# Fourier transform infrared microspectroscopy for the characterisation of human embryonic and human induced pluripotent stem cells

## Julie Oanh Cao

### BSc. (Hons)

Department of

Anatomy and Developmental Biology

Faculty of Medicine,

Nursing and Health

Sciences

Monash University, Australia

Submitted in total fulfilment

of the requirements of the degree of

Doctor of Philosophy

2013

# ADDENDUM

P4, par 1, after 'but further studies are required to assess their true efficacy.' insert the following:

'A more recently reported phenomenon that is claimed to confound the interpretation of spectral information in transflectance measurements is the electric field standing wave (EFSW) artefact (Bassan et al., 2013). However, these effects have been shown to be significantly reduced after applying appropriate spectral pre-processing methodologies such as the computation of derivative spectra, and vector normalisation. (Miljković et al., 2013). The EFSW is discussed more extensively in Chapter 5 of this dissertation.'

P163, ref 2, change 'Journal of pharmaceutical and biomedical analysis' to 'Journal of Pharmaceutical and Biomedical Analysis'

P163, ref 4, change 'Stem cells and development' to 'Stem Cells and Development'

P164, ref 4, change 'Bone Marrow Mesenchymal Stem Cells in Patients with Beta Thalassemia Major: Molecular Analysis with Attenuated Total Reflection-Fourier Transform Infrared Spectroscopy Study as a Novel Method' to 'Bone marrow mesenchymal stem cells in patients with beta thalassemia major: molecular analysis with Attenuated Total Reflection-Fourier transform infrared spectroscopy study as a novel method'

P164, ref 7, change 'Human Feeder Layers for Human Embryonic Stem Cells' to 'Human feeder layers for human embryonic stem cells'

P164, ref 12, change 'Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines' to 'Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines'

P165, ref 13, change 'Recombinant Vitronectin Is a Functionally Defined Substrate That Supports Human Embryonic Stem Cell Self-Renewal via αVβ5 Integrin' to 'Recombinant vitronectin is a functionally defined substrate that supports human embryonic stem cell self-renewal via αVβ5 integrin'

P165, ref 17, change 'Journal of biophotonics' to 'Journal of Biophotonics'

P165, ref 18, change 'Characterization and Differentiation of Human Embryonic Stem Cells' to 'Characterization and differentiation of human embryonic stem cells'

P166, ref 20, change 'Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures' to 'Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures'

P166, ref 25, change 'Optical Spectroscopy for Noninvasive Monitoring of Stem Cell Differentiation' to 'Optical spectroscopy for non-invasive monitoring of stem cell differentiation'

P167, ref 31, change 'Analytica chimica acta' to 'Analytica Chimica Acta'

P167, ref 33, change 'Persistent Donor Cell Gene Expression among Human Induced Pluripotent Stem Cells Contributes to Differences with Human Embryonic Stem Cells' to 'Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells'

P167, ref 32, change 'Investigative ophthalmology & visual science' to 'Investigative Ophthalmology and Visual Science'

P168, ref 35, change 'Tissue Engineering--Current Challenges and Expanding Opportunities' to 'Tissue engineering-current challenges and expanding opportunities'

P168, ref 36, change 'Chromatin Structure and Gene Expression Programs of Human Embryonic and Induced Pluripotent Stem Cells' to 'Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells'

P169, ref 44, change 'The Use and Misuse of FTIR Spectroscopy in the Determination of Protein Structure' to 'The use and misuse of FTIR spectroscopy in the determination of protein structure'

P169, ref 46, change 'Sequential Injection/Mid-Infrared Spectroscopic Analysis of an Acetone-Butanol-Ethanol Fermentation: Analyte Cross-Correlation Effects.' to 'Sequential injection/mid-infrared spectroscopic analysis of an acetone-butanol-ethanol fermentation: analyte cross-correlation effects.'

P169, ref 49, change 'Stem cells and development' to 'Stem Cells and Development'

P170, ref 50, change 'Applied spectroscopy' to 'Applied Spectroscopy'

P170, ref 53, change 'Tissue engineering' to 'Tissue Engineering'

P171, ref 59, change 'Transcriptional Signature and Memory Retention of Human-Induced Pluripotent Stem Cells' to 'Transcriptional signature and memory retention of human-induced pluripotent stem cells'

P171, ref 61, change 'Journal of pharmaceutical and biomedical analysis' to 'Journal of Pharmaceutical and Biomedical Analysis'

P171, ref 63, change 'Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures' to 'Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures.'

P171, ref 64, change 'Biomedical Applications of Infrared Microspectroscopy Using Synchrotron Radiation' to 'Biomedical applications of infrared microspectroscopy using synchrotron radiation'

P172, ref 66, change 'Mie-Type Scattering and Non-Beer-Lambert Absorption Behavior of Human Cells in Infrared Microspectroscopy' to 'Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy'

P172, replace ref 67 with the following:
'Nakamura, T., Kelly, J.G., Trevisan, J., Cooper, L.J., Bentley, A.J., Carmichael, P.L., Scott, A.D., Cotte, M., Susini, J., Martin-Hirsch, P.L., Kinoshita, S, Fullwood, N.J and Martin, F.L (2010). Microspectroscopy of spectral biomarkers associated with human corneal stem cells. Molecular Vision *16*, 359.'

P172, ref 68, change 'Comparison of Human Induced Pluripotent and Embryonic Stem Cells: Fraternal or Identical Twins?' to 'Comparison of human induced pluripotent and embryonic stem cells: fraternal or identical twins?'

P172, ref 69, change 'Demountable Liquid/Flow Cell for in Vivo Infrared Microspectroscopy of Biological Specimens' to 'Demountable liquid/flow cell for in vivo infrared microspectroscopy of biological specimens', and 'Applied spectroscopy' to 'Applied Spectroscopy'

P172, ref 71, change 'Nature protocols' to 'Nature Protocols'

P173, ref 73, change 'Directed Differentiation of Human Embryonic Stem Cells as Spin Embryoid Bodies and a Description of the Hematopoietic Blast Colony Forming Assay' to 'Directed differentiation of human embryonic stem cells as spin embryoid bodies and a description of the hematopoietic blast colony forming assay'

P174, ref 79, change 'Applied spectroscopy' to 'Applied Spectroscopy'

P174, ref 81, change 'Nature biotechnology' to 'Nature Biotechnology'

P174, ref 83, change 'Identification of Spectral Modifications Occurring during Reprogramming of Somatic Cells' to 'Identification of spectral modifications occurring during reprogramming of somatic cells', and 'PloS one' to PloS ONE.

P174, ref 85, change 'Analytical chemistry' to 'Analytical Chemistry'

P175, ref 89, change 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors' to 'Induction of pluripotent stem cells from adult human fibroblasts by defined factors'

P175, ref 92, change 'Medical science monitor' to 'Medical Science Monitor'

P176, ref 94, change 'Embryonic Stem Cell Lines Derived from Human Blastocysts' to 'Embryonic stem cell lines derived from human blastocysts'

P176, ref 99, change 'Molar Absorptivity in the 1000–4000 cm$^{-1}$ Range and Quantitative Infrared Spectroscopy of Aqueous Solutions' to 'Molar absorptivity in the 1000–4000 cm$^{-1}$ range and quantitative infrared spectroscopy of aqueous solutions'

P177, ref 104, change 'Nucleic acids research' to 'Nucleic Acids Research'

P177, ref 101, change 'Fourier Transform Infrared Microspectroscopy Identifies Symmetric PO$_2$- Modifications as a Marker of the Putative Stem Cell Region of Human Intestinal Crypts' to 'Fourier transform infrared microspectroscopy identifies symmetric PO$_2$- modifications as a marker of the putative stem cell region of human intestinal crypts'

P178, ref 109, change 'Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences' to 'Human induced pluripotent stem cells free of vector and transgene sequences'

P178 ref 110, change 'Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells' to 'Induced pluripotent stem cell lines derived from human somatic cells'

P105, par 3 replace:

'The mean thickness of the day 4 mesendodermal differentiated MIXL1 [GFP/w] cells originally cultured on MEFs in KOSR-based medium were 0.93 μm ±0.05, and 0.91 μm ±0.07 for the, cells originally cultured on Matrigel in mTESR based medium.' with:

'The day 4 mesendodermal differentiated MIXL1 [GFP/w] cells were very thin, with thicknesses of 0.93 μm ± 0.05, and 0.91 μm ± 0.07 for the cells cultured on MEFs in KOSR-based medium, and on Matrigel in mTESR based medium respectively. Due to their thinness, the position of the nuclei in these cells would have been approximately in the same position for each experiment, and therefore any EFSW that was probing the nucleus would have been of the same intensity. This was supported by the AFM imaging of our cytospun cell populations where populations of cells from both feeder grown and Matrigel grown cells had flat cylindrical morphologies of uniform thickness, without evidence of protruding nuclei (Supplementary Figure 3.1).'

P105, par 3, add the following to the end of paragraph 3, after '(Bassan et al., 2013).':
'Further, any remaining contributions of the EFSW would be largely nullified by our earlier spectral pre-processing. For instance, the computation of spectral derivatives and vector normalization were shown to significantly decrease the effects of the EFSW in a study by Miljković and colleagues. (Miljković et al., 2013). Certainly the prominent spectral differences observed in second derivative spectra between the experimental conditions reported here appear unlikely to have been due to EFSW effects, a view supported by the replication of these observations across different experiments, stem cell strains and the range of measurement modalities (FPA, SR-FTIR, transflection and transmission, dried and living cells). A good example is the high lipid absorbance (C-H stretching bands 3050-2800 $cm^{-1}$) observed consistently in undifferentiated human embryonic and reprogrammed stems cells in both dried transflection measurements and SR-FTIR transmission measurements on living cells, compared to differentiated progeny cell spectra (Figure 2.4 & Figure 4.2).'

