0.1845	0.8883	0.1278	0.91809759	FALSE
betaine_Day 1	Composite score	0.5314	9	-0.2051	0.8837	0.1409	0.91809759	FALSE
tyrosine_Day 1	Composite score	-0.5078	9	-0.8764	0.2358	0.1628	0.91809759	FALSE
uk15_Day 1	Composite score	0.4909	9	-0.257	0.8711	0.1796	0.91809759	FALSE
uk13_day3	Composite score	-0.4756	9	-0.8661	0.2755	0.1957	0.91809759	FALSE
alanine_Day2	Composite score	-0.4697	9	-0.8642	0.2825	0.202	0.91809759	FALSE
cytidine_Day 1	Composite score	0.4483	9	-0.3074	0.8572	0.2262	0.91809759	FALSE
uk11_Day 1	Composite score	0.44	9	-0.3167	0.8544	0.236	0.91809759	FALSE
lactate_day3	Composite score	0.4369	9	-0.3201	0.8534	0.2397	0.91809759	FALSE
uk12_day3	Composite score	0.4295	9	-0.3283	0.8509	0.2487	0.91809759	FALSE
uk6_Day2	Composite score	0.4195	9	-0.3391	0.8475	0.261	0.91809759	FALSE
leucine _day3	Composite score	-0.4025	9	-0.8416	0.3571	0.2828	0.91809759	FALSE
uk14_Day2	Composite score	0.3997	9	-0.36	0.8407	0.2866	0.91809759	FALSE
lactate_Day2	Composite score	-0.3976	9	-0.8399	0.3622	0.2893	0.91809759	FALSE
glucose_day3	Composite score	-0.3929	9	-0.8383	0.367	0.2956	0.91809759	FALSE
glucose_Day 1	Composite score	0.3916	9	-0.3683	0.8378	0.2973	0.91809759	FALSE
uk7_Day2	Composite score	-0.3855	9	-0.8357	0.3745	0.3055	0.91809759	FALSE
arginine_Day2	Composite score	-0.3807	9	-0.834	0.3793	0.3121	0.91809759	FALSE
uk7_day3	Composite score	-0.3805	9	-0.8339	0.3795	0.3124	0.91809759	FALSE
acetic_acid_Day2	Composite score	-0.3656	9	-0.8285	0.3943	0.3333	0.91809759	FALSE
uk10_Day 1	Composite score	0.3607	9	-0.399	0.8268	0.3403	0.91809759	FALSE
acetic_acid_Day 1	Composite score	-0.3568	9	-0.8253	0.4028	0.3459	0.91809759	FALSE
uk3_Day 1	Composite score	0.3556	9	-0.4039	0.8249	0.3477	0.91809759	FALSE
succinate_Day2	Composite score	-0.3524	9	-0.8237	0.407	0.3523	0.91809759	FALSE
myo_inositol_Day 1	Composite score	0.3476	9	-0.4115	0.822	0.3593	0.91809759	FALSE
glutamine_Day2	Composite score	0.3388	9	-0.4198	0.8187	0.3725	0.91809759	FALSE
succinate_Day 1	Composite score	0.3364	9	-0.422	0.8178	0.376	0.91809759	FALSE
pyruvate_day3	Composite score	-0.335	9	-0.8173	0.4233	0.3782	0.91809759	FALSE
valine_day3	Composite score	-0.3349	9	-0.8173	0.4234	0.3783	0.91809759	FALSE
pyruvate_Day 1	Composite score	0.3308	9	-0.4272	0.8157	0.3846	0.91809759	FALSE
uk4_Day 1	Composite score	0.3295	9	-0.4284	0.8152	0.3866	0.91809759	FALSE
isoleucine _day3	Composite score	-0.3198	9	-0.8116	0.4372	0.4015	0.91809759	FALSE
cytidine_Day2	Composite score	0.3169	9	-0.4398	0.8105	0.406	0.91809759	FALSE
uk1_Day 1	Composite score	0.312	9	-0.4441	0.8086	0.4137	0.91809759	FALSE
phenylalanine_Day 1	Composite score	0.3099	9	-0.446	0.8078	0.4171	0.91809759	FALSE
uk6_Day 1	Composite score	-0.2937	9	-0.8015	0.4602	0.443	0.91809759	FALSE
uk5_Day2	Composite score	-0.2891	9	-0.7997	0.4641	0.4505	0.91809759	FALSE
formate_Day2	Composite score	0.276	9	-0.4753	0.7945	0.4723	0.91809759	FALSE
creatine_Day2	Composite score	-0.2734	9	-0.7935	0.4774	0.4766	0.91809759	FALSE
asparate_Day2	Composite score	0.2679	9	-0.482	0.7913	0.4858	0.91809759	FALSE
glutamine_Day 1	Composite score	0.267	9	-0.4827	0.7909	0.4874	0.91809759	FALSE
threonine_Day2	Composite score	0.2645	9	-0.4848	0.7899	0.4916	0.91809759	FALSE
leucine _Day 1	Composite score	0.2638	9	-0.4853	0.7896	0.4927	0.91809759	FALSE
lactate_Day 1	Composite score	-0.2633	9	-0.7894	0.4857	0.4936	0.91809759	FALSE
uk2_day3	Composite score	-0.262	9	-0.7889	0.4868	0.4959	0.91809759	FALSE
asparate_Day 1	Composite score	0.2603	9	-0.4882	0.7882	0.4987	0.91809759	FALSE

