RAMAN SPECTROSCOPIC STUDY OF INDUCED PLURIPOTENT STEM CELLS: CHARACTERIZATION, IDENTIFICATION, AND DISCRIMINATION

by

YAN TAN

B.Sc.(Honors), Peking University, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2012

© Yan Tan, 2012
Abstract

Raman microspectroscopy is a non-destructive, label-free technique that offers information-rich molecular analysis of living cells. This work is the first reported Raman spectroscopic study of human induced pluripotent cells (hiPSCs), a very promising new source of non-embryonic pluripotent stem cells for drug screening, toxicity assessment, regenerative therapies, and clinical research. The Raman signatures of hiPSCs and human embryonic stem cells (hESCs) were found to be highly similar, and both distinguishable from differentiated hESCs in terms of relative Raman peak intensities and variances. Principal component analysis (PCA) of the Raman spectra demonstrated a clear distinction between hiPSCs and differentiated hESCs. Additionally, the effects of culture confluencies and cell line differences on Raman spectra of hiPS cells was investigated. It was confirmed that the spectral similarity between hiPSCs and hESCs, along with the dissimilarity between hiPSCs and differentiated hESCs were qualitatively consistent over various cell culture confluencies, and between the two available hiPS cell lines. Therefore, the results suggested that the overall cellular composition of hiPSC was more similar to that of the hESC that these cells were designed to resemble than the somatics cell from which they were derived. It is suggested that the observed spectral differences between hiPSC and hESC may be due to factors relating to reprogramming (rather than cell density difference or cell line artifacts).

Attempts were also made to investigate how Raman features of hiPS cells change during their differentiation. The pluripotent and differentiated iPSCs exhibited significantly different Raman spectral profiles; these differences were qualitatively similar to, but less marked, than differences between pluripotent and differentiated hESC. Overall, this work contributed important new data and practical insights into the utility of Raman microspectroscopy for characterization, identification, and discrimination of iPSCs and hESCs.
Preface

A version of chapter 2 and chapter 3 has been recently submitted as **Y.Tan, S.O. Konorov, H.G.Schulze, J.M.Piret, M.W. Blades and R.F.B. Turner**, Comparative study using Raman microspectroscopy reveals spectral signatures of human induced pluripotent cells more closely resemble human embryonic stem cells than differentiated cells.

I was the main contributor of this research and wrote the manuscript. I conducted all the testing and data analysis except for Raman spectra of embryonic stem cells and differentiated embryonic stem cells, collected by Stanislav Konorov.
Table of contents

Abstract ........................................................................................................................................... ii
Preface ............................................................................................................................................... iii
Table of contents ................................................................................................................................. iv
List of tables ....................................................................................................................................... vi
List of figures ..................................................................................................................................... vii
List of abbreviations ......................................................................................................................... xi
Acknowledgements ......................................................................................................................... xii
Dedication .......................................................................................................................................... xiii

1 Introduction .................................................................................................................................... 1
  1.1 Raman spectroscopy ................................................................................................................ 1
  1.1.1 Raman scattering .................................................................................................................. 1
  1.1.2 Raman signal enhancement strategies ................................................................................ 4
  1.2 Applications of Raman spectroscopy in biological systems .................................................... 7
  1.3 Human induced pluripotent cells and embryonic stem cells ................................................... 8
  1.4 Data processing methods ......................................................................................................... 9
    1.4.1 Multivariate methods in cell discrimination ...................................................................... 9
    1.4.2 The student’s t-test .......................................................................................................... 10
  1.5 Previous work .......................................................................................................................... 11
  1.6 Thesis content and organization .............................................................................................. 12

2 A comparative study of hiPS cells with hES and differentiated hES cells using Raman
microspectroscopy .......................................................................................................................... 14
  2.1 Introduction ............................................................................................................................. 14
  2.2 Material and methods ............................................................................................................. 15
    2.2.1 Cell culture ....................................................................................................................... 15
    2.2.2 Raman microspectroscopy of hiPS and hES cells ........................................................... 17
    2.2.3 Fluorescence microscopy ............................................................................................... 17
    2.2.4 Data analysis ................................................................................................................... 18
  2.3 Results and discussion .............................................................................................................. 21
List of tables

Table 2.1 Two-sample $t$-test results generally show the intensities of major peaks in Raman spectra of hiPSC and hES cells to be not statistically different. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.$^{[78]}$ .......................................................... 22

Table 2.2 Two-sample $t$-test results show significant differences for the intensities of major peaks in Raman spectra of hiPSC and differentiated hES cells. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.$^{[78]}$ .......................................................... 30

Table 3.1 Two-sample $t$-test results generally show the intensities of major peaks in Raman spectra of EOS hiPS cells (n=17) and MSCiPS1 cells (n=26) to be not statistically different. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.$^{[78]}$ ........................................................................................................ 38

Table 3.2 Two-sample $t$-test results generally show the intensities of major peaks in Raman spectra of MSCiPS1 (n=26) and hES cells (n=48) to be not statistically different. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.$^{[78]}$ .......... 39

Table 4.1 Correlation coefficients of relative intensities of major Raman peaks...................... 53

Table 4.2 Two sample $t$-test results generally show the intensities of major peaks in vector normalized Raman spectra of undifferentiated and differentiated hiPSCs to be statistically different. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.$^{[78]}$ ........................................................................................................ 61

Table 5.1 Raman peaks of interest in the Raman spectra$^{[46]}$ .................................................. 78
List of figures

Figure 1.1 Diagram of Rayleigh scattering, Raman scattering and infrared absorption processes\textsuperscript{[17]} .................................................................................................................................................. 2

Figure 2.1 Left: GFP expression of hiPS cells. Right: Bright field image of the same growth area. Images were captured after the Raman signal acquisition from hiPSCs 3 days after sub-culturing. Before taking the GFP image, the basal medium for Raman spectroscopy was replaced by PBS in order to reduce the background fluorescence emission. ...................... 18

Figure 2.2 Sample raw spectra from hiPS cells (3 days after sub-culturing on the mirror, same set of data was used in the following comparisons). The colored spectra were collected from different spots in a growth area; the bottom bold one was the average of 4 spectra from the cell-free area (2 before the cell spectra collection and 2 after). ..................................... 19

Figure 2.3 Background subtracted spectra using the raw spectra in Figure 2.2 .................. 20

Figure 2.4 Baseline flattened and smoothed spectra using the background subtracted spectra in Figure 2.2. Three spectra were rejected as indicated by the arrows. ...................... 20

Figure 2.5 Means of normalized Raman spectra, taken from living hES (N=49) and living hiPS cells (N=17) were highly similar................................................................. 21

Figure 2.6 The standard deviations of normalized Raman spectra, taken from living hES (N=49) and hiPS (N=17) cells, show reduced metabolic activity of hiPS cells compared to hES cells ......................................................................................................................... 23

Figure 2.7 Growth curves of hiPS cells, (a) a growth curve generated from a seeding density of 3.69×10\textsuperscript{5} cells/well. Exponential regression result is based on the 2nd to 4th data points. The calculated doubling time is 22.0±2.9 hours. (b) a growth curve generated from a seeding density of 1.50×10\textsuperscript{5} cells/well. Exponential regression result is based on the 2nd to 6th data points. The calculated doubling time is 24.0±2.2 hours. .......................................................... 24

Figure 2.8 Glycogen metabolism was the major contributor to spectral variances as shown by the first principal component (PC1) of the combined data. The second principal component (PC2) was more complex and included lipid and protein-related peaks. (a) PC1 and PC2 of the combined normalized spectra of hiPSCs and hESCs. (b) First principal component (PC1) of the combined data from hiPS cells and hES cells, compared with Raman spectra from solid D-glycogen................................................................. 26
Figure 2.9 A PC2/PC1 score plot revealed a tendency for hiPS and hES cells to cluster independently along the PC2 dimension. Thus, the Raman spectra of hiPS cells fairly resembled undifferentiated hESCs of the hESC line, but subtle differences nevertheless could be detected.

Figure 2.10 Comparison between hiPS cells and differentiated hES cells. A) The means of normalized Raman spectra, taken from living differentiated hES (n = 10) and living hiPS cells (n = 17) were notably different. B) Similarly, standard deviations showed marked differences between hiPS and differentiated hES cells. C) The spectral variances as shown by the first two (PC1, PC2) principal components of the combined data were dominated by protein and lipid-related bands. D) A PC2/PC1 score plot revealed a tendency for hiPS and differentiated hES cells to cluster independently along the PC1 dimension. The dispersion of PC score data points from spontaneously differentiated hES cells likely reflected a variety of cell types after non-specific differentiation. Thus, the Raman spectra of hiPS cells, which closely resembled undifferentiated hES cells, were markedly different from those of differentiated hES cells.

Figure 2.11 First principal components (PC1) of combined data from differentiated ES cells and iPS cells, and standard deviation of normalized spectra from differentiated ES cells. Scaled for clarity.

Figure 3.1 The means of normalized Raman spectra, taken from living hiPS EOS (n = 17) and MSChiPS1 cells (n = 27) were highly similar.

Figure 3.2 Standard deviations of normalized EOS iPS and MSCiPS1 cell spectra.

Figure 3.3 A PC1/PC2 score plot indicated a tendency for iPS cells (EOS line and MSCiPS1 line) and hES cells to cluster independently along the PC2 dimension, not distinguishable along the PC1 dimension. EOS iPS and MSCiPS1 cells are not distinguishable in terms of Raman spectra.

Figure 3.4 Comparison between hiPS EOS cells of various confluences, hES cells and differentiated hES cells. Spectra were taken from hiPS cells 1-5 days after sub-culturing (Day 1, the number of spectra taken n=14; Day 2, n=19; Day 3, n=17; Day 4, n=17; Day 5, n=18). Before principal component analysis these spectra were combine with spectra from hES cells (n=48) and differentiated hES cells (n=10), used in previous comparisons. A) PC2/PC1 score plot revealed the similarity of all the iPSCs and hES cells. B) PC2/PC1 score
plot showed the tendency for pluripotent hiPS to cluster along with cells and differentiated hES cells to cluster independently along the PC1 dimension.

Figure 3.5 Mean spectra of hiPS EOS cells from different days after sub-culturing. All the spectra were normalized to the nucleic acid line at 783 cm\(^{-1}\). All spectra had the R4 (757 cm\(^{-1}/783 \text{ cm}^{-1}\)) differentiation state markers consistent with those of undifferentiated ES cells.

Figure 3.6 PC2/PC1 scores plot of Raman spectra from hiPS EOS cells on the 1st-5th day after sub-culturing.

Figure 4.1 Mean spectra of hiPSCs after 0, 2, 4, 8, 15, and 20 days of differentiation. Spectra were normalized to the 783 cm\(^{-1}\).

Figure 4.2 Temporal profiles of relative intensities of major Raman peaks in the mean spectra. (A) Relative intensities of 852 cm\(^{-1}\), 900 cm\(^{-1}\), 938 cm\(^{-1}\), and 1031 cm\(^{-1}\). (B) Relative intensities of 757 cm\(^{-1}\), 876 cm\(^{-1}\), and 1002 cm\(^{-1}\). Other peaks show fluctuation over the 20 days period (not shown).

Figure 4.3 PC3/PC4/PC1 score plot of differentiated iPSCs.

Figure 4.4 (A) PC loads of differentiated data of hiPS cells (B) PC scores of differentiated hiPS cells.

Figure 4.5 (A) PC1/PC3/PC2 scores plot of differentiated iPSCs (2, 4, 8, 15 and 20 days of differentiation) and undifferentiated iPSCs (1, 2, 3, 4 and 5 days after sub-culturing), CA1 cells (5 days after sub-culturing) and differentiated CA1 cells (20 days of differentiation). (B) A simplified color scheme of Figure 4.4 A, with all differentiated iPSCs represented by blue crosses, all undifferentiated iPSCs represented by cyan circles. Inserted red-edged figure is a zoomed-in view of cyan dashed rectangular. There are 42.7\% (91 data points out of 217 in total) differentiated iPSCs falling into the orange oval “pluripotent area”.

Figure 4.6 Mean spectra of differentiated iPSCs (2, 4, 8, 15 and 20 days after differentiation) and undifferentiated hiPSCs (3 days after sub-culturing designated Day 0 in the legend).

Figure 4.7 Mean spectra of differentiated (20 days) and undifferentiated iPS cells after vector normalization.

Figure 4.8 PC1/PC2/PC3 scores plot of undifferentiated and differentiated iPS cells (20 days differentiation).
Figure 4.9 (A) PC1/PC2 loading plot (B) PC3/PC4 loading plot. Contributions of each PC are PC1, 72.5%; PC2 8.8%; PC3 4.5%; PC4 2.5%. The first four PCs captured 88.3% of the total variance. .......................................................... 63

Figure 4.10 PC1/PC2/PC4 scores plot of the vector normalized Raman spectra from undifferentiated iPSCs and differentiated iPSCs (2, 4, 8, 15, and 20 days of differentiation). The dashed line separates undifferentiated zone (below) and differentiated zone. The number of differentiated cells below the line: 14 after 2 days differentiation, 12 after 4 days of differentiation, 5 after 8 days of differentiation, 0 after 15 days of differentiation, 1 after 20 days of differentiation. ........................................................................................................ 64

Figure 4.11 PC loadings of Raman spectra from undifferentiated and differentiated iPS cells (0 day, 2 day, 4 days, 8 days, 15 days, and 20 days of differentiation) (A) PC1/PC2 loadings plot (B) PC3/PC4 loadings plot. Contributions of each PC are PC1, 49.8%; PC2 11.8%; PC3 7.1%; PC4 5.1%. The first four PCs captured 88.3% of the total variance. ......................... 65
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPS cell</td>
<td>induced pluripotent stem cell</td>
</tr>
<tr>
<td>hiPS cell</td>
<td>human induced pluripotent stem cell</td>
</tr>
<tr>
<td>ES cell</td>
<td>embryonic stem cell</td>
</tr>
<tr>
<td>hES cell</td>
<td>human embryonic stem cell</td>
</tr>
<tr>
<td>IR spectroscopy</td>
<td>infrared spectroscopy</td>
</tr>
<tr>
<td>UV-Vis spectroscopy</td>
<td>ultraviolet-visible spectroscopy</td>
</tr>
<tr>
<td>RS</td>
<td>Raman spectroscopy</td>
</tr>
<tr>
<td>RRS</td>
<td>resonance Raman spectroscopy</td>
</tr>
<tr>
<td>SERS</td>
<td>surface enhanced Raman spectroscopy</td>
</tr>
<tr>
<td>TERS</td>
<td>tip-enhanced Raman spectroscopy</td>
</tr>
<tr>
<td>CARS</td>
<td>coherent anti-Stokes Raman microscopy</td>
</tr>
<tr>
<td>SRS</td>
<td>stimulated Raman scattering</td>
</tr>
<tr>
<td>SORS</td>
<td>spatially offset Raman spectroscopy</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>LDA</td>
<td>linear discriminant analysis</td>
</tr>
<tr>
<td>HCA</td>
<td>hierarchical cluster analysis</td>
</tr>
<tr>
<td>RµS</td>
<td>Raman microspectroscopy</td>
</tr>
<tr>
<td>EGFP</td>
<td>enhanced green fluorescent protein</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>rtPCR</td>
<td>real time polymerase chain reaction</td>
</tr>
<tr>
<td>FTIR spectroscopy</td>
<td>Fourier transform infrared spectroscopy</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to first thank my supervisors, Dr. Michael W. Blades and Dr. Robin F.B. Turner for their guidance and all the insights during my studies. It is them who leaded me into the Raman world and help me make every little step forward.