P171 Insert the following after ref 64:
'65. Miljković, M. ; Bird, B.; Lenau, K.; Mazur, A.I.; Diem, M. (2013). Spectral cytopathology: new aspects of data collection, manipulation and confounding effects. Analyst, *138*, 3975-3982.'

P171-178, references 65 (Mitalipova et al.)-112 (Zelig et al.) should be renumbered to 66-113 i.e. reference ref 65 is now ref 66, ref 66 is now 67 and so forth.

On the following pages, replace the following instances of the word 'artefact' with 'signatures.': P23 par 3, P154 par 2 & 3, P155 par 1 & 2, P156 par 1 & 2, P157 par 4, and P161 par 1.
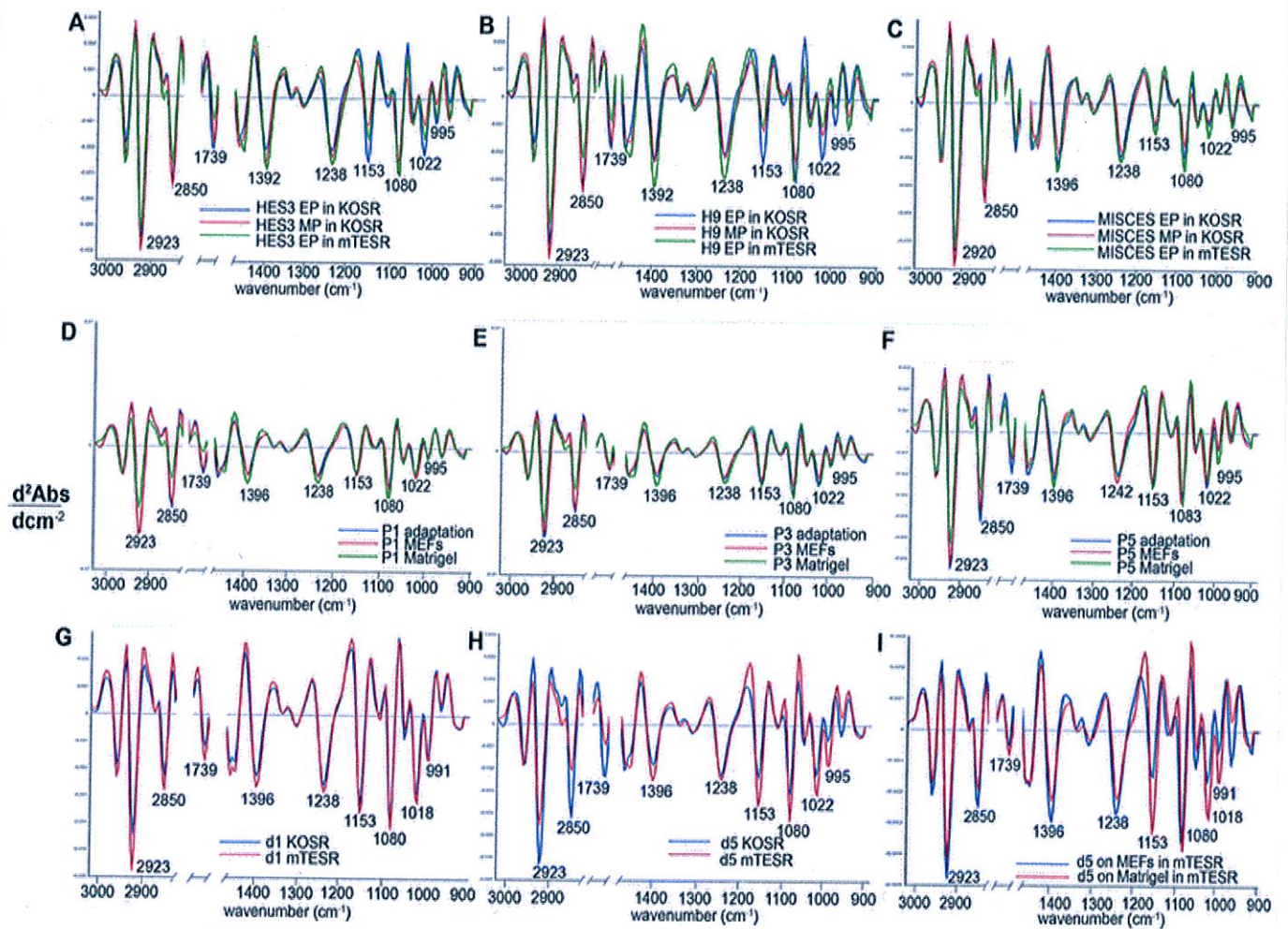
Change Supplementary Figure 3.1 to Supplementary Figure 3.2. Insert the following image as Supplementary Figure 3.1. AFM image of a monolayer of human embryonic stem cells cytospun onto a MirrIR slide.
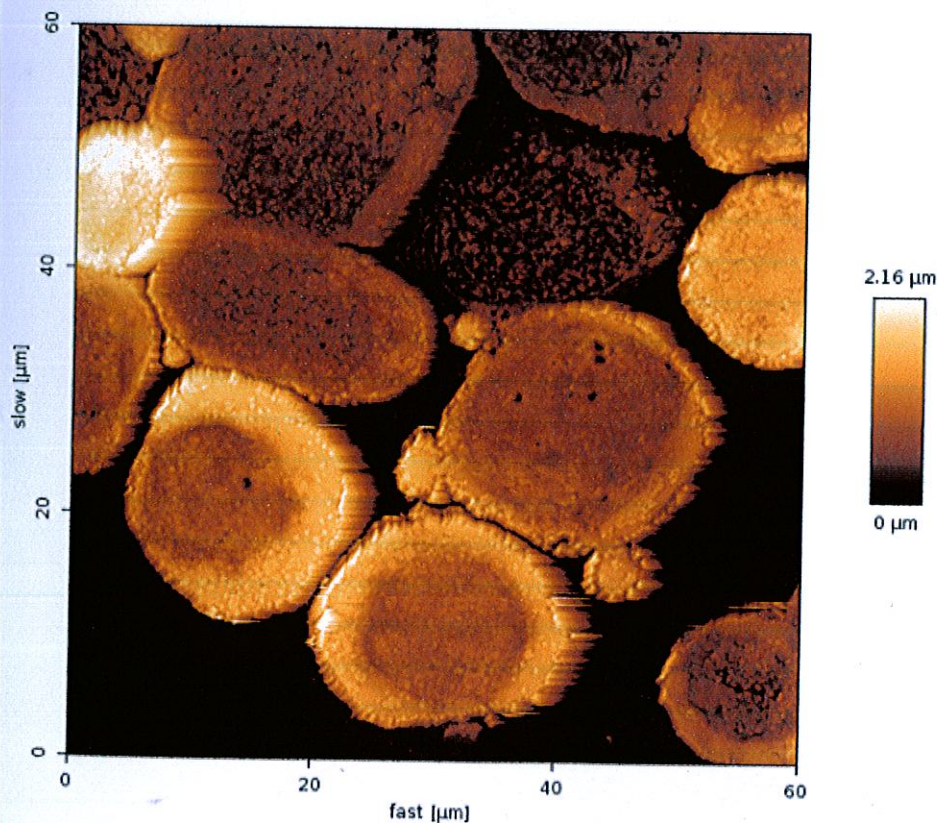
Replace Figure 1.3 with the following image:
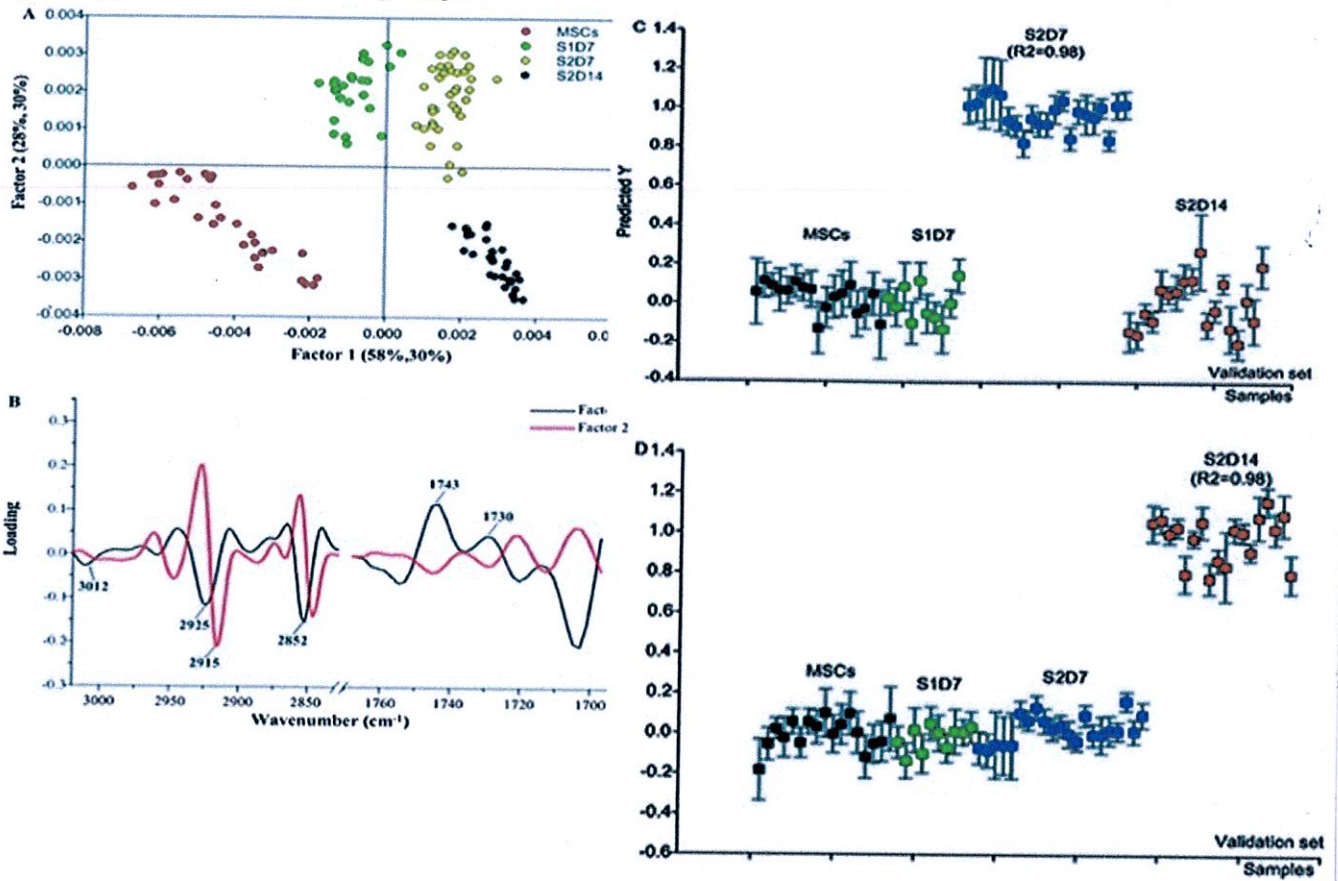