Table C.4 continued from previous page

acetic_acid_day3	Composite score	-0.2591	9	-0.7877	0.4892	0.5008	0.91809759	FALSE
uk2_Day 1	Composite score	0.2579	9	-0.4902	0.7872	0.5029	0.91809759	FALSE
uk5_Day 1	Composite score	0.2466	9	-0.4993	0.7826	0.5224	0.91809759	FALSE
uk10_day3	Composite score	0.2446	9	-0.5009	0.7817	0.5259	0.91809759	FALSE
uk6_day3	Composite score	-0.2435	9	-0.7813	0.5018	0.5278	0.91809759	FALSE
myo_inositol_day3	Composite score	0.2407	9	-0.504	0.7801	0.5328	0.91809759	FALSE
betaine_Day2	Composite score	0.232	9	-0.5108	0.7765	0.548	0.91809759	FALSE
fructose_Day 1	Composite score	-0.2268	9	-0.7743	0.5149	0.5574	0.91809759	FALSE
arginine_Day 1	Composite score	0.2268	9	-0.5149	0.7743	0.5574	0.91809759	FALSE
asparate_day3	Composite score	0.2256	9	-0.5158	0.7738	0.5594	0.91809759	FALSE
uk12_Day2	Composite score	-0.2249	9	-0.7735	0.5163	0.5606	0.91809759	FALSE
fructose_Day2	Composite score	-0.2172	9	-0.7702	0.5223	0.5746	0.91809759	FALSE
uk14_Day 1	Composite score	0.2078	9	-0.5294	0.7662	0.5916	0.91809759	FALSE
uk8_Day2	Composite score	-0.206	9	-0.7654	0.5308	0.595	0.91809759	FALSE
uk3_Day2	Composite score	0.2019	9	-0.5338	0.7636	0.6025	0.91809759	FALSE
uk5_day3	Composite score	0.1965	9	-0.5378	0.7613	0.6124	0.91809759	FALSE
phenylalanine_Day2	Composite score	0.1937	9	-0.5398	0.7601	0.6175	0.91809759	FALSE
valine_Day2	Composite score	-0.1889	9	-0.7579	0.5434	0.6264	0.91809759	FALSE
myo_inositol_Day2	Composite score	-0.1825	9	-0.7551	0.548	0.6383	0.91809759	FALSE
pyruvate_Day2	Composite score	0.1792	9	-0.5504	0.7536	0.6446	0.91809759	FALSE
uk15_Day2	Composite score	-0.1787	9	-0.7534	0.5508	0.6454	0.91809759	FALSE
glucose_Day2	Composite score	0.177	9	-0.552	0.7526	0.6487	0.91809759	FALSE
glutamine_day3	Composite score	0.1756	9	-0.553	0.752	0.6513	0.91809759	FALSE
uk4_day3	Composite score	-0.1596	9	-0.7448	0.5643	0.6816	0.9199125	FALSE
isoleucine _Day 1	Composite score	0.1595	9	-0.5644	0.7447	0.6819	0.9199125	FALSE
creatine_day3	Composite score	-0.1584	9	-0.7442	0.5652	0.684	0.9199125	FALSE
uk4_Day2	Composite score	0.157	9	-0.5662	0.7436	0.6867	0.9199125	FALSE
arginine_day3	Composite score	-0.1543	9	-0.7423	0.5681	0.6919	0.9199125	FALSE
uk8_day3	Composite score	-0.1371	9	-0.7344	0.5798	0.725	0.95043	FALSE
uk9_day3	Composite score	0.134	9	-0.5819	0.7329	0.7311	0.95043	FALSE
fructose_day3	Composite score	-0.1204	9	-0.7264	0.591	0.7577	0.9545625	FALSE
proline_Day2	Composite score	0.11	9	-0.5978	0.7214	0.7782	0.9545625	FALSE
uk12_Day 1	Composite score	0.1058	9	-0.6005	0.7194	0.7865	0.9545625	FALSE
uk1_day3	Composite score	-0.1033	9	-0.7182	0.6021	0.7914	0.9545625	FALSE
uk8_Day 1	Composite score	0.0894	9	-0.611	0.7113	0.819	0.9545625	FALSE
uk7_Day 1	Composite score	-0.0863	9	-0.7098	0.6129	0.8252	0.9545625	FALSE
leucine _Day2	Composite score	-0.0855	9	-0.7094	0.6134	0.8268	0.9545625	FALSE
ethanol_Day2	Composite score	0.0819	9	-0.6157	0.7075	0.8341	0.9545625	FALSE
threonine_day3	Composite score	-0.0796	9	-0.7064	0.6172	0.8388	0.9545625	FALSE
isoleucine _Day2	Composite score	-0.0787	9	-0.7059	0.6177	0.8406	0.9545625	FALSE
succinate_day3	Composite score	-0.077	9	-0.7051	0.6188	0.844	0.9545625	FALSE
proline_day3	Composite score	0.0765	9	-0.6191	0.7048	0.8449	0.9545625	FALSE
uk9_Day 1	Composite score	0.0751	9	-0.6199	0.7041	0.8477	0.9545625	FALSE
uk11_day3	Composite score	-0.0747	9	-0.7039	0.6202	0.8485	0.9545625	FALSE
uk13_Day 1	Composite score	-0.0685	9	-0.7007	0.624	0.861	0.9594	FALSE
uk15_day3	Composite score	0.0544	9	-0.6326	0.6935	0.8894	0.962245946	FALSE
cytidine_day3	Composite score	-0.0534	9	-0.6929	0.6332	0.8915	0.962245946	FALSE
glycine_day3	Composite score	0.0476	9	-0.6367	0.6899	0.9032	0.962245946	FALSE
uk10_Day2	Composite score	-0.0465	9	-0.6893	0.6373	0.9054	0.962245946	FALSE
valine_Day 1	Composite score	0.0453	9	-0.638	0.6887	0.9079	0.962245946	FALSE
alanine_day3	Composite score	-0.0428	9	-0.6874	0.6395	0.9129	0.962245946	FALSE
alanine_Day 1	Composite score	0.0334	9	-0.645	0.6824	0.932	0.969546903	FALSE
uk3_day3	Composite score	-0.0313	9	-0.6812	0.6463	0.9364	0.969546903	FALSE