I would like to thank Dr. H. Georg Schulze for his great advice and countless help with the data analysis and interpretation. I would also like to thank all of my lab mates, Dr. Stanislav O. Konorov, Kadek Okuda, Chad G. Atkins, Dr. Christopher J. Addison, and C.W. Tony Yang for their constant support and help through my studies.

My gratitude is also devoted to Dr. James M. Piret and for allowing me to use the cell culture facility, the entire Piret group especially Chris Sherwood for their help in cell line maintenance and sharing the knowledge of stem cell biology.

Special thanks to my boyfriend, Ran Chao for his patience, support and always being there for me. I would also like to thank my lifelong friends, Hao Huang, Jie Xu, Huifang Li, Shuai Zhao, Suwen Wang, Niankun Rao, Cai Long, Cuiling Xu, Ling Zhang, Yang Cao, Wu Yang Jin, Cong Gao, Qi Qian, Eric Yao, Sida Zhou, Xing Chen, John Chaotai Liang, Kajin Lee, Jannu Casanova Moreno and so many of them that I can never express my thanks enough.

Last of all, I am very grateful to my parents, to whom I owe the most of love. I would not be whom I am today without their tremendous spiritual care, protection and support throughout these years.
Dedication

For my parents Xingqi Chai and Zhonghua Tan
1 Introduction

In this work, near infrared Raman microscopy was employed to characterize human induced pluripotent stem (hiPS) cells, which are a type of engineered cells that originate from somatic cells but behave similar to human embryonic stem (hES) cells. Raman spectroscopy has shown promise for non-invasively charactering cell status for hES\textsuperscript{[1-9]} and a variety of other cell types\textsuperscript{[10-13]}. Although few differences in, for example, gene expression signature, \textsuperscript{[14]} differentiation abilities\textsuperscript{[15]}, were noticed between hiPS and hES cells using biological methods (e.g. gene expression profiling\textsuperscript{[14]} and DNA methylation mapping\textsuperscript{[16]}), the characteristic Raman spectra of undifferentiated and differentiated iPS cells have not been reported. It is therefore of interest to characterize the Raman signatures of hiPS cells and how they change during differentiation.

For the work reported in this thesis, Raman spectra were collected from hiPS cells and their non-specifically differentiated progeny, and compared to those from hES cells and their non-specifically differentiated progeny. Factors such as cell line-specific features, and culture confluency, which may affect Raman spectra, were also examined. The results revealed considerable similarities between hiPS cells and hES cells but mostly dissimilarities between hiPS cells and differentiated hES cells.

Background information will be presented in this chapter to help readers understand this work and to place it in context with similar studies.

1.1 Raman spectroscopy

1.1.1 Raman scattering

When light strikes molecules, the photons can be absorbed or scattered. If a molecule absorbs the light, it can be excited from the ground state to an excited state when the energy
difference between these two states matches the energy of the incident photon. Absorption spectroscopic methods, such as infrared spectroscopy (IR) and ultraviolet-visible spectroscopy (UV-Vis spectroscopy), are based on this phenomenon and they detect the radiation loss as a function of frequency. Besides being absorbed, some photons are scattered by molecules. In this case, the incident photon interacts with the molecule and distorts the electron cloud; consequently, the molecule is excited to a short-lived “virtual state”, from which a photon is quickly emitted as the molecule relaxes from this virtual state. Most of the scattered photons only interact with the electron cloud, which is relatively light, so the interactions are highly unlikely to change the energy of these photons. This elastic scattering (no net energy transfer) is called Rayleigh scattering. Inelastic scattering (also called Raman scattering) where nuclear motion (vibration) is involved, leads to a frequency shift of the scattered photon (there is a net energy transfer). It is a very low-probability process, only occurring for one per $10^6$-$10^8$ scattered photons.

![Diagram of Rayleigh scattering, Raman scattering and infrared absorption processes](image17)

Figure 1.1 Diagram of Rayleigh scattering, Raman scattering and infrared absorption processes

17
Figure 1.1 shows a schematic energy level diagram of absorption and different scattering types. Because a virtual state is created during the interaction between light and the molecule, its energy is determined by the frequency of incident light. In Rayleigh scattering, the scattered photon comes out with the same energy as the incident one. There are two types of Raman scattering, Stokes and anti-Stokes. If, as a result of the scattering process, the molecule is excited from the ground vibrational state $v_0$ to a higher excited vibrational state $v_1$ by obtaining energy from the photon, it is called Stokes scattering; if, instead, the molecule donates energy to the photon, thus dropping from a higher (i.e. excited) vibrational state down to a lower one, it is called anti-Stokes scattering. The intensities of Stokes scattering and anti-Stokes scattering depend on the population density of the relevant vibrational states. Normally, anti-Stokes scattering is less intense at room temperature because there are fewer molecules in excited vibrational states. Therefore, Stokes scattering is usually measured to obtain Raman spectra. However, in the presence of strong fluorescence interference, which only occurs at the lower energy side (Stokes side), the measurement of anti-Stokes scattering is preferred.

The energy of a vibrational transition between two energy states is reflected in the shift between $v_0$ (the frequency of incident light) and $v_r$ (the frequency of scattered light). This frequency shift is called the Raman shift. Raman spectroscopy is complementary to infrared spectroscopy in terms of the molecular vibrational transitions that they probe. The vibrational modes depend on chemical bonds, therefore Raman spectroscopy can provide information about chemical structure of molecules, their conformations, and intermolecular/intramolecular interactions. It should be noted that not every vibrational mode is Raman active. Only those vibrational modes that change the polarizability of the molecule
can exhibit Raman signal. In conventional Raman spectroscopy, besides the intensity of incident light, the signal intensity also depends on its frequency, varying according to the fourth power of the frequency \( (\nu_0^4) \).

Raman spectroscopy offers unique advantages over other techniques. First, Raman scattering from water is relatively weak, which is an important advantage for biomolecular spectroscopy over infrared absorption spectroscopy where water contributes strong interference due to its high absorbance at IR wavelengths \(^{[18]} \). Another advantage of Raman spectroscopy (RS) over other techniques, such as mass spectroscopy and chromatography, is the feasibility of non-destructive analysis with minimum sample preparation (i.e. no use of dyes, labels or contrast agents). Therefore, it is able to generate fingerprint information rapidly and cheaply. Moreover, RS can be combined with imaging methods such as confocal microscopy\(^{[19]} \) to provide high spatial resolution (\( \sim 1 \mu m \)). It can also be coupled with micromanipulation techniques such as laser tweezers\(^ {20, 21} \) and gain rich information in the single-molecule measurement. A major drawback of Raman spectroscopy is that the Raman signal is usually very weak (compared with fluorescence spectroscopy, for example), and thus can be easily interfered with or even completely obscured by autofluorescence.

1.1.2 Raman signal enhancement strategies

To circumvent the problem of the relatively weak signal, several enhancement methods have been developed, such as resonance Raman spectroscopy (RRS), surface enhanced Raman spectroscopy (SERS), tip-enhanced Raman spectroscopy (TERS), and coherent anti-Stokes Raman microscopy (CARS). In RRS, when the excitation frequency is close to that of an electronic transition, an enhancement of up to six orders of magnitude for a subset of Raman-
active modes is observed. [22] Because the excitation light is tuned to an electronic absorption wavelength, only the vibrational modes in the spectral region where the laser matches appropriate electronic transitions are enhanced. In most applications, only vibrations coupled to the chromophoric group that satisfy the resonance condition are enhanced and therefore they can be observed in preference to those of other groups. However, being unable to see the global picture of the Raman signature also limits the RRS applications to some extent. Additionally when the molecule is excited to a higher electronic state, it can also undergo a fluorescence emission process to the lower electronic state. The suppression of fluorescence is the main challenge in RRS unless the excitation light is at a sufficiently short wavelength that the fluorescence occurs at longer wavelengths than the Raman shifts of interest. Because RRS is normally used in the deep UV region (200-300nm) for biological molecules, fluorescence is largely avoided, however potential UV damage can be a serious problem and can limit the application of RRS. Recent achievements in RRS have been critically reviewed in several papers [22–24].

In surface-enhanced Raman spectroscopy (SERS), an up to 15 orders of magnitude enhancement (compared to the conventional Raman scattering) [25] has been observed for molecules placed on some metal surfaces. This enhancement is mainly generated via two mechanisms: (1) the electromagnetic enhancement, in which the Raman signal is enhanced by a strong surface plasmon electromagnetic field in the vicinity of the surface (2) and a chemical effect, such as a charge transfer between the metal and absorbed molecule that leads to an increase in molecular polarizability. [26] Due to the heterogeneity in microstructures and roughness of the SERS substrate, the enhancement factors are usually difficult to reproduce, leading to variation in the intensity of Raman signal for the same
analyte; this is the main obstacle preventing wider application of SERS for quantitative analysis.

Tip-enhanced Raman scattering (TERS) is able to address the quantitative difficulty of SERS by replacing the treated metal surface with a sharp metal tip. It can be considered a variation of SERS using a single metal tip instead of a rough surface. The electromagnetic field is spatially confined to the apex of the sharp tip providing spatially resolved molecular information on a nanometer scale.\textsuperscript{[27]} The high spatial resolution allows its application in characterizing single-wall carbon nanotubes\textsuperscript{[28]} and single-macromolecular measurements\textsuperscript{[29]}. Additionally, because the enhancement only occurs at the tip, semi-quantitative and quantitative experiments, which are hard to achieve in SERS, become possible in TERS. Nowadays the main challenge in TERS is the maintenance and mass production of durable and reproducible tips\textsuperscript{[27]}.

Coherent anti-Stokes Raman microscopy (CARS) is a four-wave mixing process where a pump (ω_p), a Stokes (ω_s), and a probe beam (ω_pr) are employed to induce a molecular polarization and generate an anti-Stokes radiation (ω_{as} = ω_p - ω_s + ω_pr). The same laser is often used to serve both pump and probe field (ω_pr = ω_p , ω_{as} = 2ω_p - ω_s). When ω_p - ω_s coincides with a molecular vibration of molecule, the CARS signal will be significantly amplified\textsuperscript{[30]} by up to three orders of magnitude. CARS has recently shown promising applications in the field of fast, high-resolution cell and tissue imaging.\textsuperscript{[31–34]}

In addition to RR, SERS, TERS, CARS, there are other techniques based on Raman scattering that have been developed, including, stimulated Raman scattering (SRS)\textsuperscript{[35]}. Furthermore, different collection geometries have been pursued to optimize Raman spectroscopy. In combination with optical fibers\textsuperscript{[36–38]}, such as in spatially offset Raman
spectroscopy (SORS)\(^{39}\), higher sample accessibility for RS is enabled. Simultaneous collection of epi-collected signal and reflected forward collected signal also enhance overall Raman signal by roughly 4 times.\(^{40}\) Other approaches combine RS with optical traps to enable the simultaneous micromanipulation and characterization of molecules on the nanometer scale.

1.2 Applications of Raman spectroscopy in biological systems

The Raman effect was first discovered in 1928\(^{41}\), however it was not until the development of laser technology in last four decades that RS has developed into a powerful analytical tool for the characterization of biological systems.\(^{33, 42–46}\) In the area of life science, the applications of RS have been numerous and diverse, such as analysis of pharmaceutical products,\(^{47}\) identification of microorganisms,\(^{20, 42, 48}\) characterization of mineralized tissues and diagnosis of bone fractures\(^{49}\), analysis of biological fluids (i.e. urine and blood)\(^{21, 50}\), and for disease diagnostics.\(^{51–54}\) Raman spectroscopy/microscopy is ideally suited to investigate biological systems for several reasons. First, the spatial resolution (~1 \(\mu m\)) is on a small enough scale to investigate subcellular organelles. Secondly, biomolecules (such as DNA, RNA, protein, and lipids) exhibit characteristic Raman peaks, therefore external labels are not necessary to observe these species, hence there is no concern that dyes or staining methods may affect the viability and metabolism of investigated cells. In addition, Raman spectra provide fingerprint information that reflects relative composition, structure, conformation and intermolecular information of the molecules in biological systems. However, some potential disadvantages must also be considered in some applications. Laser damage is perhaps the main practical concern. Careful selection of laser wavelength and
power is important. It has been found that near infrared (NIR) radiation (785-1064nm) is relatively less harmful to biomolecules. NIR excitation also generates minimal auto-fluorescence interference and offers maximal penetration depth, which can benefit applications in tissues imaging.

1.3 Human induced pluripotent cells and embryonic stem cells

Embryonic stem cells (ES cells), are pluripotent cells derived from the inner cell mass of early mammalian embryos and are able to replicate indefinitely and differentiate into cells of all three germ layers, consisting of ectoderm, endoderm and mesoderm. Most animal embryos differentiated into these three germ layers and then further differentiate into organs and tissues of the body. Human ES cells have triggered significant interest because of their potential applications in cell-based therapy, drug discovery, and pre-clinical proof-of-principle experiments. However, besides the strong ethical concerns about the destructive use of human embryos, immunologic rejection after transplantation also limits the application of hES cells for cell-based therapeutic applications.