Replace Figure 3.3 with the following:

*'Believe you can and you're halfway there'*

Theodore Roosevelt

# Table of Contents

# Abstract

Fourier transform infrared (FTIR) microspectroscopy shows potential as a benign, objective and rapid tool to screen clinically destined pluripotent and multipotent stem cells. In this work, we explored the utility of this technique to distinguish between several human embryonic and human induced pluripotent stem cell lines and their differentiated progeny, based on their macromolecular chemistry. We observed both intra-class and inter-class biochemical variation within the undifferentiated stem cell cohorts, in addition to spectroscopic differences between both classes during their differentiation towards mesendodermal and ectodermal lineages. Chemometric comparisons of the progeny cells alone, indicated that they were phenotypically distinct, suggesting that current reprogramming methods do not yield human induced pluripotent stem cells that are equivalent to human embryonic stem cells, in spite of immunohistochemistry results suggesting the antithesis. However, whether these phenotypical changes are indicative of the inadequacies of the reprogrammed cells to form therapeutically useful lineage committed progeny needs to be established.

To elucidate the basis of the phenotypic differences observed between these pluripotent stem cells, we decided to investigate the role of environmental factors influencing these spectral 'signatures', by varying their growth environments. Accordingly, we employed Fourier transform infrared (FTIR) microspectroscopy to acquire spectra from several hESC lines and their derived cell types maintained in both feeder dependent and feeder independent conditions, in KOSR based or mTESR based hESC medium, and either mechanically or enzymatically dissociated. It was found that hESC lines grown under different conditions possessed unique FTIR spectroscopic 'signatures' and that these spectroscopic differences persisted even upon differentiation towards mesendodermal lineages. The results from this study illustrates the power of this modality for defining macromolecular phenotypic differences between several lines of human embryonic stem cells and human induced pluripotent stem cells and their lineage committed progeny. The technique has been shown to be highly sensitive, as indicated by its ability to detect biochemical differences even between stem cell lines of the same class. Further, the results indicate that the spectroscopic phenotypes are sensitive to a combination of genetics and environmental factors.

# Publications and Submitted Manuscripts

Julie Cao, Elizabeth S. Ng, Don McNaughton, Edouard G. Stanley, Andrew G. Elefanty, Mark J. Tobin, Philip Heraud (2012). Fourier transform infrared microspectroscopy reveals unique phenotypes for human embryonic and induced pluripotent stem cell lines and their progeny (published in the Journal of Biophotonics)

Julie Cao, Elizabeth S. Ng, Don McNaughton, Edouard G. Stanley, Andrew G. Elefanty, Mark J. Tobin, Philip Heraud (2013). Fourier transform infrared microspectroscopy reveals that tissue culture conditions affect the macromolecular phenotype of human embryonic stem cells (published in the Analyst)

Julie Cao, Don McNaughton, Mark J. Tobin, Andrew G. Elefanty, Philip Heraud (2013). The characterisation of pluripotent and multipotent stem cells using Fourier transform infrared microspectroscopy (published in the International Journal of Molecular Sciences)

<div align="center">

**Monash**

**University**

# General Declaration

</div>

In accordance with Monash University Doctorate Regulation 17/ Doctor of Philosophy and Master of Philosophy (MPhil) regulations the following declarations are made: I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes three original manuscripts, two of which have been published in peer reviewed journals. The third manuscript has been submitted and is currently under review. The core theme of the thesis is the utilisation of laboratory and synchrotron based FTIR microspectroscopy to characterise human embryonic and human induced pluripotent stem cells and their various lineage committed progeny. The ideas, development, provision of data and writing up of all the papers in the thesis were partly the responsibility of myself, the candidate, working within the Monash Immunology and Stem Cell Laboratories (MISCL), Monash University under the supervision of Professor Andrew Elefanty, Professor Donald McNaughton, Dr Philip Heraud and Dr Mark Tobin.

The inclusion of co-authors reflects the fact that the work came from the active collaboration between researchers and acknowledges input into team-based research.

In the case of Chapter 2, contributions to the work involved the following: The $MIXL1^{GFP/w}$ targeted hESC line described both in Chapters 2 and 3 was generated by Dr. Richard Davis; the $NKX2.5^{GFP/w}$ targeted hESC line was generated by Dr. David Elliott; the ENVY line was generated by Dr. Magdaline Costa; the HES3 and H9 lines were obtained from WiCell; the MEL1, ES4CL1 and IMR90C2 lines were provided by the Australian Stem Cell Centre, and the AUS line was donated by Professor Paul Verma.

In the case of Chapter 3, the MISCES line was generated by Dr. Anna Michalska's research group.

The differentiation methods described in Chapter 2 and 3 were developed by Elizabeth Ng, who also provided assistance with the spin EB experiments.

Overall, my contribution to the work included in this thesis was 90%.

| Thesis chapter | Publication title | Publication status | Nature and extent of candidate's contribution |
|---|---|---|---|
| 1 | The characterisation of pluripotent and multipotent stem cells using Fourier transform infrared microspectroscopy | Submitted | Researching the literature and writing the manuscript |
| 2 | Fourier transform infrared microspectroscopy reveals unique phenotypes for human embryonic and induced pluripotent stem cell lines and their progeny | Published | Performing experiments, data analysis and writing the manuscript |
| 3 | Fourier transform infrared microspectroscopy reveals that tissue culture conditions affect the macromolecular phenotype of human embryonic stem cells | Published | Performing experiments, data analysis and writing the manuscript |

Julie Oanh Cao

June 2013

# Acknowledgements

First and foremost, I would like to gratefully acknowledge my supervisors Philip Heraud, Andrew Elefanty, Donald McNaughton and Mark Tobin, for taking me on as a PhD student, and providing me with the opportunity to carry out research in the nascent field of stem cell biospectroscopy. I am indebted to the guidance and wisdom which you imparted to me throughout my candidature. Thanks to everybody in the E/S lab, who have all provided me with advice and technical support at various times, with an extra special thank-you to the members of Stem Core, in particular Robyn Mayberry, Amanda Bruce, Sue Mei Lim and Sheetal Saini for teaching me about tissue culture, and providing me with stem cells for my spectroscopy experiments. I am grateful to everybody at the Centre of Biospectroscopy, for passing on their spectroscopy knowledge, with a special mention to Finlay Shanks for teaching me how to acquire spectra and for fixing the spectrometer when it wasn't working. Thank you to the Australian synchrotron for granting me beam time for my live cell experiments, and for funding my conference attendances. Finally, I would like to acknowledge my wonderful family and friends for their endless emotional and/or IT support over the past four years. I am so grateful to have you all in my life.