Table C.4 continued from previous page

uk1_Day2	Composite score	-0.0206	9	-0.6755	0.6525	0.9581	0.979774138	FALSE
formate_Day 1	Composite score	-0.0145	9	-0.6722	0.6559	0.9704	0.979774138	FALSE
uk14_day3	Composite score	-0.014	9	-0.6719	0.6562	0.9714	0.979774138	FALSE
ethanol_day3	Composite score	0.0045	9	-0.6616	0.6666	0.9908	0.9908	FALSE

APPENDIX D

SUPPLEMENTARY DATA FOR CHAPTER 4

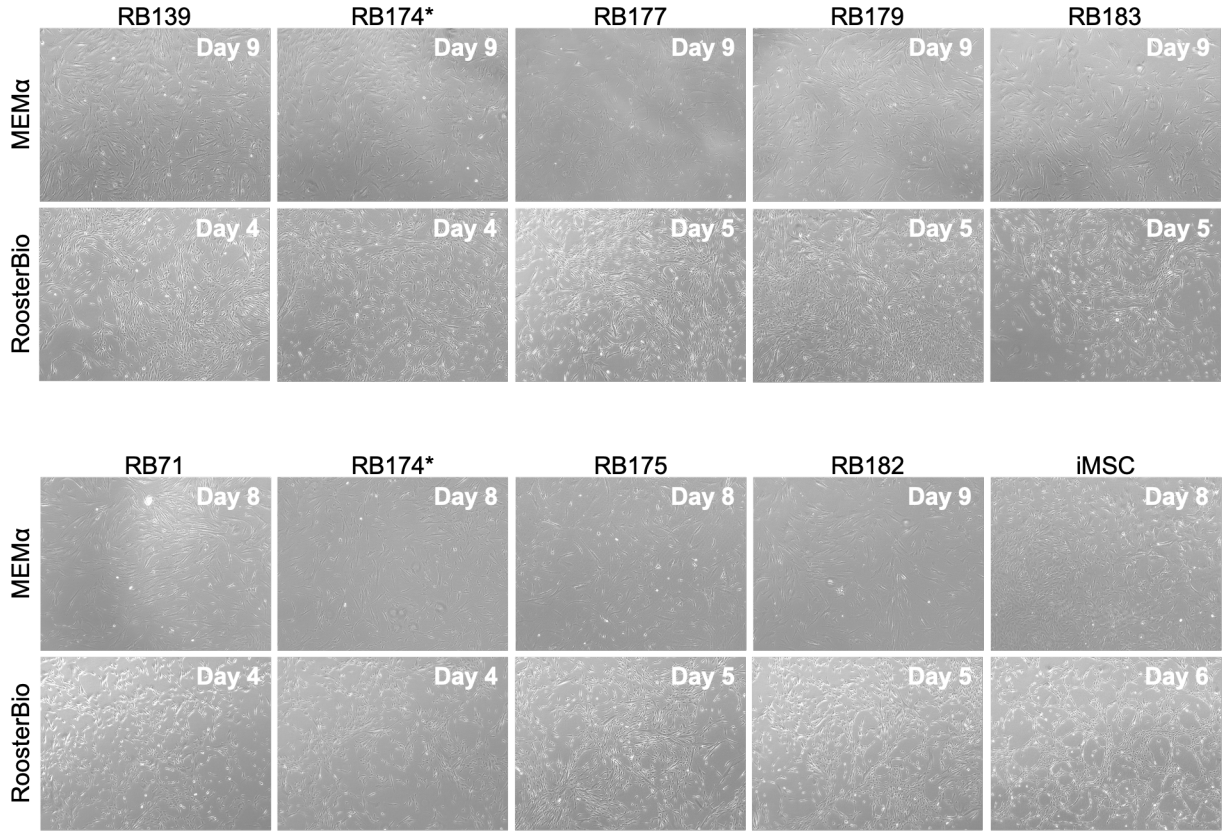
Table D.1: Population doubling level of each cell line

Donor	Media used for expansion	Initial cells seeded (Ci)	Final cell yield (Cf)	Initial population doubling level (PDL0)	Final PDL
RB139	aMEM	3.75E+05	8.10E+06	12.2	16.6
RB174.1	aMEM	3.75E+05	9.80E+06	13.06	17.8
RB177	aMEM	3.75E+05	4.23E+06	15.1	18.6
RB179	aMEM	3.75E+05	8.20E+06	12.6	17.1
RB183	aMEM	3.75E+05	5.30E+06	12	15.8
RB139	RoosterBio	3.75E+05	1.51E+07	12.2	17.5
RB174.1	RoosterBio	3.75E+05	2.55E+07	13.06	19.1
RB177	RoosterBio	3.75E+05	2.12E+07	15.1	20.9
RB179	RoosterBio	3.75E+05	2.66E+07	12.6	18.7
RB183	RoosterBio	3.75E+05	7.80E+06	12	16.4

Table D.1 continued from previous page

RB71	aMEM	3.75E+05	7.50E+06	12.72	17.0
RB174.3	aMEM	3.75E+05	8.43E+06	13.06	17.6
RB175	aMEM	3.75E+05	6.24E+06	12	16.1