Generation of patient-specific pluripotent cells from mature cells is one potential way to overcome rejection problems. This has recently been achieved by reprogramming of somatic cells into pluripotent cells. Human induced pluripotent cells (hiPSCs) have attracted much attention and interest. Indeed, many questions about the true functional equivalence of iPS and ES cells have emerged. It is therefore important to explore and characterize all of the similarities and differences between iPSCs and ESCs to help researchers better understand the processes of reprogramming and de-differentiation, the nature of induced pluripotency, and to accelerate the development of iPSC applications.
One major difference between iPS cells and ES cells is the introduction and expression of exogenous reprogramming genes in iPS cells, which may cause genetic and/or phenotypic instabilities.\textsuperscript{[67]} Genome-wide DNA methylation patterns of iPS cells also showed some deviations from ES cells\textsuperscript{[68]}.

1.4 Data processing methods

1.4.1 Multivariate methods in cell discrimination

Multivariate methods, including supervised and unsupervised methods, are based on the assumption that the analyte spectra have subtle but reproducible variations associated with changes in analyte concentrations and/or properties\textsuperscript{[69]}. Supervised approaches require reference knowledge of class membership that yields a model for subsequent new data. When reference information is not available, unsupervised approaches are useful to group the data points according to covariance in the data matrix.

An inherent property of cells and tissues is their heterogeneity and complexity. Their Raman spectra often contain many overlapping bands; therefore advanced multivariate analysis strategies are helpful for better data interpretation. It has been shown that, in combination with multivariate methods, such as principal component analysis (PCA)\textsuperscript{[70–72]}, linear discriminant analysis (LDA),\textsuperscript{[73]} and hierarchical cluster analysis (HCA)\textsuperscript{[9, 74]}, amongst others, Raman spectroscopy becomes a very promising fingerprinting tool for cell discrimination. Indeed, such combinations have been used in a significant number of studies to distinguish between normal and various pathological states.\textsuperscript{[75]} One of the most common unsupervised methods is PCA. It is often used to examine ‘natural differences’ by virtue of identifying independent variances and random fluctuations in spectra\textsuperscript{[76]}. In PCA, the data
matrix $\mathbf{X}$ is decomposed into several principal components (PCs). The result of PCA is a product of PC scores $\mathbf{T}$ and PC loadings $\mathbf{P}$ matrices plus the residue $\mathbf{E}$: \[ \mathbf{X} = \mathbf{TP}^T + \mathbf{E} = t_1p_1^T + t_2p_2^T + \ldots + t_Ap_A^T + \mathbf{E} \]
in which $\mathbf{A}$ is the number of computed PCs. When dealing with spectroscopic data, $\mathbf{X}$ is an $M \times N$ matrix that contains the intensities from $M$ spectra; the column number $N$ is the number of wavenumber points for each spectrum. The orthogonal PCs, that each assemble correlated variances, are calculated and ranked by contributions to the total variance. The latter are provided by scalars computed during the singular value decomposition of $\mathbf{X}$, an important step in PCA, and can be seen as PC eigenvalues. Thus, the first few PCs capture the most variance in the spectral data set, while the latter PCs are most likely to contain random error. Furthermore, each spectrum can be approximated by a linear combination of a few PCs, with the expansion coefficients, called “scores”, indicating the contribution of each PC to each spectrum. In this way, similar spectra will exhibit a set of similar scores and form a cluster in a PC scores plot, and can be distinguished from those having different PC scores.

1.4.2 The student’s $t$-test

The $t$-test is a statistical test using the $t$-distribution. In this work, we employ the commonly used Student’s $t$-test to compare the means of two samples. The statistical null hypothesis is that the means of two measurements from two samples, drawn from normally distributed populations, are equal while the experimental hypothesis is that they differ. The test statistic, $t_{\text{calc}}$, is obtained from

\[
t_{\text{calc}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]

(Eq. 1)
where $\bar{x}_1$ and $\bar{x}_2$ are the means of the two samples, $n_1$ and $n_2$ the numbers of measurements for each sample, and $\sigma_1$ and $\sigma_2$ the standard deviations. The probability $p$ is analyzed which means, for example, when $p=0.05$, that as large a difference as that observed between the means $\bar{x}_1$ and $\bar{x}_2$ would only be expected in 5% of occasions. In other words, when

$$t_{\text{calc}} > t_{\text{critical}}$$

($t_{\text{critical}}$ is the critical value under certain degrees of freedom and the selected significance level $\alpha$, usually $\alpha=0.05$), or $p < \alpha$, the null hypothesis is not considered tenable, and is thus rejected; in practical terms the two samples are considered to have statistically different means and the experimental hypothesis is considered tenable. Otherwise, the null-hypothesis is tenable, and the two samples are not considered to have statistically different means. Because large differences between sample means could nevertheless be observed by chance, leading to incorrect rejection of the null hypothesis, $p$-values have to be adjusted when performing multiple $t$-tests to avoid the accumulation of such chances and erroneous rejection of the null hypothesis. The easiest, and a conservative way, to guard against erroneous null hypothesis rejection is to divide the required significance level by the number of $t$-tests performed, an approach known as the Bonferroni adjustment.\(^{[78]}\)

### 1.5 Previous work

Characterization of human stem cells and stem cell populations has been performed by a number of different techniques. Each of them has certain pros and cons, and provides a specific type of information. One of the common techniques is to assess the expression profile of various genetic markers (e.g. Oct, SSEA3/4 or Nanog) and validate the cells’ ability to differentiate into the derivatives of all three germ layers.\(^{[79, 80]}\) However, these existing biological methods, i.e. immunocytochemistry and fluorescence activated cell
sorting (FACS), are time-consuming, destructive and laborious and typically require a large cell population. There are other genomic and proteomic methods available\textsuperscript{[79, 80]}, but they also suffer from similar problems.

As mentioned, the Raman spectral "fingerprint" of cells can be used to characterize the status of cells under particular conditions. This fingerprint may change over time with cellular composition that originated from cell-specific metabolic processes. It hence provides additional information that can be used to distinguish certain cells from other cells. It is of interest to characterize overall Raman signatures of cells and interpret, at least in a general sense, the sources of variances in the spectra. In the case of hES cells, some studies have been completed toward understanding the typical Raman features and how they vary under normal maintenance conditions,\textsuperscript{[3]} with temperature variations,\textsuperscript{[7]} and subsequent to a differentiation stimulus.\textsuperscript{[1, 3]} So far, there are no Raman spectral studies on induced pluripotent cells, which could be considered as an artificial counterpart of hES cells.

1.6 Thesis content and organization

In this study, we obtained Raman spectra, excited at 785 nm, of hiPS cells, hES cells and differentiated hES cells. Results showed that by using multivariate analysis it was possible to find out the similarity and subtle differences between hiPS cells and hES cells, and discriminate hiPS cells and differentiated hES cells (Chapter 2). Following these results, the influences of different factors on the discrimination of these cells are discussed in Chapter 3. The Raman signatures of different hiPS cell lines was investigated, with results indicating that this did not influence the chemometric discrimination. Then, spectral variations caused by different cell confluencies were examined, leading to the conclusion that
these two factors do not change the similarity between hiPS cells and hES cells, or the
dissimilarity between iPS cells and differentiated hES cells. Chapter 4 presents a Raman
study of differentiated hiPS cells, and several attempts were made to discriminate their
Raman spectra from those from undifferentiated iPS cells. Chapter 5 provides the
conclusions from this work and suggests future investigations in this area.
2 A comparative study of hiPS cells with hES and differentiated hES cells using Raman microspectroscopy

2.1 Introduction

Human embryonic stem (hES) cells can be cultured and expanded for many passages in vitro, without losing their ability to differentiate into three embryonic germ layers. The same is true for induced pluripotent stem (iPS) cells, which are obtained by reprogramming somatic cells using ectopic expression of the transcription factors Oct4, Sox2, Klf4 and c-Myc\[61\] or alternative reprogramming cocktails. These iPS cells have very promising applications in development of patient-derived expandable cell source\[81\], cell-based therapies\[82\], and drug screening\[81\] with reduced ethical concerns. All of these applications require the selection and characterization of cells lines that reliably, efficiently and stably differentiated into disease-relevant cell types. For the maintenance purpose, it would be important to monitor the pluripotent status of these hES and hiPS cells. Also, a convenient analytical method is needed to measure the differentiation states of hES and hiPS cells.

Raman spectroscopy (RS) is a vibrational technique that measures the inelastic scattering of incident photons by molecular bonds. Thus, Raman scattered photons exhibit frequency shifts with respect to the incident photons corresponding to the energies of allowable vibrational modes of the sample. The spectrum of Raman shifts can serve as a "fingerprint" of the molecular composition and structure of the sample. RS can be conveniently implemented using a microscopy platform for the excitation and collection optics to obtain spatially resolved spectral data at high magnification. This so-called "Raman microspectroscopy" (RµS) has become a widely used tool for investigating living cells due to a combination of features that include: high spatial (~1 μm) and spectral resolution (< 0.1cm^{-1}), \[12, 83\] the ability to perform relatively fast (few minutes), non-invasive, and label-
free measurements, especially in the near infrared region where there is little evidence that the excitation causes cellular damage,[1, 8, 84] and very weak interference by water bands. RμS has been used to aid investigation of mouse[1, 85] and human[1, 8] ES cells, but thus far there have been no Raman studies reported involving iPS cells.

An increasing number of differences between hESCs and hiPSCs have been reported, including evidence of genetic instability,[67] as well as epigenetic differences leading to differential gene silencing in hES cells and hiPS cells suggesting hiPS cells retain a transcriptional memory of the original cells. [67, 68, 86] Hence, even in the pluripotent state, we hypothesize that hiPS cells may exhibit detectable and meaningful spectral differences compared with hES cells. It is known, for example, that differentiated hES cells have larger protein to nucleic acid ratios than hES cells, which are clearly reflected in the comparatively stronger protein-related Raman bands in differentiated hES cells. [1] It is therefore of interest to determine whether hiPSCs have reduced protein-related bands. Specifically, one could ask whether the Raman signatures of hiPS cells are qualitatively more similar to those they are designed to resemble (pluripotent hES cells) or more similar to those that they were derived from (mature cells). In this chapter, we report the Raman spectral signatures of hiPS cells, and the variances observed within these signatures, with particular emphasis on the comparison with hES cells.

2.2 Material and methods

2.2.1 Cell culture

The EOS hiPSC line was generously provided by Dr. James Ellis (The Hospital for Sick Children, Toronto, ON, Canada). This cell line was created using the EOS (Early
Transposon promoter and Oct-4 and Sox2 enhancers) lentiviral vector containing an EGFP-IRES-\textit{Puro}^{R} sequence serving as a fluorescence reporter of pluripotency (and puromycin resistance to aid selection).  \textsuperscript{24} These cells were maintained with mTeSR1 medium (STEMCELL Technologies, Vancouver, BC, Canada) containing an additional 1 \( \mu \)g/mL puromycin (Invitrogen, Burlington, ON, Canada) on Matrigel (BD Biosciences, Mississauga, ON, Canada) coated tissue culture dishes with a daily medium change. The hiPSCs were passaged when 85-95\% confluent every 10-12 days by using accutase (Sigma–Aldrich, Gillingham, UK).

In order to estimate the doubling time of hiPSCs, 8 tissue culture wells seeded with approx. 150,000 hiPSCs were prepared. The cells were maintained at 37\degree C in a humidified incubator with a daily medium change. At six different time points during the exponential growth phase, one well was treated with trypsin (Invitrogen) for 10 min and the cells counted.

The CA1 hESC line was graciously provided by Dr. Andras Nagy (Mount Sinai Hospital, Toronto, ON, Canada) and maintained in mTeSR1 medium on Matrigel-coated (BD Biosciences) tissue culture dishes with a daily medium change. The cells were passaged every 6 or 7 days when 85-95\% confluent using a combination of dispase (STEMCELL Technologies) and physical sheer to detach hESCs aggregates according to the manufacturer’s protocol. Differentiated CA1 cells were obtained by culture in DMEM/F12 medium with 10\% fetal bovine serum (FBS) for 20 days.

Cell maintenance cultures were plated on tissue culture dishes (Starstedt, Newton, NC); for Raman spectroscopic use, cells were plated on mirrors seeded with a starting cell count of about 20,000 (hESCs) or 50,000 (hiPSCs). The mirrored substrates were used in
order to improve the Raman signal as previously described.[88] Each mirror consisted of a glass-encapsulated 100 nm gold thin-film on 12.7-mm diameter, 6-mm thick, glass discs (ThorLabs Inc, Newton, NJ).

2.2.2 Raman microspectroscopy of hiPS and hES cells

Raman spectra were obtained at room temperature from cells growing on gold mirrors using a Raman microscope system (RM 1000 for differentiated hES cells, inVia for all others, Renishaw, Gloucestershire, U.K.) equipped with a Renishaw 785nm diode laser. The laser output power varied from 195-205 mW for this series of experiments. Prior to Raman analysis, maintenance medium was replaced by basal medium (DMEM/F12) to reduce background interference from culture medium (mTeSR). For Raman measurements, a mirror immersed in a 35mm Petri dish filled with basal medium was placed under the microscope. A 40x water immersion objective was used to focus the laser beam on the cells on the mirror, which was placed on a custom-made temperature controller to keep the temperature at 37°C. About 15-20 spectra were taken from each mirror, with up to a 100 s exposure time/spectrum, covering the range from 687cm⁻¹ to 1073cm⁻¹.

2.2.3 Fluorescence microscopy

While in the pluripotent state, the hiPS EOS line expresses the EGFP-IRES-\textit{Puro}^R sequence, but upon differentiation, EGFP-IRES-\textit{Puro}^R expression is extinguished.\textsuperscript{24} Thus, pluripotent cells/colonies are marked with EGFP (enhanced green fluorescent protein) and are easily identified using fluorescence microscopy. EOS-EGFP expression by hiPSCs cultured in our lab was confirmed with fluorescence microscopy. Live cell images were captured using Leica Application Suite software (LAS AF6000, Leica Microsystems,
Weitzlar, Germany) that was integrated with the Raman microscope system. After Raman spectra were taken, the basal medium (DMEM/F12) was replaced with phosphate buffered saline (PBS) to eliminate the strong fluorescence background from the basal medium. Using a 40x water immersion objective, bright-field (3ms-8ms exposure times) and fluorescence (100ms-300ms exposure times) cell images were taken. In both cases, the exposure time was determined automatically by the acquisition software (LAS AF, Leica Microsystems, Weitzlar, Germany) to achieve the best image quality.