# List of Abbreviations

| | |
|---|---|
| AFCs | Amniotic fluid cells |
| APEL | Albumin polyvinylalcohol essential lipids |
| ANN | Artificial neural networks |
| BFGF | Basic fibroblast growth factor |
| ATR-FTIR | Attenuated total reflectance Fourier transform infrared microspectroscopy |
| BM | Bone marrow |
| BMP | Bone morphogenetic protein |
| CBC | Crypt base columnar |
| CDB | Cell dissociation buffer |
| $COO^-$ | Carboxylate |
| CXCR4 | C-X-C chemokine receptor type 4 |
| DNA | Deoxyribose nucleic acid |
| DSP | Distal side population |
| EBs | Embryoid bodies |
| E-CADHERIN | Epithelial cadherin |
| EFSW | Electric field standing wave |
| EMSC | Extended multiplicative scatter correction |
| ESNCs | Embryonic stem cell derived neural cells |
| FACS | Fluorescence Activated Cell Sorting |
| FGF | Fibroblast growth factor |
| FPA | Focal plane array |
| FPA-FTIR | Focal plane array Fourier transform infrared spectroscopy |
| FTIR | Fourier transform infrared spectroscopy |
| GFP | Green fluorescence protein |
| hESC | Human embryonic stem cells |
| hiPSC | Human induced pluripotent stem cells |
| hMSC | Human mesenchymal stem cells |
| HMSC | Haematopoietic stem cells |
| IR | Infrared |

| | |
|---|---|
| IRENI | Infrared environmental imaging |
| KOSR | Knock out serum replacer |
| LDA | Linear discriminant analysis |
| LRCs | Label retaining cells |
| MCT | Mercury cadmium telluride |
| MEFs | Murine embryonic fibroblast cells |
| MIXL1 | Mixl1 homeobox-like 1 |
| mRNA | Messenger ribonucleic acid |
| MSC | Multiplicative scatter correction |
| MSCs | Mesenchymal stem cells |
| MVA | Multivariate analysis |
| NPCs | Neural progenitor cells |
| OSLN | Oct4, Sox2, Lin28 and Nanog vectors |
| PBS | Phosphate buffered saline |
| PC | Principal component |
| PCA | Principal component analysis |
| PDGFRα | Platelet derived growth factor alpha |
| PLS-DA | Partial least squares discriminant analysis |
| PLSR | Partial least squares regression |
| PSP | Proximal side population |
| QCL | Quantum cascade laser |
| RNA | Ribonucleic acid |
| RT PCR | Real time polymerase chain reaction |
| SC | Stem cell |
| SIMCA | Soft independent modeling of class analogy |
| SNRS | Signal to noise ratios |
| SP | Side population |
| SRC | Synchrotron radiation centre |
| SR-FTIR | Synchrotron radiation Fourier transform infrared microspectroscopy |
| SSEA-4 | Stage specific embryonic antigen-4 |
| SVMA | Support vector machines |
| TA cells | Transiently amplified cells |

| | |
|---|---|
| TD cells | Terminally differentiated |
| TGF-β3 | Transforming growth factor beta 3 |
| TRA-1-60 | Trafalgar-1-60 |
| UHCA | Unsupervised hierarchical cluster analysis |
| VEGF | Vascular endothelial growth factor |

# CHAPTER 1

Literature review

## 1.1. Introduction

Owing to their unique potential to differentiate into the three embryonic germ layers and give rise to all the cell types of the embryo proper, pluripotent stem cells hold much promise in the field of regenerative medicine. They include embryonic stem cells (Thomson et al., 1998), so called because of their derivation from blastocysts stage embryos, and induced pluripotent stem cells (Takahashi et al., 2006), aptly named because pluripotency has been 'induced' via the reprogramming of their somatic cell progenitors. In contrast, adult stem cells are mostly 'multipotent', which means that they are able to give rise to a subset of cell lineages, and of these, the hematopoietic stem cells are the best characterised population (Wagers and Weissman, 2004).

Before their envisaged clinical applications, numerous challenges pertaining to the isolation, identification, enrichment and purification of differentiated stem cells must be overcome (Griffith and Naughton, 2002; Langer, 2007; Taupin, 2006). The development of more robust screening processes is critical because the inadvertent transplantation of populations of undifferentiated stem cells has been known to lead to the formation of non-malignant tumours called teratomas. At present, stem cell differentiation is monitored via the use of a number of molecular biological techniques which include in vitro and in vivo assays, flow cytometry, real time polymerase chain reaction (RT-PCR) and microarray technologies (Hoffman and Carpenter, 2005) (Figure 1.1). These techniques not only require time consuming sample preparation, but also involve the use of biomarkers or labels, which are absent on certain cell types such as cardiomyocytes, gastrointestinal stem cells (Walsh et al., 2009), and corneal stem cells (German et al., 2006). Furthermore, these label driven methods have been known to reduce sample integrity by causing cellular stress and damage, thereby affecting the cells' behaviour. Given the insufficiencies of these

**Figure 1.1**: Flow chart summarising conventional molecular biology techniques currently used to monitor stem cell differentiation, the parameters that they measure, and their disadvantages.

| **Conventional molecular biology techniques used to monitor stem cell differentiation** |
|---|

| In vitro and in vivo assays | Flow cytometry | Microarrays | RT-PCR |
|---|---|---|---|

| Parameters measured |
|---|

| Cell functionality | Intracellular and cell surface markers | Gene expression | Detection and quantification of DNA |
|---|---|---|---|

**Disadvantages:**

- Laborious sample preparation
- May require the use of biomarkers or labels, which are not present on all stem cell types
- Labeling methods can compromise cellular integrity

methods, there is a clear need amongst stem cell biologists, to implement an objective, label-free, non-destructive technique for the screening of stem cells and their derivatives.

The recent adoption of vibrational spectroscopic approaches to study stem cell differentiation has emerged as a feasible solution to this problem (Downes et al., 2010). One of these modalities, Fourier transform infrared (FTIR) microspectroscopy, has been the subject of preliminary studies by various groups to interrogate both pluripotent and multipotent cells. Whilst the study of biological samples using FTIR microspectroscopy has been successful for more than half a century (Blout and Mellors, 1949 and Woernley, 1952); laying the foundation for our current understanding of their IR band assignments, its application to stem cells has only taken place within the last few years. This work furthers the field by providing new knowledge, comparing human embryonic and human induced pluripotent stem cells, and exploring the effect of the growth environment on spectral phenotype.

## 1.2. FTIR microspectroscopy – a concise background

Mid infrared FTIR spectroscopy, based on radiation absorption between 2.5µm and 25µm wavelengths (4000-400 cm$^{-1}$) exploits the intrinsic property of molecular systems to vibrate in resonance with different frequencies of infrared light. In biological samples, the vibrational modes in macromolecular molecules, such as proteins, lipids, carbohydrates and nucleic acids, give rise to a series of clearly identifiable functional group bands in FTIR spectra, providing us with information about relative concentrations and specific chemical structures (Mantsch et al., 1986). Band assignments of mid-IR spectra common to biological samples are presented in Table 1.1 according to foundation publications in the literature.

The most prominent band in biological spectra is the amide I band which arises from the coupled C=O stretching and N-H bending of proteins (Byler and Susi, 1986; Jackson and Mantsch, 1995) and is sensitive to changes in protein secondary structure. However, the interpretation of protein information in this region of the spectrum needs to be done with caution, since contamination of the spectra from water vapour and light scattering can cause profound changes in this spectral region. Recent approaches that have been developed to mitigate these effects are claimed to be successful (Bassan et al., 2012; Vaccari et al., 2012), but further studies are required to assess their true efficacy.

A more recently reported phenomenon that is claimed to confound the interpretation of spectral information in transflectance measurements is the electric field standing wave (EFSW) artefact (Bassan et al., 2013; Filik et al., 2012). However, these effects have been shown to be significantly reduced after applying appropriate spectral pre-processing methodologies such as the computation of derivative spectra, and vector normalisation. (Miljković et al., 2013). The EFSW is discussed more extensively in Chapter 5 of this dissertation.

### .1.2.1. FTIR microspectroscopy instrumentation

The types of FTIR instruments that have been used to interrogate stem cells include benchtop FTIR microspectroscopic instruments which utilise either globar or synchrotron mid-infrared radiation sources. Recently, the coupling of focal plane array (FPA) detectors to FTIR microscopes has allowed for the acquisition of thousands of spectra in a single experiment (Lewis et al., 1995). Nevertheless, despite the speed advantages, such systems do so at the expense of decreased spatial resolutions and signal to noise ratios when compared to synchrotron based experiments which utilise single detectors. An exception to the latter this is the multi-beam IR synchrotron source at the IRENI (IR Environmental

Imaging) beamline located within the Synchrotron Radiation Centre (SRC) (Wisconsin, USA) (Nasse et al., 2011) which employs an FPA detector.

To achieve diffraction limited spatial resolution with single point detectors, the infrared beam size is confined to a smaller region of interest via decreasing the aperture in the IR microscope. However, doing this with a globar IR source results in degraded signal to noise ratios due to an insufficient amount of IR light reaching the detector. By using the highly brilliant, collimated beam of photons generated by synchrotron IR sources, spectra can be acquired using an aperture of 5μm×5μm because the SNR can be ~1000 times higher than can be achieved with globar IR sources (Miller, 2009). It should be noted however, that due to the effects of diffraction, the obtainable spatial resolution is a function of wavenumber, and is not necessarily defined by the microscopic aperture that is employed.

More recently, advances in IR instrumentation have seen the utilisation of quantum cascade laser sources of mid infrared light (Haibach et al., 2011), which are more brilliant than synchrotron sources and allow for FTIR measurements with high signal to noise ratios to be achieved in the laboratory, even with single living cells. However, a few major challenges need to be overcome before they can be widely adopted (Weida and Yee, 2011). The first one is concerned with how to remove the mid-IR laser speckle, which arises from the use of coherent illumination, and significantly reduces image resolution. The second major issue relates to determining the optimal way to couple the laser radiation into the microscope, so that the former issue, in addition to others, can be fully investigated. Thirdly, QCL have restricted tuning ranges and because they require scanning, data acquisition is considerably slower. As a consequence, the broadband, multiplexing nature of FTIR is lost.