RB182	aMEM	3.75E+05	3.65E+06	11.48	14.8
iMSC	aMEM	3.75E+05	1.86E+07	5.7	17.1
RB71	RoosterBio	3.75E+05	1.50E+07	12.72	18.0
RB174.3	RoosterBio	3.75E+05	1.29E+07	13.06	18.2
RB175	RoosterBio	3.75E+05	1.92E+07	12	17.7
RB182	RoosterBio	3.75E+05	1.47E+07	11.48	16.8
iMSC	RoosterBio	3.75E+05	1.90E+07	5.7	17.1

Figure D.0.1: MSC morphology at harvest time



APPENDIX E

BIOLOGICAL SKETCH

Xunan Shen was born and raised in China, where she got her primary, secondary, and college education. She graduated with a degree in Pharmacy at Nanjing University of Chinese Medicine in 2015. Then, she came to the United States for her master's degree in Pharmacology and Toxicology at the University of Kansas (KU), where she graduated in 2017. At KU, she worked in the Dr. Staudinger's Lab to investigate the Pregnane X receptor SUMOylation in primary mice brain. In 2017 she joined the University of Georgia ILS program and proceeded to join the Bioinformatics Ph.D. program. She joined the Edison Lab in 2018. At the Edison Lab, she studied the Mensenchymal stem cells immunomodulation effects and how to use metabolomics to predict MSCs potency. After graduation, she wants to join in research institution in China.