Figure 2.1 Left: GFP expression of hiPS cells. Right: Bright field image of the same growth area. Images were captured after the Raman signal acquisition from hiPSCs 3 days after sub-culturing. Before taking the GFP image, the basal medium for Raman spectroscopy was replaced by PBS in order to reduce the background fluorescence emission.

2.2.4 Data analysis

Background spectra (Figure 2.2) were taken from cell-free regions on the mirror before and after data collection from the cell growth areas. The average background spectrum for each mirror was then subtracted from every individual spectrum from each mirror (Figure 2.3). Each spectrum was then smoothed by an automated smoothing filter\cite{89} using 3 iterations, and a moving average, peak stripping, semi-automated procedure\cite{90} was used to
flatten the baseline. After pre-processing, spectra were of high quality and peaks of different cellular components were distinguishable (Figure 2.4). Spectra were processed using Matlab 7.1 (The Mathworks, Natick, MA), compared based on their means and standard deviations, and further analyzed using PCA. Excel 2007 (Microsoft) was used for t-tests.

Figure 2.2 Sample raw spectra from hiPS cells (3 days after sub-culturing on the mirror, same set of data was used in the following comparisons). The colored spectra were collected from different spots in a growth area; the bottom bold one was the average of 4 spectra from the cell-free area (2 before the cell spectra collection and 2 after).
Figure 2.3 Background subtracted spectra using the raw spectra in Figure 2.2

Figure 2.4 Baseline flattened and smoothed spectra using the background subtracted spectra in Figure 2.2. Three spectra were rejected as indicated by the arrows.
2.3 Results and discussion

2.3.1 Comparison between hiPS and hES cells

Raman spectra of hiPSCs were highly similar to hES cells (Figure 2.5). Specifically, the mean spectra of hiPS cells and hES cells (normalized to overlapping nucleic acid bands near 783 cm\(^{-1}\)) were nearly indistinguishable. In fact, with the exception of the 700 cm\(^{-1}\) lipid and 852 cm\(^{-1}\) glycogen bands, the mean intensities of major Raman peaks from hiPS and hES cells were statistically not different (Table 2.1). Detailed peak assignments can be found in Appendix.

![Raman spectra of hiPSCs and hES cells](image_url)

**Figure 2.5** Means of normalized Raman spectra, taken from living hES (N=49) and living hiPS cells (N=17) were highly similar.

Note also that, although the intensities of the 783 cm\(^{-1}\) bands were equalized because of the normalization, the peak profiles were nevertheless quite similar. 783 cm\(^{-1}\) is a combined band of 782 cm\(^{-1}\) (C\(_5\)'-O-P-O-C\(_5\)') and 788 cm\(^{-1}\) (thymine, uracil, and cytosine). The ratio between the intensities of these two bands, which were used to reflect the structure of
chromatin\cite{93}, will change the profile of 783 cm\textsuperscript{-1} peak. Thus, structural differences between hiPSC and hESC chromatin were not evident at this level of analysis. Because of the overall similarity between the spectra, the Raman differentiation state markers were also similar. The previously characterized\cite{1} R4 marker (the ratio of the 757 cm\textsuperscript{-1} band intensity to the 783 cm\textsuperscript{-1} band intensity) was 0.46 for hiPS cells with a 95\% confidence interval (CI) of 0.44 - 0.50 (n=17) in agreement with the R4 ratio of 0.46 (n=48) and 95\% CI of 0.43 -0.48 for hESCs. In differentiated cells, this marker is typically ~ 1 or larger\cite{1} (further comparisons with differentiated cells follow below). Besides the bands discussed above, there are a few small peaks, i.e 984 cm\textsuperscript{-1}, between 783 cm\textsuperscript{-1} and 852 cm\textsuperscript{-1}, as well as 852 cm\textsuperscript{-1} and 937 cm\textsuperscript{-1}. However the intensities of these peaks are very subjective to baseline flatten technique, thus not as reliable as other peaks.

Table 2.1 Two-sample t-test results generally show the intensities of major peaks in Raman spectra of hiPSC and hESC cells to be not statistically different. Because multiple t-tests were used, the $p$-values were based on the Bonferroni inequality\cite{78}.

<table>
<thead>
<tr>
<th>Raman peak</th>
<th>Assignment</th>
<th>Mean ± Standard error</th>
<th>hiPS</th>
<th>hES</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>700 cm\textsuperscript{-1}</td>
<td>Lipid</td>
<td>0.217±0.076</td>
<td>0.138±0.045</td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>717 cm\textsuperscript{-1}</td>
<td>Lipid</td>
<td>1.05±0.13</td>
<td>0.92±0.11</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>757 cm\textsuperscript{-1}</td>
<td>Protein</td>
<td>0.450±0.052</td>
<td>0.459±0.082</td>
<td></td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>852 cm\textsuperscript{-1}</td>
<td>Glycogen and protein</td>
<td>0.89±0.12</td>
<td>1.05±0.30</td>
<td></td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>937 cm\textsuperscript{-1}</td>
<td>Glycogen</td>
<td>1.13±0.17</td>
<td>1.27±0.46</td>
<td></td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>1002 cm\textsuperscript{-1}</td>
<td>Protein</td>
<td>2.94±0.28</td>
<td>3.02±0.20</td>
<td></td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>1031 cm\textsuperscript{-1}</td>
<td>Protein</td>
<td>0.622±0.062</td>
<td>0.629±0.080</td>
<td></td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>
The dynamic properties of hES cells, rather than measurement error, are the main determinants of their standard deviations.\textsuperscript{[1]} The standard deviations of the normalized spectra for both hiPS cells and hES cells (Figure 2.6) showed the major spectral variances to be related to the 850 cm\(^{-1}\), 861 cm\(^{-1}\) and 938 cm\(^{-1}\) glycogen bands. Glycogen variances are known to occur under normal maintenance conditions in hES cells.\textsuperscript{[3]} However, greater variances in glycogen peaks suggested a faster changing glycogen concentration in the hES cells compared to the hiPS cells. In hES cells, glycogen band intensities and their variances have been related to growth rates.\textsuperscript{[1]} Two trials of growth curve measurements show closely matched results, which also indicates the relative stability of hiPS cells within a short period in terms of growth rate. The doubling time of the hESCs (approx. 20h\textsuperscript{[1]}) is shorter than that of iPS cells (approx.23h, Figure 2.7), and this higher proliferation rate is likely to be the main cause of the higher standard deviations of hESCs relative to hiPSCs in terms of the glycogen bands.

Figure 2.6 The standard deviations of normalized Raman spectra, taken from living hES (N=49) and hiPS (N=17) cells, show reduced metabolic activity of hiPS cells compared to hES cells.
Figure 2.7 Growth curves of hiPS cells, (a) a growth curve generated from a seeding density of 3.69×10^5 cells/well. Exponential regression result is based on the 2nd to 4th data points. The calculated doubling time is 22.0±2.9 hours. (b) a growth curve generated from a seeding density of 1.50×10^5 cells/well. Exponential regression result is based on the 2nd to 6th data points. The calculated doubling time is 24.0±2.2 hours.

PCA was performed here on the combined normalized spectra of the hiPSCs and hESCs. The first PC (PC1; Figure 2.8 a) closely matched the Raman spectra from solid D-glycogen (Figure 2.8 b). In other words, it confirmed that the largest independent variance for both the hiPSCs and hESCs derived from the changing glycogen concentrations. Although proteins are the most abundant species in cells, and they dominate the mean spectra (Figure 2.5), the protein variance was not as remarkable as that of glycogen. However, the
second PC (PC2; Figure 2.8a) of the combined data showed contributions from protein (ca. 1002 cm$^{-1}$) and lipids (699 cm$^{-1}$, 718 cm$^{-1}$ and 983 cm$^{-1}$) bands. The RNA band 814 cm$^{-1}$ in PC2, shows a variance in the ratio of RNA/DNA in the data set, while the normalization offsets the difference at the 783 nucleic acid band. Finite data points were taken from two groups of cells, and that induced a spectral offset (~0.1 cm$^{-1}$-0.5 cm$^{-1}$) between the two. When combining the smoothed data sets, this offset is manually zeroed. It may therefore explain some of the small shifts (within 1 cm$^{-1}$) observed at several bands in Figure 2.8a. Since the position of a Raman band is also affected by secondary structure, the small shifts present at several bands may be caused by the natural differences in composition or conformation of cellular constituents. PC1 and PC2 assembled most of the variance in the data set (86% of total variance).
Figure 2.8 Glycogen metabolism was the major contributor to spectral variances as shown by the first principal component (PC1) of the combined data. The second principal component (PC2) was more complex and included lipid and protein-related peaks. (a) PC1 and PC2 of the combined normalized spectra of hiPSCs and hESCs. (b) First principal component (PC1) of the combined data from hiPS cells and hES cells, compared with Raman spectra from solid D-glycogen.

When plotting the PC2 vs. PC1 scores of the spectra (Figure 2.9), it was apparent that the data points in Figure 2.9 tended to segregate into two adjacent clusters with each cluster populated mostly by a single cell type. Thus, subtle differences may have existed between the specific hiPSC and the hES cell lines used here, allowing some degree of discrimination between these groups. The result is in agreement with results from genetic assays.\[16\] On the
PC1 dimension, which stands for the glycogen variance, hES cells show a more dispersed population than hiPS cells. It might be also explained by the faster dividing rate of hES cells thus caused a more heterogeneous glycogen distribution across the growth area. Most of the data points of hiPS cells fall on to the positive side of PC2, in contrast to most of the hES cells on the negative side. As the PC2 mainly resembles protein and lipid features, thus it reflects that the variance of Raman spectra of hiPS cells and hES cells correlate differently to protein and lipid. This will be discussed further below.

Figure 2.9 A PC2/PC1 score plot revealed a tendency for hiPS and hES cells to cluster independently along the PC2 dimension. Thus, the Raman spectra of hiPS cells fairly resembled undifferentiated hESCs of the hESC line, but subtle differences nevertheless could be detected.

Overall, these results showed the similarities between the Raman spectra of hiPS cells and undifferentiated hES cells and that spectral variances of hiPS cells were consistent with the variances of undifferentiated hES cells (but less robust than those of hES cells). Nevertheless, small differences, captured primarily by PC2, were evident but it is unclear whether these differences were due to effects related to the reprogramming process or whether they only reflected the differences between two cell lines. Because there is a lot
more heterogeneity among different iPS cell lines (different source cells, different derivation methods) than among ES cell lines, another hiPS cell line, MSCiPS1 cell line was also used to make further comparisons in Chapter 3.

2.3.2 Comparison between undifferentiated hiPS cells and differentiated hES cells

The results above were further confirmed via the comparison of normalized spectra from undifferentiated hiPSCs with normalized spectra from hES cells after 20 days of growth in media stimulating non-specific differentiation. Mean spectra (Figure 2.10 A) also showed clear differences between hiPS cells and differentiated hES cells. All the mean intensities of hiPS and hES cell peaks were statistically different (Table 2.3). Thus, hiPS cells did not exhibit the spectral features typical of a mixed population of fully differentiated hES cells. Differentiated hES cells were clearly dominated by protein (757 cm\(^{-1}\), 876 cm\(^{-1}\), 1003 cm\(^{-1}\), 1032 cm\(^{-1}\)) and lipid bands (699 cm\(^{-1}\), 717 cm\(^{-1}\)) and, although normalized to the overlapping nucleic acid band near 783 cm\(^{-1}\), showed a somewhat different shape for this band as well. Furthermore, the standard deviations of Raman spectra from differentiated hES cells (which differed remarkably from those of hiPS cells, showing a much higher variance in the spectra) were attributed to the diverse cells types after differentiation and their diverse metabolisms (Figure 2.10 B).
Figure 2.10 Comparison between hiPS cells and differentiated hES cells. A) The means of normalized Raman spectra, taken from living differentiated hES (n = 10) and living hiPS cells (n = 17) were notably different. B) Similarly, standard deviations showed marked differences between hiPS and differentiated hES cells. C) The spectral variances as shown by the first two (PC1, PC2) principal components of the combined data were dominated by protein and lipid-related bands. D) A PC2/PC1 score plot revealed a tendency for hiPS and differentiated hES cells to cluster independently along the PC1 dimension. The dispersion of PC score data points from spontaneously differentiated hES cells likely reflected a variety of cell types after non-specific differentiation. Thus, the Raman spectra of hiPS cells, which closely resembled undifferentiated hES cells, were markedly different from those of differentiated hES cells.
Table 2.2 Two-sample $t$-test results show significant differences for the intensities of major peaks in Raman spectra of hiPSC and differentiated hES cells. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.\[78\]

<table>
<thead>
<tr>
<th>Raman peak</th>
<th>Assignment</th>
<th>Mean ± Standard error</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hiPSC</td>
<td>differentiated hES</td>
<td></td>
</tr>
<tr>
<td>700 cm$^{-1}$</td>
<td>Lipid</td>
<td>0.217±0.076</td>
<td>0.89±0.33</td>
</tr>
<tr>
<td>717 cm$^{-1}$</td>
<td>Lipid</td>
<td>1.05±0.13</td>
<td>2.71±0.63</td>
</tr>
<tr>
<td>757 cm$^{-1}$</td>
<td>Protein</td>
<td>0.450±0.052</td>
<td>1.51±0.55</td>
</tr>
<tr>
<td>852 cm$^{-1}$</td>
<td>Glycogen and protein</td>
<td>0.89±0.12</td>
<td>2.49±0.71</td>
</tr>
<tr>
<td>937 cm$^{-1}$</td>
<td>Glycogen</td>
<td>1.13±0.17</td>
<td>3.25±0.96</td>
</tr>
<tr>
<td>1002 cm$^{-1}$</td>
<td>Protein</td>
<td>2.94±0.28</td>
<td>10.72±3.00</td>
</tr>
<tr>
<td>1031 cm$^{-1}$</td>
<td>Protein</td>
<td>0.622±0.062</td>
<td>1.82±0.53</td>
</tr>
</tbody>
</table>

A PCA on the combined data (hiPS cells and differentiated hES cells) revealed that both PC1 and PC2 were dominated by protein and lipid-associated Raman bands (Figure 2.10C). Additionally, the PC1 resembled most of the spectral features of the standard deviation for differentiated hES cells (Figure 2.11). Thus, the major variances in this data set were attributed to results of the differentiation.
Furthermore, the PC2/PC1 score plot (Figure 2.10D) formed two well-separated clusters that each consisted of a single cell group. In addition, the differentiated hES cells cluster was much more widely dispersed than the hiPS cell cluster. This dispersion was also consistent with the non-specific differentiation stimulus giving rise to various phenotypes. The data points of hES and hiPS cell group differently, mainly on the PC1 dimension.