## 1.3. Interpreting features of the FTIR spectrum

Since the identification of most spectral features is difficult in the raw spectrum due to bands from multiple vibrational modes being superimposed, it is a common practice to convert the spectra to their derivatives, often the second derivative (Savitzky and Golay, 1964), apply a de-noising filter such as smoothing, and correct for path length differences using normalisation techniques such as vector normalization, or more powerful methods such as multiplicative scatter correction (MSC) or extended multiplicative scatter correction (EMSC) (Martens and Stark, 1991). The resultant spectra possess a flatter baseline, and for second derivative spectra, negative peaks, which are directly aligned to the centre of the underivatised spectrum's absorbance peak, enabling previously overlapping bands to be resolved. Whilst comparing the average second derivative spectra can be useful in discerning differences in band intensities between different spectral datasets, it can be misleading especially when dealing with a small number of spectra, or if there are different levels of heterogeneity in the sample dataset. Accordingly, spectroscopists tend to employ more objective, multivariate approaches for the analysis of spectral trends. Typical methods used in these studies include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA), Unsupervised Hierarchical Cluster Analysis (UHCA) and Artificial Neural Networks (ANNs).

### 1.3.1. Principal Component Analysis

Owing to the intrinsic multi-dimensionality of IR spectra, the task of interpreting trends within the dataset can be difficult. PCA allows for the reduction in the dimensionality of spectral data sets, thereby facilitating the identification of any clustering patterns that may exist (Wold et al., 1987). In PCA, each spectrum is represented as a single point on a

scores plot with axes called principal components (PCs) These PCs comprise a new set of variables, maintaining as much of the original variation as possible, with the first PC representing the vector showing the direction of greatest variation in the spectral data set and subsequent PCs showing the direction of next greatest variation in the data set, with these sources of variation independent to previous PCs. This condensing of the larger number of original variables to a smaller number of variables allows for the most relevant analytical information to be presented. In PCA, a loadings plot or 'pseudo spectrum' is calculated for each PC, which displays the variance at each wavenumber in the spectral dataset (Figure 1.2).

### 1.3.2. Linear Discriminant analysis

LDA is a factor analysis method which involves the decomposition of a matrix of spectra into matrices which consist of loading spectra and scores (Geladi and Kowalski, 1986). The original spectra can be thought of as linear combinations of the loading spectra and the loadings' contributions are denoted by the scores. This technique ensures that inter-class separation is maximised whereas any intra-class separation is minimised. Often, a cross-validation step is implemented, where the model is validated by using a supervised training data set, followed by classification of an independent validation test set (Figure 1.3).

### 1.3.3. Partial Least Squares Discriminant Analysis

PLS analysis is another multivariate analysis technique that is used to decompose and identify structure in a large data sets (Geladi and Kowalski, 1986), and is often used in conjunction with LDA. PLS involves the identification of a set of components called latent vectors, which decomposes the X (predictor) and Y (dependent) matrices simultaneously and is followed by a regression step where the decomposition of X is used to predict Y. In PLS-DA the calibration data matrix consists of the spectral dataset (multivariate X) and a

**Figure 1.2.** PCA results of subsets of 'stem cell like' renal epithelium carcinoma cells. Included in the analysis were the sub-side populations, distal side population (DSP) and proximal side population (PSP) versus non-side population (Non-SP) cell spectra. (A) The scores plot of PC1, PC2 and PC3 and (B) corresponding loadings of PC1 and (C) PC2 are shown. Key biochemical differences are outlined in lipid, phosphodiester and carbohydrate absorption bands (Hughes et al. 2010).
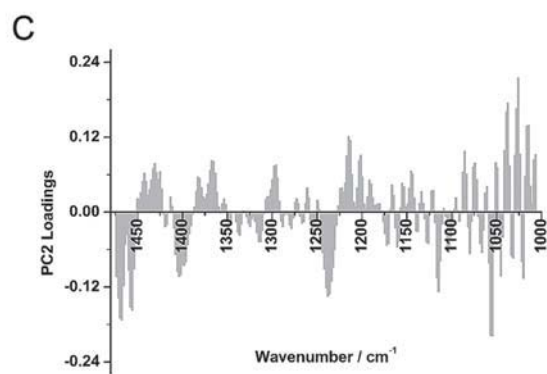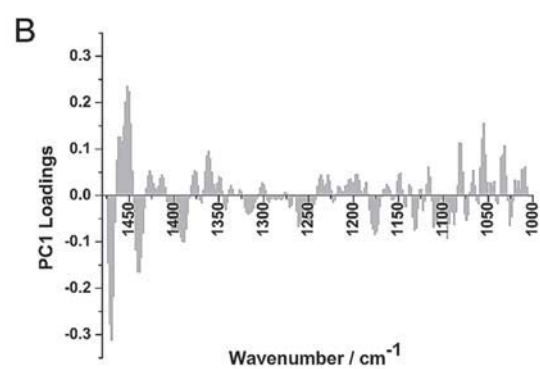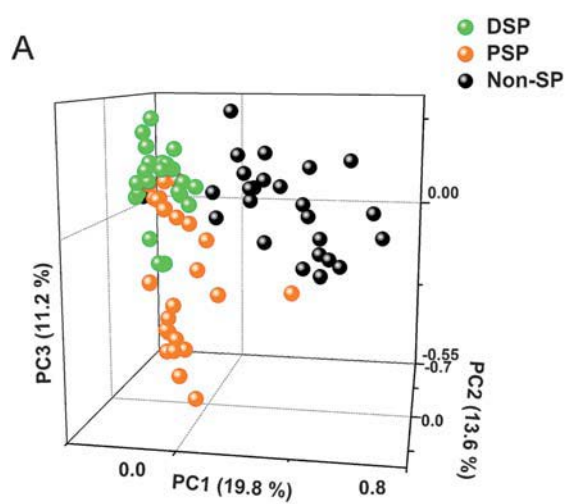
**Figure 1.3.** (A) Scores and loadings plots from the PLS-DA of FTIR spectral data acquired at different stages of hepatic differentiation (A) Scores plot showing factor 1 and 2, explaining 58% and 28% of the sample variance, respectively; (B) loading plot for factors 1 and 2 showing the most variable spectral regions explaining the PLS-DA. PLS discriminant analysis (PLS-DA) results of spectra drawn from the four investigated cell classes: undifferentiated rBM-MSCs, early stage cells (S1D7), mid-stage cells (S2D7) and late stage cells (S2D14) (C-D). The correlation coefficients ($R2 > 0.98$ for all calibration sets) indicated that all the datasets were well modelled. PLS-DA correctly classified and discriminated all of the validation spectra (Ye et al. 2012).
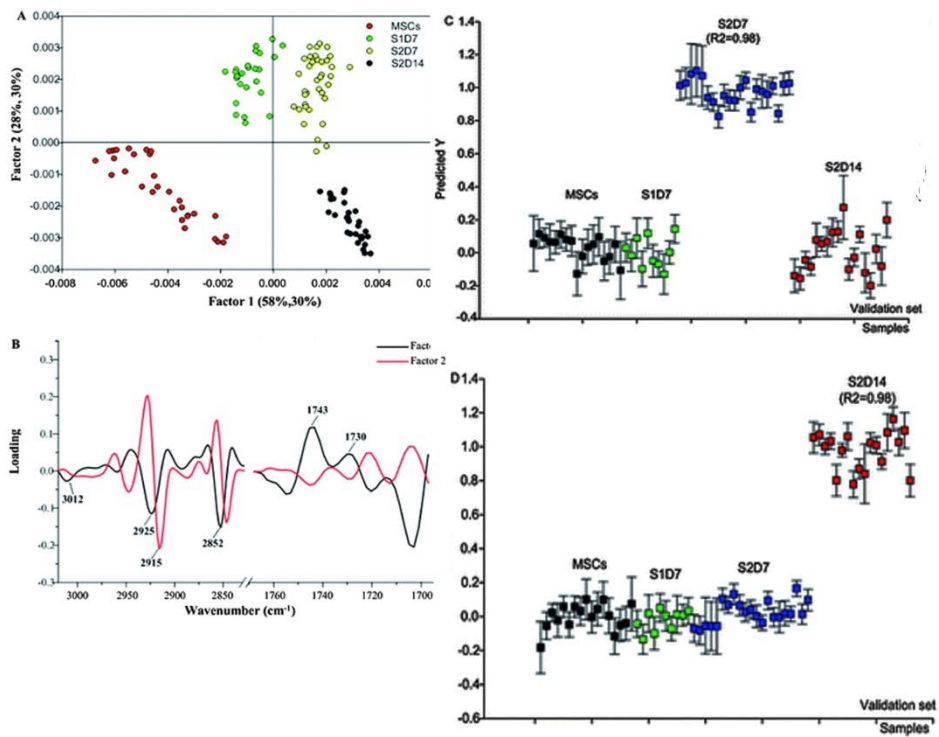
Y matrix containing variables with integer values of 0 or 1 coding for the each of the modelled spectral classes. Classification of the dataset is then carried out by predicting a Y value for each spectrum in an independent validation using PLS models that had been generated from the calibration set. Correct classification of each class are arbitrarily assigned to samples with predicted $Y > 0.5$ for respective spectra.

### 1.3.4. Unsupervised Hierarchical Cluster Analysis

In UHCA, spectral distances are measured using the pre-processed dataset to elucidate the degree of similarity between two spectra or clusters, with more similar spectra having smaller distances (Tanthanuch et al., 2010). The clustering process itself is performed by an algorithm, such as Ward's algorithm, which utilises a matrix defining inter-spectral distances to determine the two most similar IR spectra via combining the two clusters with the smallest degree of heterogeneity into a new cluster or hierarchical group (Figure 1.4).