Because the separation between the differentiated hES cell cluster and the hiPS cell cluster (Figure 2.10D) was more evident than that between the hES cell cluster and the hiPS cell cluster (Figure 2.9), we conclude that, in terms of Raman features, the hiPS cells were spectropically closer to hES cells than to differentiated hES cells.

2.4 Conclusion

The presented content in this chapter is the first spectroscopic comparison between hiPS and hES cells. Raman microspectroscopy was used to characterize the spectral features
of hiPS cells and to compare them to those of hES cells and non-specifically differentiated hES cells. This comparison demonstrated that (1) hiPS cells are spectroscopically similar in terms of relative peak intensities and variances in peak intensities, to pluripotent hES cells, and (2) that hiPS cells are spectroscopically distinguishable from hES-derived differentiated cells using the same criteria as used to distinguish hES cells from differentiated hES cells.

Thus, reprogramming produced a reduction in the protein to nucleic acid ratio of reprogrammed somatic cells to bring this ratio to the levels characteristic of embryonic pluripotent cells. The implication is that either the reprogramming process or the effect of any genetic instabilities or epigenetic differences, relative to ESCs, lead to changes in overall steady-state protein expression levels.

Although reprogramming does not result in phenotypic differences compared to ESCs that are obvious in the raw spectra, subtle differences between the iPSCs and hESCs used in this study could be detected using PCA. This is an intriguing result in light of emerging evidence that iPSCs do not necessarily respond to differentiation stimuli in the same way as hESCs. However, our observed spectroscopic differences may be a cell line artefact and a study with broader objectives would be needed to confirm this. In any case, future work should attempt to elucidate what causes the subtle spectral differences between hiPSC and hESC. It would also be useful to examine further how the Raman signatures of hiPSCs differ from those of hESCs under different conditions (e.g., temperature, differentiation stimuli, etc.), especially if they could be correlated with genetic and epigenetic factors. Nevertheless, the present results clearly demonstrated the similarity in the Raman signatures of the pluripotent cells, thus providing further evidence of the utility of Raman microscopy to
discriminate non-invasively between pluripotent cells, whether embryonic or induced, and cells differentiated from hESCs.
3 Factors that influence the discrimination between hiPS cells, hES and differentiated hES cells

3.1 Introduction

3.1.1 Difference between cell lines

A cell line is an culture established relatively stable cell culture of animal or plant origin that is able to proliferate for a long time when appropriate medium and other environmental conditions are provided[94]. Cell lines of the same cell type deviate in many aspects, for example, apoptosis behaviors[95], control mechanism of cell cycle progression[96], cellular pathways (ATP synthesis, cell signaling, etc)[97] and so on, which eventually show up as differences in metabolism or, more generally, their detailed chemical composition. There are many established hES cell lines derived from different sources and approved (in Canada) for research. Although all hES cell lines share the expression of characteristic pluripotency markers, many differences are emerging between lines, including specific pluripotency marker molecules, transcriptional profiles, genetic and epigenetic stability.[98] These variations may be associated with the wide range of culture conditions or inherent genetic variation.

It was reported that epigenetic and transcriptional variations existed among human pluripotent (hES and hiPS) cell lines using genetic assays.[16] In fact, more significant variation has been observed among various hiPS cell lines than among hESC lines.[16] Besides genetic methods, analytical methods have also been broadly exploited to investigate cell-line-specific features. These methods are relatively rapid, convenient and robust, and able to achieve in vitro/in vivo measurements; however, in most cases they are not able to
obtain detailed genetic/phenotypic information. They are, to some extent, all limited by the inability to fully distinguish among complex biomolecules.

Several spectroscopic studies have been performed in this area. FTIR results indicated that the discrimination between two prostate cell lines was mainly due to inherent biochemical differences between the two cells lines rather than different growth media or different nuclear-to-cytoplasm ratios\cite{99}. Raman spectroscopy was employed to distinguish between cancerous and non-cancerous cell lines\cite{100}, and identify a cell type in a mixed cell population via its spectral signature\cite{101}. Using a hybrid multivariate method of PCA (Principal component analysis) and LDA (Linear discrimination analysis),\cite{102} this technique can accurately differentiate among prostatic adenocarcinoma (CaP) cell lines of varying degrees of biological aggressiveness, based on the difference of glycogen and nucleic acid concentration.

Considering the existence of cell-line specific features, in this Chapter a different hiPS cell line, MSCiPS1, was investigated in order to provide evidence that the findings reported in Chapter 2 are not merely attributable to cell line differences. Rather, it is shown here that at least one other hiPS cell line in addition to the EOS-iPS cell line investigated in Chapter 2, exhibits similarly distinguishable spectral features when compared to hESC and differentiated hESC.

### 3.1.2 Effect of cell density on Raman spectra

After being sub-cultured (i.e., a new culture started from a low density of cells harvested from a previous culture), cells go through three growth phases. First, during the lag phase, the cells adapt themselves to the growth condition and the cell number relatively remains constant. After that, in the exponential phase, the cell number increases
exponentially with time and most of the cells are in the S or G₂ phases of the cell cycle. The culture then enters the plateau phase where the maximum cell number in the culture has been reached and the proliferative rate drops to almost zero. During this phase, most of the cells remain in the G₁/G₀ phase of the cell cycle\textsuperscript{[103]} For anchorage-dependent cells that do not form multi-layers, this phase is where the cells approach a confluent monolayer. The culture will end in a death phase after plateau phase if the space and medium condition become limited. A plot of cell number or cell density versus time is called growth curve.

Spectroscopic techniques have been used to investigate phase-specific features of cell growth. IR spectra were captured from mammalian cells at different growth stages\textsuperscript{[104]} (exponential and plateau stages), showing significant changes in major biochemical components (nucleic acids, glycogen, protein and lipids) by decomposing the cell spectra into a linear combination of spectra from each single type of component. Different Raman spectra were observed from cancer cells in the plateau and exponential phases of growth.\textsuperscript{[105]} These differences included a reduced lipid concentration and an increase in protein and RNA concentration for non-tumorigenic exponential cells. Indeed, biochemical differences introduced by cells growing in cultures of different confluency (in other words, for anchorage-dependent cells, different positions on the growth curve) are related to the metabolism of the cells and would be expected lead to variability in the Raman spectra\textsuperscript{[106]}.

Therefore, it is important to make sure that the similarity between Raman spectra of hiPSCs and hESCs, as well as the dissimilarity between hiPSCs and differentiated hESCs, are observable regardless of culture confluency. Because the cell cultures used in this study were not synchronized in terms of the cell cycle, the culture phase actually reflects a distribution of cells in each phase of cell cycle. Raman spectra from cultures at different
confluencies (c.a. 5%-90%) were collected from the hiPSCs (EOS line), and again compared with those from hESCs, and differentiated hESCs. It was found that, although the culture confluencies had an effect on the Raman spectra, the Raman spectra all showed a high degree of similarity to those from hESCs and distinct from those of differentiated hESCs.

3.2 Material and methods

The MSCiPS1 cell line was generously provided by Dr.George Q.Daley (Children’s Hospital Boston, MA). This cell line was derived from mesenchymal stem cells by ectopic expression of four transcription factors (Oct4, Sox2, Klf4 and Myc)[64]. These cells were maintained with mTeSR medium (STEMCELL Technologies, Vancouver, Canada) on Matrigel (BD Biosciences, Mississauga, ON, Canada, diluted to 0.083 mg/mL in mTESR) coated tissue culture dishes with a daily medium change. Non ES-like growth areas were selectively removed from the culture every day before the medium change. The MSCiPS1 cells were passaged when 70% confluent every 10-12 days using dispase (Sigma-Aldrich, Gillingham, UK).

Maintenance conditions and protocols for EOS hiPS and CA1 hES cell lines were the same as described in Chapter 2. Raman measurements and data analysis procedures are also the same as described in Chapter 2.

3.3 Results and discussion

3.3.1 Raman spectra of MSCiPS1 cells

In order to ascertain whether the difference between hiPS cells and hES cells is more significant than the difference between two iPS cell lines, Raman spectra of EOS and
MSCiPS1 cells were compared. The mean spectra were normalized to the nucleic acid peak at 783 cm$^{-1}$ as in Chapter 2 (Figure 3.1).

![Raman spectra comparison](image)

**Figure 3.1** The means of normalized Raman spectra, taken from living hiPS EOS (n = 17) and MSChiPS1 cells (n = 27) were highly similar.

**Table 3.1** Two-sample $t$-test results generally show the intensities of major peaks in Raman spectra of EOS hiPS cells (n=17) and MSCiPS1 cells (n=26) to be not statistically different. Because multiple $t$-tests were used, the $p$-values were based on the Bonferroni inequality.$^{[78]}$

<table>
<thead>
<tr>
<th>Raman peak cm$^{-1}$</th>
<th>Assignment</th>
<th>Mean ± Standard error</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>Lipid</td>
<td>0.217±0.076</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>717</td>
<td>Lipid</td>
<td>1.05±0.13</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>757</td>
<td>Protein</td>
<td>0.450±0.052</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>852</td>
<td>Glycogen and protein</td>
<td>0.89±0.12</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>937</td>
<td>Glycogen</td>
<td>1.13±0.17</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>1002</td>
<td>Protein</td>
<td>2.94±0.28</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>1031</td>
<td>Protein</td>
<td>0.622±0.062</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>
Two-sample t-test results generally show the intensities of major peaks in Raman spectra of MSCiPS1 (n=26) and hES cells (n=48) to be not statistically different. Because multiple t-tests were used, the p-values were based on the Bonferroni inequality.\[^{78}\]

<table>
<thead>
<tr>
<th>Raman peak cm(^{-1})</th>
<th>Assignment</th>
<th>Mean ± Standard error MSCiPS1</th>
<th>Mean ± Standard error hES</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>Lipid</td>
<td>0.235±0.089</td>
<td>0.138±0.045</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>717</td>
<td>Lipid</td>
<td>1.09±0.20</td>
<td>0.92±0.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>757</td>
<td>Protein</td>
<td>0.53±0.12</td>
<td>0.459±0.082</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>852</td>
<td>Glycogen and protein</td>
<td>1.27±0.61</td>
<td>1.05±0.30</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>937</td>
<td>Glycogen</td>
<td>1.67±0.94</td>
<td>1.27±0.46</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>1002</td>
<td>Protein</td>
<td>2.93±0.04</td>
<td>3.02±0.20</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>1031</td>
<td>Protein</td>
<td>0.646±0.093</td>
<td>0.629±0.080</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>

The mean spectra of MSCiPS1 cells contained similar features to those of EOS iPS cells, however showing a higher intensity at glycogen bands (852 cm\(^{-1}\) and 938 cm\(^{-1}\)). Interestingly, EOS iPSCs have a unique protein band at 983 cm\(^{-1}\) (arginine and tyrosine)\[^{107}\] that may be used as a Raman maker to distinguish between MSCiPS1 and EOS iPS cells. Two-sample t-test results (Table 3.1) revealed all other bands have statistically not different means except for 757 cm\(^{-1}\) and 852 cm\(^{-1}\). The pluripotency R4 marker of MSCiPS1 cells was 0.53 with a 95% confidence interval (CI) of 0.48-0.58 (n=26), which is higher than previously characterized pluripotent cells, but still in the pluripotent range.\[^{1}\] Likewise, Table 3.2 showed the two-sample t-test results comparing peaks from Raman spectra of MSCiPS1 and the CA1 line of hES cells. Except for a glycogen band at 937 cm\(^{-1}\), and protein bands at 1002 cm\(^{-1}\) and 1031 cm\(^{-1}\), other bands in the mean spectrum of MSCiPS1 were statistically different from those of hES cells. Glycogen concentration is known to vary in cell culture of different confluences, however the reason why two proteins bands differ is unclear. Larger standard deviation of MSCiPS1 cell spectra compared to that of EOS iPS cell spectra was
observed as shown in Figure 3.2, indicating more prominent metabolic variation in MSCiPS1 culture.

![Figure 3.2 Standard deviations of normalized EOS iPS and MSCiPS1 cell spectra](image)

Using PCA, the calculated scores computed from the spectra of EOS iPS cells, MSCiPS1 cells and hES cells (Figure 3.2), reveals a more obvious separation between hiPS and hES cells compared to Figure 2.9. It should be noticed that the two iPS cell lines were not clearly separated in Figure 3.1, but form an “hiPS cell cluster” distinct from the hES cluster. These two closely neighboring clusters differ in the PC2 dimension that represents only 3.8% of the total variance, but are not distinguishable on the PC1 dimension that contributes to 86.7% of the total variance. Therefore, one can conclude that these two iPS cell lines EOS and MSCiPS1 are highly similar in terms of their Raman spectra, both showing subtle differences compared to hES cells captured by PC2. Importantly, the MSCiPS1 line exhibits similar differences compared to hES cells as observed for EOS iPSC line. This difference has been shown to be more significant than that between the two cell lines. It was further confirmed that hiPS and hES cells were spectrally alike in most respects (PC1). The apparent discrimination between hiPS and hES cells using PC2 may reflect
compositional changes that are either caused by, or not completely reversed by the reprogramming process. This is, however, merely a supposition at this point and much more research would be required to draw any conclusion.

Figure 3.3 A PC1/PC2 score plot indicated a tendency for iPS cells (EOS line and MSCiPS1 line) and hES cells to cluster independently along the PC2 dimension, not distinguishable along the PC1 dimension.

EOS iPS and MSCiPS1 cells are not distinguishable in terms of Raman spectra.

3.3.2 Effect of EOS iPSC culture confluence on Raman spectra

Raman spectral variability was found to arise from cell cycle progression and changes in cell culture confluency of human tumor cells.\textsuperscript{[106]} It was hypothesized that various confluencies may reduce the previously observed similarity between hiPSCs and hESCs (Figure 2. 8). Therefore, an additional comparison of the hiPS cells, undifferentiated and differentiated hES cells was carried out by considering iPS cells of different confluencies. Spectra from EOS hiPSCs from different days (from Day 1 to Day 5, of which data from Day 3 was used for comparisons in Chapter 2.) after sub-culturing with a starting number of ca. 50,000 cells, were grouped along with those of hESCs (Figure 3.4 A). The clusters from hiPS cells on each day overlapped with the hES cluster, regardless of cell confluency. Likewise, all the spectra
from hiPS cells, undifferentiated hES cells and differentiated hES cells were grouped for PCA (Figure 3.4 B). It was evident that differentiated hESCs were distinguishable from the ‘pluripotent cluster’ consisting of hiPS and hES cells. Therefore, the results here were consistent with the previously observed similarity between hES and hiPS cells.
Figure 3.4 Comparison between hiPS EOS cells of various confluences, hES cells and differentiated hES cells. Spectra were taken from hiPS cells 1-5 days after sub-culturing (Day 1, the number of spectra taken n=14; Day 2, n=19; Day 3, n=17; Day 4, n=17; Day 5, n=18). Before principal component analysis these spectra were combine with spectra from hES cells (n=48) and differentiated hES cells (n=10), used in previous comparisons. A) PC2/PC1 score plot revealed the similarity of all the iPSCs and hES cells. B) PC2/PC1 score plot showed the tendency for pluripotent hiPS to cluster along with cells and differentiated hES cells to cluster independently along the PC1 dimension.
Figure 3.5 presents the mean spectra from hiPS cells on each day after sub-culturing, of which the day 3 data was used in the previous comparisons in Chapter 2. It was clear that all the mean spectra exhibited similar mean Raman intensities in the major peaks, except for a large variance in the two glycogen peaks 852 cm$^{-1}$ and 938 cm$^{-1}$. The mean spectrum from Day 4 has significantly higher peak intensities at these two glycogen bands which was attributed to a relatively high cell density in the colonies. Day 4 has lower peak intensities at glycogen bands than Day 5. The reason might be that overgrowing cells started to die.

![Figure 3.5 Mean spectra of hiPS EOS cells from different days after sub-culturing. All the spectra were normalized to the nucleic acid line at 783 cm$^{-1}$. All spectra had the R4 (757 cm$^{-1}$/783 cm$^{-1}$) differentiation state markers consistent with those of undifferentiated ES cells.](image)

The conventional R4 markers (757 cm$^{-1}$/783 cm$^{-1}$) for hiPSCs were 0.409±0.051, 0.40±0.12, 0.450±0.052, 0.537±0.047, 0.418 ±0.059 for days 1 to 5 respectively, and all of them were, close to the R4 of pluripotent hES cells (0.459±0.082). PCA was performed on the combined data of these hiPS cells. According to PC2/PC1 score plot in Figure 3.6, these hiPSCs were not distinguishable at this level of analysis.
Figure 3.6 PC2/PC1 scores plot of Raman spectra from hiPS EOS cells on the 1st-5th day after sub-culturing

3.4 Conclusion

The effects of using hiPS cells from a different cell line, and using hiPS cells of different culture confluences were examined. Raman spectra of EOS hiPS cells and MSCiPS1 cells were separated in the PC score plot, which supported the ability of Raman microscopy to discriminate between two cell lines of the same cell type (i.e., induced pluripotent cells). The ability to discriminate among other iPS cell lines was not tested due to unavailability of other iPS cell lines for this preliminary study. More data would clearly be desirable, however it is significant that at least the two available cell lines were spectrally consistent. Although Raman spectroscopy was able to distinguish between EOS iPS and MSCiPS1 cells, when grouping the data from these two hiPS cell lines together with the previously used undifferentiated and differentiated hES cells, it was evident in the score plot that, data points of hiPS and hES cells were clustered together and distinct from the those of differentiated cells. Mean spectra of MSCiPS1 cells presented similar features to those of the pluripotent cells (EOS iPS and hES cells), especially with the R4 marker around 0.5. Taken together, the difference between these two hiPS cell lines were relatively less prominent than that between
them and hES cells (both undifferentiated and differentiated). Therefore, the segregation among hiPS, undifferentiated and differentiated hES cells was not merely attributable to cell line differences, but most likely to the intrinsic differences.

Raman spectra were also collected from hiPS cell cultures of various confluencies. PCA results for these iPS cells indicated that they were not distinguishable from each other. There existed a distribution of different cell densities across the colony, thus, it was possible that each Raman spectrum included some the cell density bias within a certain spot. Overall, however there was no clear trend over the number of days after sub-culturing. PCA were performed on the data matrix combining all the spectra from hiPS cells of various confluencies, together with undifferentiated and differentiated hES cells, in which the pluripotent cells including all iPS and hES cells, and clearly separated from the differentiated cells.

The counterpart experiments on hESCs of various confluences were not presented here because the main idea was to find out that hiPS and hES cells were still similar in terms of Raman features when their cell density were close. The hiPS cell experiments covered the range of confluencies over which there would be one with a close cell density compared to the hES culture. One can conclude that the various confluences was not the main reason that cause the separation of hiPS, undifferentiated and differentiated hES cell in the Raman spectra.
4 Monitoring of the differentiation of induced pluripotent stem (iPS) cells with Raman microspectroscopy

4.1 Introduction

Under specific conditions, and using an appropriate differentiation stimulus, embryonic stem (ES) cells will differentiate and generate progeny consisting of derivatives of the three germ layers (mesoderm, endoderm, and ectoderm). They offer the potential for providing a source of cell types for cell replacement therapy. However, it is important to characterize the differentiation status of ES-derived cells carefully before clinical use, because if some cells remain in the undifferentiated state, a cancer-like teratoma could emerge. Currently the means to efficiently and non-invasively control and monitor stem cell differentiation remains one of the greatest challenges in stem cell technology.

Without the problems of ethical concerns and immune rejection, iPS cells hold great promise as a patient-specific cell source for regenerative medicine and therapy. Through careful control of differentiation conditions, functional cardiovascular cell lineages have been derived from iPS cells with a similar efficiency as from ES cells. Moreover, the mouse iPS cell derived cardiac myocytes were found to be able to repair and regenerate infarcted myocardia with improved cardiac function in vivo. It has also been reported that epidermal melanocytes generated from iPSCs can be used for the study of melanocyte developmental biology and diseases. In addition, neuronally differentiated iPS cells have been used for the discovery of patient-specific therapeutics and for identifying pathogenetically relevant targets in patients.
Many important problems associated with iPS cells have been suggested. For example, iPS cells derived by transcription-factor-based reprogramming were found to maintain an epigenetic memory of the origin cells. These cells were observed to favour differentiation along lineages related to the donor cell and hence served to restrict alternative cell fates, thus, undermining efforts at directed differentiation for applications in disease modeling and treatment. Just as for ES cells, rigorous characterization of differentiation is necessary for safe use of cells derived from iPS cells.

Immunocytochemical staining is a common method used to study stem cell differentiation. It detects proteins in cells using specific binding of antibodies. Nevertheless, this technique has a number of drawbacks, such as the requirement of fluorescent biomarkers (to visualize antigen-antibody interaction) that may perturb the cell and change the cell chemistry. Also, fluorescent marker photo-bleaching affects the ability to perform long-term experiments. Mass spectrometry and reverse transcription polymerase chain reaction (rtPCR) are two other widely used cell characterization techniques, but they are also destructive. There is a need for a robust, non-invasive, and label-free technique which can be applied to characterizing the behaviours of stem cells, especially their differentiation status. Several vibrational spectroscopy methods are potential candidates to distinguish undifferentiated stem cells from their derivatives. Fourier transform infrared (FTIR) spectroscopy has been demonstrated to distinguish stem cells from mature cells and has been applied to monitor the osteogenic differentiation of human mesenchymal stem cells. A classification model based on linear discriminant analysis (LDA) was developed to separate spectra of undifferentiated cells with glycogen accumulation from those of differentiated cells with high expression of calcium phosphate salts. A higher DNA absorbance caused by the unique open
chromatin structure was found to be associated with hematopoietic stem cells (HSCs) compared with bone marrow (BM) cells. However, it should be noted that FTIR application is limited to fixed or dried cells because of the strong interference from water.

As has been amply demonstrated in the previous chapters, Raman spectroscopy is a suitable tool to investigate living stem cells and differentiated cells. For example it has been shown that differentiated murine and human embryonic stem cells exhibited an intensity decrease in DNA/RNA Raman peaks, which agreed with the well-known reduction in the nucleus to cytoplasm ratio during differentiation. Significant variations have been observed in the peak intensities of protein bands between undifferentiated and differentiated cells. More importantly, the intensity ratio of 757 cm$^{-1}$ (tryptophan) / 784 cm$^{-1}$ (DNA/RNA composition) has been demonstrated to be a reliable maker for discriminating undifferentiated and differentiated ES cells. The major advantage of Raman over FTIR is the capability to investigate live cells; the major advantage of FTIR over Raman is its rapid detection.

In this chapter, Raman spectra were acquired from hiPS cells after they were induced to differentiate using 10% FBS at 2, 4, 8, 15 and 20 days. The resulting spectra were compared with the Raman spectra from undifferentiated hiPSCs (3 days after sub-culturing, same data as used in Chapter 2), undifferentiated ESCs (same data as used in Chapter 2), and differentiated hiPSCs (20 days of differentiation, as in Chapter 2).

4.2 Material and methods
The EOS hiPS cells were passaged and maintained in the same way, as described in Chapter 2. Prior to the Raman measurements, the undifferentiated hiPSCs were plated on mirrors with
a starting cell count of about 50,000 and cultured for 1-2 days, in order to allow the cells adapt to the new environment. Subsequently, differentiation mTeSR1 medium that contains 10% FBS was used to stimulate non-specific differentiation. For differentiated cells, due to a decrease in metabolism, a medium change was performed every other day. Raman spectra were taken from 2, 4, 8, 15, and 20 days differentiated hiPSCs, and analysed as described in Chapter 2.

### 4.3 Results and discussion

![Figure 4.1 Mean spectra of hiPSCs after 0, 2, 4, 8, 15, and 20 days of differentiation. Spectra were normalized to the 783 cm\(^{-1}\).](image)

Figure 4.1 is a plot of the normalized mean spectra (normalized to the 783 cm\(^{-1}\) nucleic acid band) of hiPS cells after 2, 4, 8, 15 and 20 days of differentiation as well as the mean spectrum of undifferentiated hiPS cells (Day 0) as a reference. The Raman spectra consisted of peaks corresponding to molecular vibrations of all major cellular components, nucleic acid
(783 cm\(^{-1}\), 811-813 cm\(^{-1}\) and 827 cm\(^{-1}\)), proteins (757 cm\(^{-1}\), 811-813 cm\(^{-1}\), 827 cm\(^{-1}\), 876 cm\(^{-1}\), 1002 cm\(^{-1}\), 1031 cm\(^{-1}\), 1061 cm\(^{-1}\)), lipid (700 cm\(^{-1}\) and 717 cm\(^{-1}\)), and carbohydrates (852 cm\(^{-1}\) and 938 cm\(^{-1}\)). Comparing the spectra of differentiated cells and undifferentiated cells, some changes can be observed. Several protein band intensities 757 cm\(^{-1}\), 827 cm\(^{-1}\), 876 cm\(^{-1}\), 1002 cm\(^{-1}\), 1031 cm\(^{-1}\) and 1061 cm\(^{-1}\) in the differentiated cells, as well as the pluripotency marker 757 cm\(^{-1}\)/783 cm\(^{-1}\) increased by about 50% after 20 days of differentiation and reached ~0.67. However, this value was smaller than the reported ratio (>1) for terminally differentiated hES cells. It was also observed that in the colonies of these differentiated iPS cells there were still GFP-active spots indicating undifferentiated iPS cells (Figure not shown). This observation may suggest an incomplete differentiation of hiPSCs with mTeSR and 10% FBS. The hypothesis of incomplete differentiation is in agreement with a reported result that some intracellular factors of iPS cells can inhibit terminal differentiation, with a 10 times lower efficiency compared to ES cells.[15] The inhibition factor(s) may be related to the four transgenes (Oct3/4, Klf4, Sox2, and c-Myc) that were initially used to derive iPS cells. In the establishment of iPS cells, these genes were silenced, so the artificial promoters may prevent these genes from resuming natural cellular control of expression during differentiation.[15] An intensity increase of phospholipid bands (717 cm\(^{-1}\) and 876 cm\(^{-1}\)) was also observed which was in contrast of the decreasing trend observed during ES cell differentiation. [119] Subtle changes of the band associated with the O-P-O stretching mode of RNA at 811-813 cm\(^{-1}\) were observed in the spectra, in terms of the peak position and intensity. Because the spectra are initially normalized to the DNA and RNA (but mainly DNA) band at 783 cm\(^{-1}\), the changes in the region 811 cm\(^{-1}\) to 813 cm\(^{-1}\) (RNA and protein assignment) probably indicates a change in the ratio of RNA to DNA, which can be
rationalized by different levels of mRNA translation during the differentiation.\cite{9} Similarly, a hybrid DNA/RNA band around 900 cm$^{-1}$ became prominent from day 4, perhaps indicating an increased level of mRNA translation from day 4. As discussed before, the intensities of glycogen bands 852 cm$^{-1}$ and 938 cm$^{-1}$ are indicative of the proliferation rate and cell density\cite{3} and it is known that glycogen accumulation occurs in a crowded slowly-growing colony. For each mirror, the initial seeding cell number was about 50,000 per mirror. When the cells differentiated, their proliferation rate decreased while, at the same time, the total cell number and cell density still increased but at a lower rate. That may explain why high levels of glycogen concentration were observed in the differentiated cell spectra.

Figure 4.2 Temporal profiles of relative intensities of major Raman peaks in the mean spectra. (A) Relative intensities of 852 cm$^{-1}$, 900 cm$^{-1}$, 938 cm$^{-1}$, and 1031 cm$^{-1}$. (B) Relative intensities of 757 cm$^{-1}$, 876 cm$^{-1}$, and 1002 cm$^{-1}$. Other peaks show fluctuation over the 20 days period (not shown).