### 1.3.5. Artificial neural networks

An ANN is an example of a *non-linear* pattern recognition approach that can be used in FTIR microspectroscopy of stem cells, and is analogous to biological neural systems (Agatonovic-Kustrin and Beresford, 2000). Their working environment consists of numerous interconnected processing elements (neurons) that work in cohesion to gain, predict, interpret and represent data in order to solve multifaceted user-specific problems. The architecture of an ANN consists of several 'neuronal' layers, adjoined by 'synapses'. Data is first sent to the input layer, which sends this via the synapses to the middle layers, and then to the output layer. For instance, in the case of FTIR data, the number of input neurons often equates to the number of individual wavenumber measurements in the spectrum, whereas the output neurons correspond to the different classification groupings. Parameters called "weights", which manipulate the data in the calculations are stored in the

**Figure 1.4:** UHCA results of mouse embryonic stem cells and progenitor cells at different stages of neural differentiation. Cluster analysis employed Ward's algorithm using second derivatives, vector normalised spectra, over the spectral ranges 3000-2800 cm$^{-1}$ to 1750-900 cm$^{-1}$ (Tanthanuch et al., 2010).

~ 16 ~
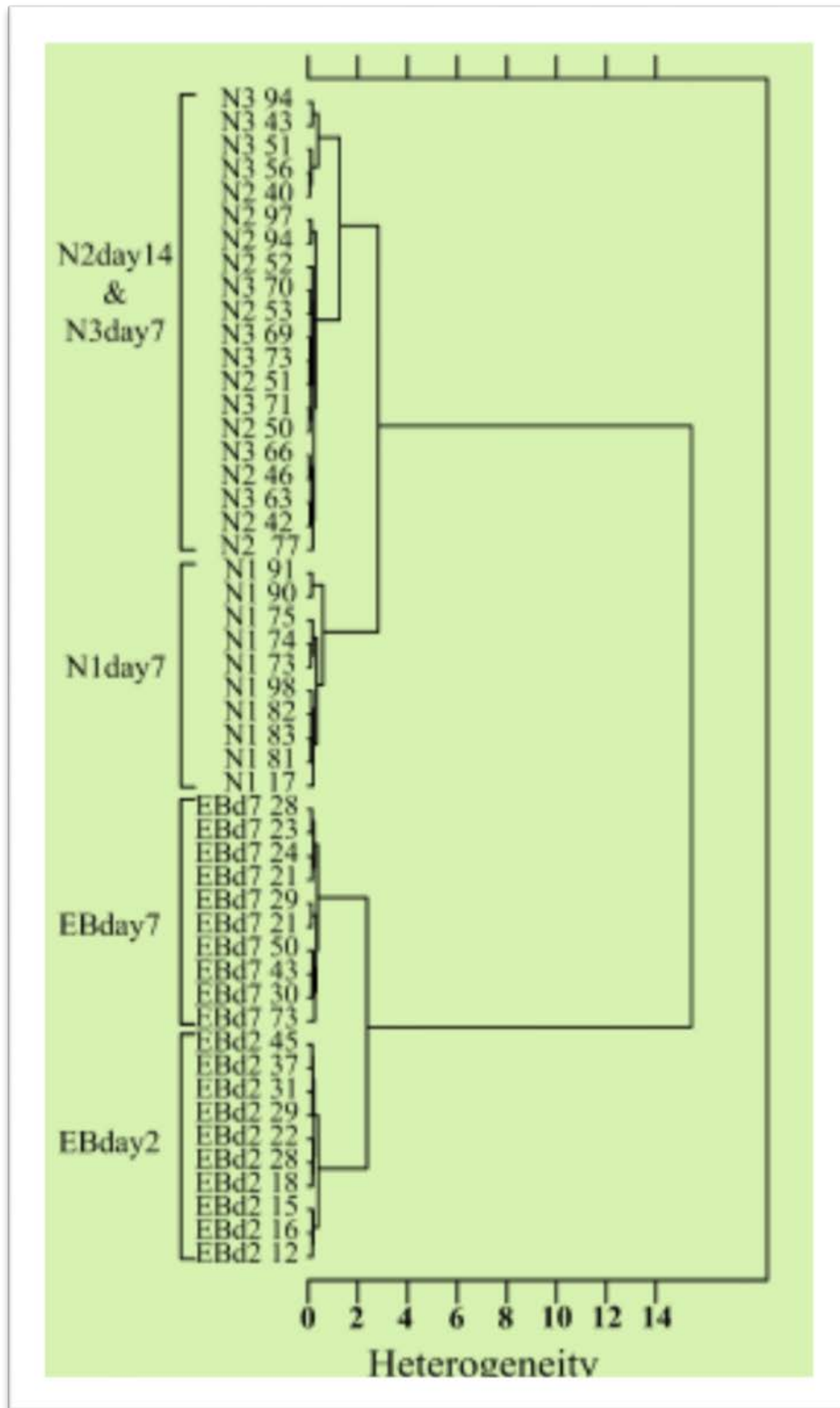
synapses, and can be optimised during the training phase of the ANN via back propagation, which adjusts these, in addition to other biases via calculating the gradient of the error.

All of these approaches to spectral classification have been widely used in biospectroscopy in general, including stem cell spectroscopy work (pertinent examples are :(Pijanka et al., 2010), PCA; (Walsh et al., 2009), LDA; (Tanthanuch et al., 2010), and UHCA (Heraud et al., 2010).

## 1.4. FTIR microspectroscopy can discriminate between stem cells of different potencies

The applicability of FTIR microspectroscopy for discriminating between stem cells with different intrinsic potencies based on their spectral 'signatures' was demonstrated by one laboratory which compared the spectral signatures of individual pluripotent stem cells (hESCs) from individual multipotent stem cells (hMSCs) (Pijanka et al., 2010). Examination of their IR spectra revealed differences in spectral regions associated with lipids. For instance, the hESC spectra displayed more intense peaks at 2920 $cm^{-1}$ and 2850 $cm^{-1}$ and at 1740 $cm^{-1}$, which were red shifted by approximately 4 $cm^{-1}$, relative to the single cell spectra of the hMSC. These observations were validated by lipid staining tests which revealed higher lipid concentrations in the cytoplasm of the hES cells compared with the hMS cells. This data suggested a possible link between lipid concentrations and the potency of a stem cell, which might be related to the actions of the eicosanoid pathway, although further work is required to verify this.

## 1.5. Pluripotent stem cell applications - embryonic stem cells

Various laboratories have demonstrated the utility of FTIR microspectroscopy for elucidating the biochemical differences that distinguish the spectra of murine and human

embryonic stem cells from their lineage committed progeny. In the first published hESC study (Heraud et al., 2010), our colleagues showed that the differences between spectra of hESC lines and their derived cell types could be detected after only four days of differentiation.

In the work of Heraud et al. (2010), FPA-FTIR microspectroscopy was used to acquire spectra of hES cells that were differentiated towards mesendoderm (precursor cells of mesoderm and endodermal embryonic germ layers) induced by a bone morphogenetic protein cytokine cocktail (BMP4/Act A medium) or towards ectodermal lineages (precursor cells of the skin and nervous system) in medium supplemented with fibroblast growth factor (FGF2) (Figure 1.5). Across all three experimental replicates, the undifferentiated human embryonic stem cell line, $MIXL1^{GFP/w}$, could be discriminated from its differentiated progeny due to spectra of the undifferentiated cells possessing higher IR absorbances for lipid and glycogen components. The distinctness of the spectra of the progeny cells from the spectra of the stem cells was further confirmed using various multivariate data analysis approaches including PCA and PLS-DA, in addition to ANN modelling.

Others who have exploited this tool to study embryonic stem cell differentiation have utilised mouse derived cells. One of the first groups to do so employed FTIR microspectroscopy and PCA-LDA approaches to identify the unique marker bands of spontaneously differentiated murine embryonic stem cells that occurred during known biological stages of ES cell differentiation (Ami et al., 2008). Multivariate analysis revealed an excellent separation of spectral clusters between the undifferentiated cells and cells that were differentiated at days 4, 7, 9, and 14, in the PCA-LDA score plots. The group successfully identified pronounced protein secondary structural changes, in the 1700–1600 cm$^{-1}$ wavenumber region between undifferentiated cells and their