Figure 4.2 is a plot of the relative intensities of major peaks in Figure 4.1 at different time points after stimulation of differentiation. The intensities of the peaks at 852 cm$^{-1}$, 900 cm$^{-1}$, 938 cm$^{-1}$, and 1031 cm$^{-1}$ relative to 783 cm$^{-1}$ follow similar trends (Figure 4.2A), suggesting that the same sources or correlated sources were contributing to the peak
intensities in these peaks at all time points. Correlation coefficients among these four peak intensities are in the range of 0.94-0.99 (Table 4.1, highlighting >0.85 coefficients in red), which further supports the observation of correlations between glycogen (852 cm\(^{-1}\) and 938 cm\(^{-1}\)) and protein (900 cm\(^{-1}\), 1031 cm\(^{-1}\)) bands. Protein levels are known to increase in differentiated cells (Figure 4.2 B), however the peak intensity at 1002 cm\(^{-1}\) can be the better candidate to provide sensitive detection compared to 757 cm\(^{-1}\) and 876 cm\(^{-1}\) to reveal this significant change before and after differentiation. Positive correlations exist between 757 cm\(^{-1}\) to other protein bands 852 cm\(^{-1}\), 900 cm\(^{-1}\), 1002 cm\(^{-1}\), 1031 cm\(^{-1}\) and 1061 cm\(^{-1}\). A weak correlation between 757 cm\(^{-1}\) and 1061 cm\(^{-1}\) was noticed, which may be due to a combination of nucleic acid and protein contributing to 1061 cm\(^{-1}\). The temporal profiles of relative intensities at 700 cm\(^{-1}\) and 717 cm\(^{-1}\) (lipid) do not show significant correlation with any of the protein or nucleic acid bands, therefore, cannot be used to distinguish differentiation state.

<table>
<thead>
<tr>
<th>Raman shift /cm(^{-1})</th>
<th>700</th>
<th>717</th>
<th>757</th>
<th>811</th>
<th>827</th>
<th>852</th>
<th>876</th>
<th>900</th>
<th>938</th>
<th>1002</th>
<th>1031</th>
<th>1061</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>717</td>
<td>0.20</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>757</td>
<td>-0.03</td>
<td>-0.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>811</td>
<td>0.29</td>
<td>0.38</td>
<td>-0.60</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>827</td>
<td>0.15</td>
<td>-0.05</td>
<td>0.41</td>
<td>0.21</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>852</td>
<td>-0.43</td>
<td>-0.07</td>
<td>0.88</td>
<td>-0.53</td>
<td>0.22</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>876</td>
<td>-0.37</td>
<td>-0.58</td>
<td>0.74</td>
<td>-0.58</td>
<td>0.55</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900</td>
<td>-0.41</td>
<td>0.05</td>
<td>0.87</td>
<td>-0.46</td>
<td>0.25</td>
<td>0.99</td>
<td>0.65</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>938</td>
<td>-0.54</td>
<td>0.00</td>
<td>0.77</td>
<td>-0.45</td>
<td>0.10</td>
<td>0.98</td>
<td>0.60</td>
<td>0.98</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1002</td>
<td>-0.26</td>
<td>0.28</td>
<td>0.89</td>
<td>-0.45</td>
<td>0.39</td>
<td>0.88</td>
<td>0.59</td>
<td>0.92</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1031</td>
<td>-0.43</td>
<td>0.22</td>
<td>0.83</td>
<td>-0.49</td>
<td>0.17</td>
<td>0.95</td>
<td>0.55</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1061</td>
<td>-0.48</td>
<td>-0.32</td>
<td>0.56</td>
<td>-0.75</td>
<td>-0.41</td>
<td>0.75</td>
<td>0.46</td>
<td>0.69</td>
<td>0.77</td>
<td>0.47</td>
<td>0.66</td>
<td>1.00</td>
</tr>
</tbody>
</table>
As Raman peaks originating from several cellular components overlap, it was difficult to observe small spectral differences. PCA was first carried out on the Raman spectra of differentiated hiPS cells. Inspection of the covariance matrix eigenvalues showed that the first four PCs represented 91.8% of the total variance (PC1 67.0%; PC2 15.7%; PC3: 5.5; PC4:3.6%). The greatest discrimination between the differentiated iPS cells at different days was observed in the PC3/PC4/PC1 scores plot (Figure 4.3); however the clusters still had overlapping areas suggesting that hiPS cells on each day stayed within a continuous range of differentiation levels. This might be due to the non-specific differentiation stimulus and niche formation. Day 2 data points appeared to be more clustered than other groups, because at an early stage of differentiation, the iPS cells may have biochemically closer phenotypes compared to the later stages.

![Figure 4.3 PC3/PC4/PC1 score plot of differentiated iPS cells.](image)

The PC loadings plot identified many other wavenumbers not observed in the mean spectra (Figure 4.4). Major protein and glycogen bands (757 cm\(^{-1}\), 853 cm\(^{-1}\), 1002 cm\(^{-1}\), 1031
cm\(^{-1}\) and 1061 cm\(^{-1}\)) dominated the PC1 loading plot, indicating protein and glycogen compositions as the origin of the primary variation in the spectra. Therefore, the largest contribution of the variances was attributed to two factors: the diversity of differentiation states that caused different levels of protein production, and various cell densities leading to changing glycogen concentration across the colony. Because there was no other nucleic acid band in PC1 except 899 cm\(^{-1}\) (nucleic acid and protein), this combination band should mainly be dominated by the protein contribution. The contribution from nucleic acid (827 cm\(^{-1}\), 864 cm\(^{-1}\) and 942 cm\(^{-1}\)) was mainly captured by PC2, along with lipid contribution (700 cm\(^{-1}\) and 715 cm\(^{-1}\)) and some protein features (1031 cm\(^{-1}\) and 1002 cm\(^{-1}\)). However, PC2 was most associated with intra-cluster variation, thus in the PC2 dimension, the inter-cluster segregation was not clear (Figure not shown).
Inter-cluster variations were also represented by PC3 and PC4. Because of a relatively smaller variance in nucleic acid composition, most of the nucleic acid bands (735 cm\(^{-1}\), 713 cm\(^{-1}\), 811 cm\(^{-1}\), 821 cm\(^{-1}\), 830 cm\(^{-1}\), 901 cm\(^{-1}\) and 1070 cm\(^{-1}\)) were represented in PC3 and PC4. Also one should note the negative correlation with the lipid bands (699 cm\(^{-1}\), 715 cm\(^{-1}\), and 1031 cm\(^{-1}\)) and most of the protein and glycogen bands (758 cm\(^{-1}\), 844 cm\(^{-1}\), 857 cm\(^{-1}\), 917 cm\(^{-1}\), 921 cm\(^{-1}\) and 1003 cm\(^{-1}\)). In contrast, in PC3, one of the distinct RNA band 867 cm\(^{-1}\) (ribose) correlated positively with the protein bands and lipid bands, but negatively to other nucleic acid bands and the other distinct RNA band at 811 cm\(^{-1}\) (O-P-O stretching).
The negative correlation between protein and nucleic acid is in agreement with the increasing protein concentration and reduced nucleic acid concentration during differentiation; the exception at 867 cm\(^{-1}\) is unclear but may be related to conformational changes\(^{[121]}\) that enhanced ribose vibration\(^{[21]}\).
Figure 4.5 (A) PC1/PC3/PC2 scores plot of differentiated iPSCs (2, 4, 8, 15 and 20 days of differentiation) and undifferentiated iPSCs (1, 2, 3, 4 and 5 days after sub-culturing), CA1 cells (5 days after sub-culturing) and differentiated CA1 cells (20 days of differentiation). (B) A simplified color scheme of Figure 4.4 A, with all differentiated iPSCs represented by blue crosses, all undifferentiated iPSCs represented by cyan circles. Inserted red-edged figure is a zoomed-in view of cyan dashed rectangular. There are 42.7% (91 data points out of 217 in total) differentiated iPSCs falling into the orange oval “pluripotent area”.

Raman spectra from undifferentiated and differentiated iPS cells, undifferentiated and differentiated ES cells were included in the PCA analysis in Figure 4.5 in order to
discriminate pluripotent cells and definitively differentiated cells. The undifferentiated iPS cells of different confluences were not markedly segregated from undifferentiated ES cells. Therefore, there exists a similarity between their biochemical components, which was supported by the results presented in Chapter 3. To facilitate making a general conclusion, data points of differentiated iPS cells were formatted uniformly (blue crosses) in Figure 4.5 B. These data points formed a dispersive shape in all the three dimensions, but distinct from differentiated CA1 cells on PC2 dimension. The reader should notice that there were 91 data points (42.7 % of 217 points) from differentiated iPS cells falling into the pluripotent zone (indicated by the orange oval in Figure 4.5 B). This provides further evidence of the incomplete state of differentiation and the existence of undifferentiated niches in these iPSC colonies. Similar to hES cells, the spatial complexity in hiPS cell cultures may induce heterogeneous microenvironments (niches) that influence iPS cell fates. To obtain a homogeneous cell culture and better discrimination, it is therefore necessary to have effective control of colony size and local environment or use mild and directional differentiation stimulus. Further studies of the gene expression profiles, especially the expression the four key transcription factors will be useful to confirm the observations.

Compared to ES cells, differentiated cells normally spend more time in the G1 phase compared with the S, G2, or M phases of the cell cycle, which causes a 50% reduction in the nucleic acid Raman peak at 783 cm$^{-1}$. Yet previous comparisons masked changes in the 783 cm$^{-1}$ peak because it was used to normalize the spectral intensities. In order to observe the information from this peak, vector normalization was applied as an alternative to process the mean spectra of undifferentiated and differentiated iPS cells. Because Raman spectra of iPS cells were not affected markedly by cell density (Chapter 3), only spectra from
hiPS cells after 3 days (the time point in the middle of the series) of sub-culturing are presented here. Each baseline-flattened, smoothed spectrum was normalized to the total area under the spectrum, such that the peak intensities indicate their relative percentage contribution to the total signal (Figure 4.6). The peak intensity at 783 cm\(^{-1}\) followed a general reduction trend as the cells differentiated with the exception of the point at 15 days when the laser might be mainly scanning over an undifferentiated growth area in the differentiated colony.

![Figure 4.6 Mean spectra of differentiated iPSCs (2, 4, 8, 15 and 20 days after differentiation) and undifferentiated hiPSCs (3 days after sub-culturing designated Day 0 in the legend). Inserted figure is an intensity plot of 783 cm\(^{-1}\) peak.](image)

Figure 4.6 shows the mean spectra of undifferentiated iPSCs (black curve) and 20 days differentiated iPS cells (red curve). Of interest, there is a noticeable decrease in nucleic acid bands at 783 cm\(^{-1}\) (35% reduction) and 811 cm\(^{-1}\) (53% reduction). In fact, with the exception
Figure 4.7 Mean spectra of differentiated (20 days) and undifferentiated iPS cells after vector normalization

Table 4.2 Two sample t-test results generally show the intensities of major peaks in vector normalized Raman spectra of undifferentiated and differentiated hiPSCs to be statistically different. Because multiple t-tests were used, the p-values were based on the Bonferroni inequality.\textsuperscript{[78]}

<table>
<thead>
<tr>
<th>Raman peak (cm$^{-1}$)</th>
<th>Assignment</th>
<th>Mean ± Standard Deviation /×10$^{-3}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hiPSCs(N=17)</td>
<td>Differentiated iPSCs (20 days)(N=38)</td>
<td></td>
</tr>
<tr>
<td>700 cm$^{-1}$</td>
<td>Lipid</td>
<td>2.1225 ± 0.0004</td>
<td>0.44867 ± 0.0003</td>
</tr>
<tr>
<td>717 cm$^{-1}$</td>
<td>Lipid</td>
<td>7.9430 ± 0.0006</td>
<td>5.0166 ± 0.0002</td>
</tr>
<tr>
<td>757 cm$^{-1}$</td>
<td>Protein</td>
<td>3.8401 ± 0.0000</td>
<td>3.3333 ± 0.0002</td>
</tr>
<tr>
<td>783 cm$^{-1}$</td>
<td>Nucleic acid</td>
<td>8.2278 ± 0.0007</td>
<td>5.3000 ± 0.0065</td>
</tr>
<tr>
<td>811 cm$^{-1}$</td>
<td>RNA</td>
<td>2.8369 ± 0.0011</td>
<td>1.3321 ± 0.0008</td>
</tr>
<tr>
<td>827 cm$^{-1}$</td>
<td>Protein and nucleic acid</td>
<td>2.1184 ± 0.0002</td>
<td>1.7603 ± 0.0010</td>
</tr>
<tr>
<td>852 cm$^{-1}$</td>
<td>Glycogen and protein</td>
<td>6.0691 ± 0.0002</td>
<td>8.0345 ± 0.0023</td>
</tr>
<tr>
<td>876 cm$^{-1}$</td>
<td>Protein</td>
<td>1.2572 ± 0.0001</td>
<td>2.3784 ± 0.0005</td>
</tr>
<tr>
<td>900 cm$^{-1}$</td>
<td>Nucleic acid and protein</td>
<td>1.0647 ± 0.0001</td>
<td>1.7702 ± 0.0003</td>
</tr>
<tr>
<td>938 cm$^{-1}$</td>
<td>Glycogen</td>
<td>7.7860 ± 0.0007</td>
<td>9.8168 ± 0.0053</td>
</tr>
<tr>
<td>973 cm$^{-1}$</td>
<td>Protein and nucleic acid</td>
<td>0.2116 ± 0.0000</td>
<td>0.12857 ± 0.0025</td>
</tr>
<tr>
<td>984 cm$^{-1}$</td>
<td>Protein</td>
<td>1.9693 ± 0.0002</td>
<td>0.2330 ± 0.0002</td>
</tr>
<tr>
<td>1002 cm$^{-1}$</td>
<td>Protein</td>
<td>2.1786 ± 0.0055</td>
<td>1.8938 ± 0.0436</td>
</tr>
<tr>
<td>1031 cm$^{-1}$</td>
<td>Protein</td>
<td>4.3442 ± 0.0001</td>
<td>4.6814 ± 0.0008</td>
</tr>
<tr>
<td>1061 cm$^{-1}$</td>
<td>Protein</td>
<td>2.0500 ± 0.0002</td>
<td>2.5030 ± 0.0007</td>
</tr>
</tbody>
</table>
of the 827 cm\(^{-1}\) (protein and nucleic acid) and 1031 cm\(^{-1}\) (protein) peaks, the mean intensities of major Raman peaks from undifferentiated and differentiated hiPS cells are statistically different (Table 4.2). Vector normalization produced similar results to those obtained using 783 cm\(^{-1}\) normalization in the comparisons of the mean spectra of undifferentiated and differentiated hiPSCs, that mean spectra of undifferentiated hiPSCs consist of stronger bands at 700 cm\(^{-1}\), 717 cm\(^{-1}\), 811 cm\(^{-1}\) and weaker peaks at 852 cm\(^{-1}\) and 938 cm\(^{-1}\). Figure 4.7 suggests that in differentiated iPSCs the higher relative intensities of 757 cm\(^{-1}\) and 1002 cm\(^{-1}\) (protein) seen in Figure 4.1 were possibly due to their lower DNA contribution (783 cm\(^{-1}\)) and is a normalization artifact. Conversely, under vector normalization undifferentiated hiPSCs spectra exhibited a slightly higher protein composition (15% higher at both 757 cm\(^{-1}\) and 1002 cm\(^{-1}\)). Vector normalization is problematic in the sense that it considers every spectrum to have an equal area under the curve. In the differentiated iPS cell culture, two glycogen bands at 852 cm\(^{-1}\) and 938 cm\(^{-1}\) contributed larger areas, but this is not related to pluripotency. Therefore, a smaller area under the spectrum could also cause the appearance of the “higher protein composition” in undifferentiated iPS cells.

![Figure 4.8 PC1/PC2/PC3 scores plot of undifferentiated and differentiated iPS cells (20 days differentiation)](image)
PCA analysis was carried out on the vector normalized Raman spectra from undifferentiated and differentiated hiPS cells in order to examine whether the retrieved information at 783 cm\(^{-1}\) can improve the discrimination. The first attempt was on the Raman spectra from undifferentiated and 20 days differentiated hiPS cells (Figure 4.8 and Figure 4.9).

Figure 4.9 (A) PC1/PC2 loading plot (B) PC3/PC4 loading plot. Contributions of each PC are PC1, 72.5%; PC2 8.8%; PC3 4.5%; PC4 2.5%. The first four PCs captured 88.3% of the total variance.

Two clusters were spatially well separated in PC1/PC2/PC3 plots (Figure 4.8) with only one outlier from the differentiated group (N=38) in the undifferentiated zone. PC1 captured
72.5% of the total variance, and represented a mixture of lipid (718 cm\(^{-1}\)), nucleic acid (782 cm\(^{-1}\) and 812 cm\(^{-1}\)), proteins and glycogen (broad band at 847 cm\(^{-1}\)-867 cm\(^{-1}\), 942 cm\(^{-1}\) and 1003 cm\(^{-1}\)). The protein band 757 cm\(^{-1}\) is not prominent in PC1 or PC2, suggesting the change of this peak is not significant in terms of the fraction in the total signal.

![Figure 4.10 PC1/PC2/PC4 scores plot of the vector normalized Raman spectra from undifferentiated iPSCs and differentiated iPSCs (2, 4, 8, 15, and 20 days of differentiation). The dashed line separates undifferentiated zone (below) and differentiated zone. The number of differentiated cells below the line: 14 after 2 days differentiation, 12 after 4 days of differentiation, 5 after 8 days of differentiation, 0 after 15 days of differentiation, 1 after 20 days of differentiation.](image)

Not surprisingly, PCA applied to the vector-normalized spectra leads to a better segregation between undifferentiated and 2-20 days differentiated iPS cells in Figure 4.10. There were only 33 data points (14% of 231 data points in total) from differentiated groups falling into the differentiated zone below the boundary line, of which were 1 from 20 days differentiated cells, 0 from 15 days differentiated cells, 5 from 8 days differentiated cells, 12 from 4 days differentiated cells, and 2 from 2 day differentiated cells. This semi-quantitative measurement reveals a temporal trend that less data points fall into the “undifferentiated zone” for longer differentiated times. However, the improved segregation might be
highlighting the difference in the nucleic acid contribution to the total signal. As mentioned before, the difference might not be related to changes in absolute nucleic acid concentration, but, alternatively, to changes in total signal (e.g. the influence of glycogen should be considered). Therefore, it remains unclear whether the segregation evident in Figure 4.10 actually reveals the intrinsic differences in differentiation status, or just differences in glycogen levels.

Figure 4.11 PC loadings of Raman spectra from undifferentiated and differentiated iPS cells (0 day, 2 day, 4 days, 8 days, 15 days, and 20 days of differentiation) (A) PC1/PC2 loadings plot (B) PC3/PC4 loadings plot. Contributions of each PC are PC1, 49.8%; PC2 11.8%; PC3 7.1%; PC4 5.1%. The first four PCs captured 88.3% of the total variance.
PC loading plots (Figure 4.11) looked similar to those for undifferentiated and differentiated (20 days) iPS cells (Figure 4.9). As PCA is an unsupervised method, the factors (loadings, scores and eigenvalues) are computed to maximize the variance without considering features of the data matrix. Supervised linear discriminant analysis (LDA) computes a linear combination of variables that maximize the variance between groups and minimize the variance within groups. However, in LDA, the number of variables (wavenumber position in Raman spectra) has to be smaller than the number of samples. Therefore, for future analyses, one could consider using the PC scores input to a LDA model.

[49]

4.4 Conclusion

This is the first reported Raman spectroscopic study of differentiated human induced pluripotent cells. The differentiated iPS cells exhibited significantly different Raman spectral profiles, which can be exploited for discrimination. With all the spectra normalized to a nucleic acid peak 783 cm$^{-1}$, specific biomolecular differences that can be inferred by the Raman spectra, including an increase in the relative protein and glycogen concentration relative to nucleic acids, and a decrease in lipid concentration in differentiated iPS cells. These findings were attributed to higher levels of mRNA translation, larger cell percentages in G1 phase and increased cell density in differentiated cell culture. The cause of lipid band variation is still unknown but could be attributed to higher cell density (thus higher membrane content). The results from hiPS cells subjected to a non-specific differentiation stimulus showed that, in comparison to hES cells, only a very modest degree of differentiation was induced. Thus, based on the intensity of the R4 differentiation state
indicator, the results showed that progressively more, but incomplete, differentiation of hiPS cells resulted as a function of the duration of the differentiation stimulus.

Vector normalization was attempted to examine the contributions of major peaks to the total peak area of each spectrum in order to better reveal information from the band at 783 cm\(^{-1}\) that was obscured when it was used to normalized spectral intensities. It was observed that DNA contribution to the total signal decreased during differentiation. The reason could be either that the absolute DNA concentration decreased, or the total signal increased during differentiation (larger under curve areas at glycogen bands in differentiated cell culture). PCA on vector-normalized data was able to discriminate between undifferentiated and differentiated iPS cells, revealing a temporal trend of differentiation.

It should be noted that, the improved discrimination using vector-normalized spectra could also arise from the equalization of total under-the-curve areas (the normalization technique). Error could emerge from this normalization technique, so the segregation did not definitively indicate the differentiation states. It is therefore necessary to conduct chemical marker measurements and to correlate them with spectral observations to fairly evaluate the PCA and the relative merits of each normalization approach. Regardless, these preliminary results show great promise for the development of Raman microspectroscopy coupled with multivariate analysis methods as a rapid, label-free tool to identify and monitor the differentiation of iPS cells. The results are encouraging enough that future work to establish the reasons underlining these spectral differences is justified in order to enable application to more complex in vivo and in vitro systems.
5 Conclusion and future work

The feasibility of Raman microspectroscopy for characterization of iPS cells and determination of their differentiation status was explored in this work. In chapter 2, Raman spectra of hiPS was reported for the first time, and then compared with those of hES cells and differentiated hES cells. The results demonstrated spectroscopic similarity between hiPS and hES cells, and also dissimilarity between hiPS and differentiated hES cells. In particular, the content presented in Chapter 2 leads to the inference that reprogramming leads to a reduction in the protein to nucleic acid ratio of reprogrammed somatic cells, resulting in a ratio that is similar to that measured for embryonic pluripotent cells. PCA results indicated that there are subtle differences between the iPS cells and hES cells.

To examine whether the differences are the result of a cell line artifact or spectral variance from culture confluence, an additional iPS cell line was considered in Chapter 3. Raman spectra were acquired from the iPS cell line MSCiPS1, and then compared with the EOS iPSC line for both undifferentiated and differentiated hES cells. It was found that these two iPS cell lines can be discriminated from each other, however this cell-line related difference was not as pronounced as that between hiPS and hES (difference between two cell types). Additionally, the influence of cell culture confluency was examined. It was shown that the various cell culture confluencies did not affect the segregation of hiPS with differentiated hES cells, while the similarity between hiPS and hES cells was maintained regardless of confluency.

In Chapter 4, the identification and discrimination of differentiated iPS cell states was attempted by using Raman microspectroscopy. The differentiated iPS cells exhibited significantly different Raman spectral profiles from those of undifferentiated pluripotent
cells, specifically in the reduction of nucleic acid/protein ratio. The decreased nucleic acid/protein ratio was attributed to a higher level of mRNA translation and a larger cell percentage in G1 phase. However, incomplete differentiation and evidence of the existence of undifferentiated cells in the differentiated iPS cell culture were also found. An alternative approach to analyzing the contributions of each major peak to the spectra using vector normalization was explored. PCA results on vector-normalized data indicated an improved segregation between undifferentiated and differentiated iPS cells, but it was cautioned that this could arise from the normalization technique itself due to distortions caused by the different contributions from (for example) glycogen peak areas to each spectrum. The segregation and temporal trend did not unambiguously suggest a change in differentiation state, but possibly increased cell densities.

Nevertheless, the results presented here clearly demonstrate the potential utility of Raman microspectroscopy in the identification, characterization of iPS cells, as well as discrimination between iPS and differentiated iPS cells. To extend this work, it would be interesting to further investigate how the Raman signatures of hiPS cells vary under different maintenance conditions; this would be especially useful for evaluating how the Raman signatures correlate with iPS cell differentiation. It would also be useful to stimulate specific differentiation pathways in iPS cells and to determine their temporal Raman features. Also, obtaining more information using immunochemistry or staining methods in parallel with the Raman measurements would be necessary to help interpret the spectroscopic results.
Bibliography


<table>
<thead>
<tr>
<th>Reference</th>
<th>Authors</th>
<th>Journal</th>
<th>Year</th>
<th>Volume</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>[66]</td>
<td>V. Ng, A.B.H. Choo</td>
<td>Open Stem Cell J</td>
<td>2010</td>
<td>8-17</td>
<td></td>
</tr>
</tbody>
</table>


# Appendix

## Table 0.1 Raman peaks of interest in the Raman spectra[^46]

<table>
<thead>
<tr>
<th>Raman band (cm⁻¹)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>699-702</td>
<td>Cholesterol or cholesterol esters (lipid assignment)</td>
</tr>
<tr>
<td>713</td>
<td>Cytosine (DNA/RNA assignment)[^21]</td>
</tr>
<tr>
<td>715-719</td>
<td>C–N symmetric stretching in phospholipids</td>
</tr>
<tr>
<td>735</td>
<td>Adenine ring breathing mode (DNA/RNA assignment)[^124]</td>
</tr>
<tr>
<td>755-759</td>
<td>Symmetric ring breathing in tryptophan (Protein assignment)</td>
</tr>
<tr>
<td>765-766</td>
<td>Tryptophan[^107] (Protein assignment)</td>
</tr>
<tr>
<td>783</td>
<td>782 cm⁻¹: DNA/RNA (pyrimidine → ring breathing); 788 cm⁻¹: DNA (backbone → O-P-O stretching)</td>
</tr>
<tr>
<td>811</td>
<td>RNA O-P-O stretching (RNA assignment)</td>
</tr>
<tr>
<td>821-822</td>
<td>Phosphodiester (nucleic acid assignment)</td>
</tr>
<tr>
<td>827-828</td>
<td>Proline, hydroxyproline, tyrosine (protein assignment) ν₂ PO₃⁻ stretch of nucleic acids</td>
</tr>
<tr>
<td>830-831</td>
<td>Tyrosine and O-P-O stretch of nucleic acids (protein/DNA/RNA) assignment</td>
</tr>
<tr>
<td>842-847</td>
<td>glucose, (C-O-C) skeletal mode</td>
</tr>
<tr>
<td>852-857</td>
<td>Ring breathing mode of tyrosine, protein; C-C stretch of proline ring; Glycogen; (protein and glycogen assignment)</td>
</tr>
<tr>
<td>864</td>
<td>Nucleic acid assignment (weak)[^121]</td>
</tr>
<tr>
<td>867</td>
<td>Ribose vibration, one of the distinct RNA bands</td>
</tr>
<tr>
<td>876</td>
<td>Hydroxyproline (collagen assignment), tryptophan</td>
</tr>
<tr>
<td>899-901</td>
<td>DNA/RNA hybrid band[^125], L-Proline (protein assignment)[^107]</td>
</tr>
<tr>
<td>915</td>
<td>Ribose vibration, one of the distinct bands of RNA</td>
</tr>
<tr>
<td>917-918</td>
<td>Proline, hydroproline, glycogen and lactic acid</td>
</tr>
<tr>
<td>921</td>
<td>C-C stretch of proline ring/glucose/lactic acids (protein and glycogen assignment)</td>
</tr>
<tr>
<td>932</td>
<td>Proline, hydroproline, ν(C-C) skeletal of collagen backbone</td>
</tr>
<tr>
<td>937-943</td>
<td>Proteins (collagen type I → C-C stretching, α-helix → C-C stretching); carbohydrates (glycogen) Adenine (nucleic acid assignment)</td>
</tr>
<tr>
<td>950</td>
<td>Most probably amino acids proline and valine, and polysaccharides</td>
</tr>
</tbody>
</table>

[^46]: Various assignments and references provided for each peak, indicating their specific role in the Raman spectra.
<table>
<thead>
<tr>
<th>Raman band(cm⁻¹)</th>
<th>Assignment</th>
</tr>
</thead>
</table>
| 973-974         | C-C backbone (collagen assignment)  
|                 | Ribose vibration, one of the distinct RNA bands |
| 986-987         | $\delta$(C-N-C) and $\delta$(C-N-H) of thymine$^{[126]}$ |
| 1001-1003       | Phenylalanine, protein assignment |
| 1030-1034       | C-H in-plane bending mode of phenylalanine  
|                 | phospholipids |
| 1059-1061       | C-C in-plane bending (one of C-C ring vibration to be expected in aromatic structure of xylene) |
| 1070-1090       | Symmetric PO$_2^-$ stretching of DNA |