**Figure 1.5:** (A) Scores plots for PLS analysis of the spectral data for the three treatment groups from one experiment showing replicate samples for each cell type. Each spectrum is represented as a point in PC1 vs. PC2 vs. PC3 space. (B) Regression coefficient plots used to explain the clustering observed in the PLS scores plot shown in panel A. PC1 regression coefficient loadings indicated spectral differences explaining clustering along PC1, which separates hESCs from the cells differentiated in FGF2 or BMP4/Act A. PC3 regression coefficient loadings explained the separation of the spectra from FGF2-and BMP4/Act A-treated sample groups along PC3. Limited spread in the PC2 direction did not contribute significantly to the clustering of samples (Heraud et al., 2010).
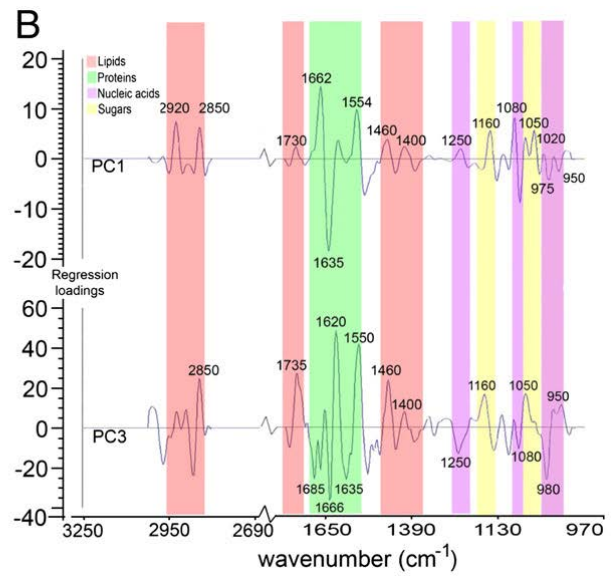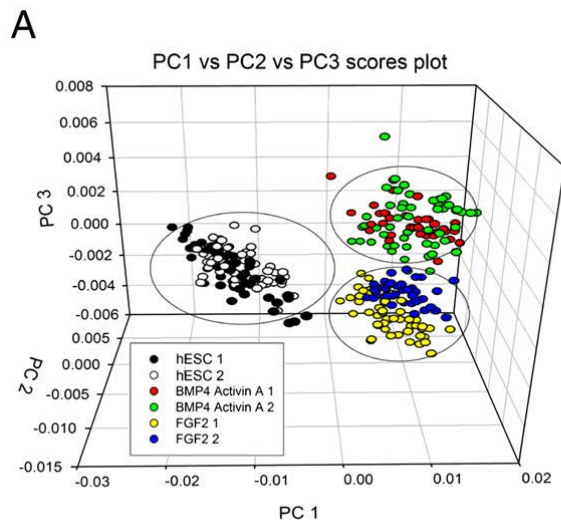
cardiomyoctye precursors. Additionally, they observed a marked increase in the expression of α-helical structures between day 0 and day 10 of differentiation, and noticed that the β-turn secondary structures were enriched, although the latter was not present in the undifferentiated cells. Further to these overall protein changes, the authors reported simultaneous intensity decreases at 994 cm$^{-1}$ and at 914 cm$^{-1}$, which were assigned to the nucleic acid absorption bands, up to 9 days after differentiation, suggestive of the occurrence of mRNA translation. Other observations were a decrease in the 966 cm$^{-1}$ band intensity, assigned to the DNA C-C stretching of the backbone and RNA ribose phosphate main chain modes in undifferentiated cells. Whereas between days 4 and 7 of differentiation, they believed that the active transcription of the genome was switched on, as evidenced by the appearance of new bands at 954 cm$^{-1}$ and at 899 cm$^{-1}$ (assigned to the C–C vibration of the A-DNA backbone and to the deoxyribose ring vibration respectively). In another successful application of FTIR microspectroscopy to discriminate different types of differentiated mESCs, synchrotron FTIR microspectroscopy was implemented to study the differentiation of stem cells into hepatocyte-like cells (Thumanu et al., 2011). The differentiation of mouse embryonic stem cells into hepatocyte-like cells is a complex, multistage process, which first consists of endoderm induction, followed by hepatic initiation, and finally, maturation into hepatocyte-like cells. Taking advantage of the high brightness of the synchrotron infrared radiation, the group was able to accurately discriminate between the three differentiating cell populations at the single cell level.

For instance, mature hepatocyte-like cells were found to have heavy loadings at 1656 cm$^{-1}$, assigned to α-helix protein secondary structures, when compared with spectra of progenitor cells, which were seen to cluster away from these groups along PC1. This finding was corroborated by the fact that mature hepatocyte like cells express proteins characteristic of mature albumin, which is rich in α-helical structures, and the α-feto-protein. Whereas

along PC2, a separation was observed between spectral clusters of progenitor cells from spectra of the mature hepatocyte cells, with the former having heavy loadings at 1627 cm$^{-1}$, assigned to $\beta$-sheet secondary structure of proteins. Lipid differences could also explain the separation observed along this PC, with spectral clusters of hepatic progenitor cells having higher absorbances for loadings bands associated with C-H stretching regions at 2852 cm$^{-1}$ and 2923 cm$^{-1}$, and the ester carbonyl stretching mode at 1740 cm$^{-1}$. Subsequent PLS modelling of the spectral datasets showed that the three differentiated cell types could be discriminated with very high sensitivity and specificity.

FPA and synchrotron based FTIR microspectroscopy (SR-FTIR) has proved useful for discriminating mouse embryonic stem cells differentiated towards embryoid bodies (EBs), from neural progenitor cells (NPCs), and embryonic stem derived neural cells (ESNCS) (Tanthanuch et al., 2010). The authors observed macromolecular chemical differences between the different differentiated cell types, mainly occurring in the spectral regions assigned to protein and lipid absorbance. During the differentiation of NPCs and ESNCs, protein secondary structural changes were observed, with increases in $\alpha$ helical and decreases in $\beta$ sheet secondary structure. These findings were in agreement with existing non spectroscopic data which have shown similar higher levels of $\alpha$ helical structure proteins occurring with neural differentiation, possibly due to alterations in cytoskeleton proteins. Further, the differentiation of neural progenitors and mature neural cells were found to correspond to changes in the peak position and spectral intensity of the asymmetric and symmetric CH$_2$ stretching modes, and ester carbonyl band ascribed to lipid components, thought to be due to increased expression of glycerophospholipids, which are involved in neural cell proliferation, differentiation, and signal transduction. Spectral differences from each stage of differentiation could be seen via interrogating spectra of both clumps, and individual cells, and the biochemical entities explaining the differences

between the EBs and the ESNCs were verified using PCA and unsupervised hierarchical cluster analysis (UHCA).

## 1.6. Pluripotent stem cell applications – induced pluripotent stem cells

Most of the pluripotent lines that that have been investigated by FTIR microspectroscopy have been of embryonic origin, with the extension of this approach to autologous cell (somatic cell derived) induced pluripotent stem cells only presented in the past year. The first group to publish in this area employed synchrotron FTIR microspectroscopy to analyse the FTIR spectra of several ES and iPSC lines (Sandt et al., 2012). In their first paper, the authors reported that spectra acquired from six ES lines and six unrelated iPSC cell lines were biochemically similar, as indicated by their spectral clusters co-localising in the scores plot, and further confirmed by the PLS-DA prediction modelling data.

Sandt et al. (2012) noticed molecular differences between the infrared spectra of the parental somatic amniotic fluid cells (AFC) and their iPSC derivatives, indicated by their spectral clusters segregating along PC1 of the PCA scores plot (Figure 1.6). The PC loadings, which best described the spectral variance causing the segregation were the bands assigned to the lipids ($1740 \text{ cm}^{-1}$, $1710 \text{ cm}^{-1}$, $1465 \text{ cm}^{-1}$, $1455 \text{ cm}^{-1}$, $1170 \text{ cm}^{-1}$), the proteome ($1650 \text{ cm}^{-1}$, $1635 \text{ cm}^{-1}$, $1550 \text{ cm}^{-1}$), and phosphorylation ($1270 \text{ cm}^{-1}$, $1074 \text{ cm}^{-1}$). These biochemical disparities were confirmed by PLS-DA modelling which could successfully discriminate spectra from both stem classes with a 100% accurate prediction. Similarly, spectral differences could also be seen between the murine iPSC cell line (M2A1) and the murine embryonic fibroblast cells (MEF) used in their generation. Again, their macromolecular dissimilarities could be seen by the separation of scores in the PCA scores plot, and by the excellent classification by PLS-DA modelling.

**Figure 1.6.** Analysis of infrared spectra obtained from human and murine ESC as compared to human and murine iPSC and the corresponding somatic cells (B) Comparison of iPSC and non-isogenic hESC lines by Synchrotron FTIR microscopy. Six ESC (H1, H9, HUES3, CL01, CL03, CL04) (blue) and six unrelated iPSC cell lines (PB03, PB04, PB08, PB09, PB10, PB13) (red) were compared by MVA. The scores plot based on PC1 and PC2 of Principal Component Analysis (PCA) shows that iPSC and ESC spectral signatures are very similar. (C) Loading plot corresponding to the comparison of iPSC and hESC. PC1 and PC2 show that the dispersion of iPSC and ESC spectra is related to changes in protein: nucleic acid, protein: glycogen contents, and in the proteome. (D) Comparison of spectral signatures of human somatic amniotic fluid cells (AFC) and their iPSC derivatives. Two independent AFC populations (blue) and their derived iPSC (PB03, PB04) (red) were compared. As can be seen in scores plot of the Principal Component Analysis AFC and iPSC derived from them could be differentiated by their spectral signatures. (E) Loadings plot corresponding to the comparison of AFC and their iPSC derivatives. PC1 and PC4 allow separating AFC and iPSC spectra from changes in the lipid signal (1740 cm$^{-1}$, 1710 cm$^{-1}$, 1465 cm$^{-1}$, 1455 cm$^{-1}$, 1170 cm$^{-1}$), in the proteome signal (1650 cm$^{-1}$, 1635 cm$^{-1}$, 1550 cm$^{-1}$), and in the cellular phosphorylation (1270 cm$^{-1}$, 1074 cm$^{-1}$). (F) Comparison of spectral signature of MEF with the M2A1 iPSC and murine ESC. Scores plot of the PLS-DA show distinct and separated spectra of MEF (green), M2A1 (red) and 4 murine ESC (CJ7, R2, D3 and GS2) (blue). M2A1 clustered with the murine ESC. (G) Loading plot corresponding to the comparison of murine iPSC and mESC. PC1 and PC2 explaining the spectral differences between murine stem cells and MEF in the proteome and nucleic acid ranges (Sandt et al. 2012).
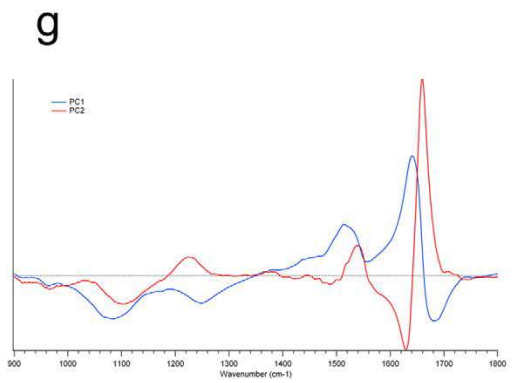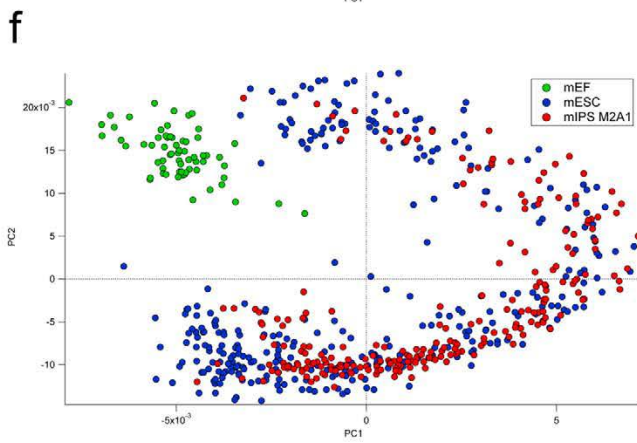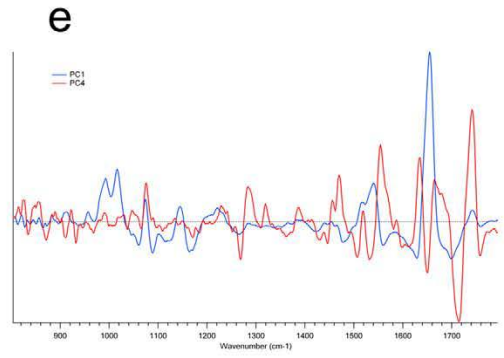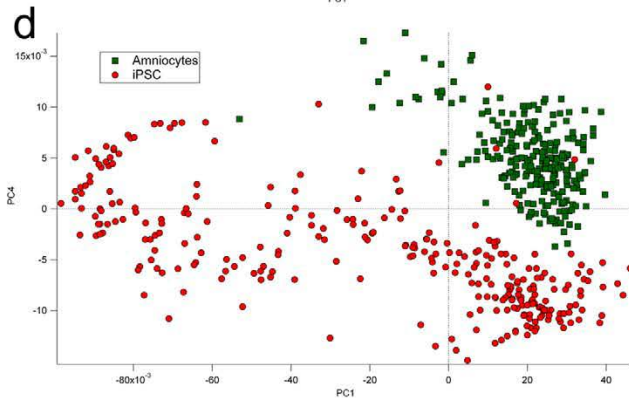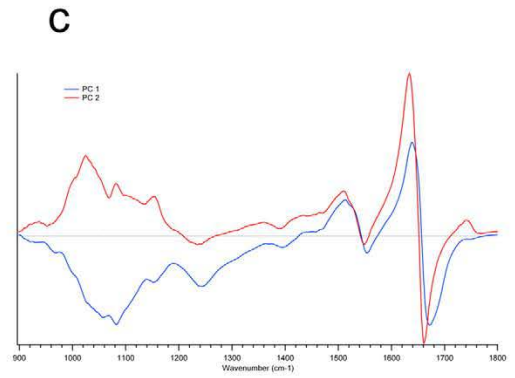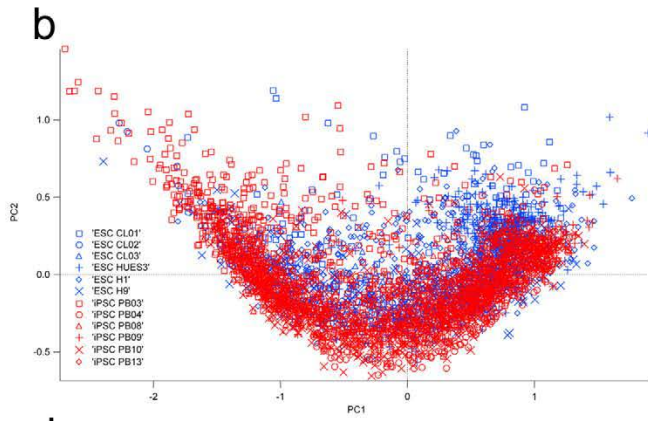
In a further comparison, the FTIR spectra of ESC-H9 mesenchymal stem cells generated from H9 cells (MSC-H9) and iPS cells derived from the MSC-H9 cells (iPSC-H9) were all analysed via multivariate analysis where it was found that the MSC-H9 spectra clustered away from spectra of the colocalising ESC-H9 and iPSC-H9 lines due to the former having higher lipid to protein, and lower nucleic acid to protein ratios. Whilst a PLS-DA model using the first 3 PCs explaining 85% of the spectral variance could not discriminate the two co-localising stem cell types from each other, a PLS-DA model using 8 PC factors, explaining 99% of the spectral variance was able to discriminate between the two co-localising stem cell types with high accuracy.

However, the most noteworthy finding from the study was the demonstration that the technique was also successful in discriminating fully programmed iPSC lines from partially reprogrammed iPSC lines, with PCA analysis showing a separation of scores pertaining to cells of the two stem cell variants along PC2 and explained by bands in loadings plots associated with increased lipid and glycogen storage in the iPSCs-PR.

Subsequent PLS-DA modelling capturing 98% of the variance (10 factors) was then able to correctly classify the iPSC-FR spectra from the iPSC-PR spectra with 100% sensitivity and specificity. Additionally, biochemical variability was observed in the spectral regions ascribed to protein, glycogen, lipid, and nucleic acids between the two cell types at different stages of reprogramming.

Nevertheless, the problem with utilising such a large number of factors is that features such as noise or spectral regions that are not of interest to the investigation but are cross correlated may be included in the prediction, thus producing an over-fitted model and reducing its robustness. This was demonstrated by researchers who upon trying to predict glucose features in their fermentation samples, found that other analytes were being

included in the predictive analyses when they used more than three PCs in spectral models (Kansiz et al., 2005).

In their most recent study, Sandt et al. (2013) compared the IR spectral profiles of two amniocyte derived IPSC lines that had been generated using different reprogramming methods. The iPSC line PB09 was transduced with four VSVG-pseudotyped lentiviruses carrying Oct4, Sox2, Lin28 and Nanog (OSLN) vectors. Whereas the iPSC line PB10 was transduced using only two exogenous transcription factors, Oct4 and Sox2. Whilst the mean spectra of both these cell lines appeared strikingly different, the groups could not be separated via the first two principal components along a PCA scores plot, although intra-class separation was observed within the PB10 line, along PC3. Further, PLS-DA modelling of spectra from the two lines showed poor prediction results, with low sensitivity or specificity. From this data, the group concluded that the use of two transcription factors appeared to be sufficient for the generation of biochemically, and thereby, spectroscopically identical iPS cells.

Whilst these findings are interesting, the authors' claims would have been strengthened by the investigation of a greater number of cell lines. Another suggestion for follow up studies would be to determine whether these observed biochemical similarities persisted upon differentiation.

## 1.7. Multipotent stem cell applications

### 1.7.1. Mesenchymal stem cells

FTIR microspectroscopy has been applied to a varied range of adult stem cell types in multipotent stem cell studies. Krafft et al. (2007) were the first group to probe human mesenchymal stem cell (hMSC) differentiation at the single cell level. In their work, the IR spectra of several hundred single human mesenchymal stem cells with or without

osteogenic stimulation were recorded, with the non-stimulated control cells mainly consisting of populations that had either high or low levels of peripheral glycogen levels, whereas stimulated cells showed alterations in protein composition and expression of octacalciumphosphate, a calcium phosphate salt (Krafft et al., 2007). However, the main discriminants between the non-stimulated and stimulated classes were attributed to protein differences. For instance, spectra of the non-stimulated hMSCs consisted of a more intense peak at 3285 cm$^{-1}$ and at 1631 cm$^{-1}$, the spectral regions which were assigned to the amide A and β sheet protein secondary structures respectively. A criticism of this study is that the absorbance at the amide A band 3285 cm$^{-1}$ band can also arise from the O-H stretching mode of water and since this wasn't controlled for, it is feasible that differences in this peak intensity could be partially ascribed to differences in the hydration state of the samples.

The spectroscopic technique has also been adopted for studies of rat bone marrow mesenchymal stem cells (rBM-MSCs) differentiated to hepatocytes (Ye et al., 2012). The main findings from the work were that the late stage differentiated cells possessed significantly higher levels of lipid compared to the stem cell progenitors and early stage differentiated cells, with notable differences in the FTIR absorbance bands at: 3012 cm$^{-1}$ (*cis* C=C stretch from unsaturated lipids), 2952 cm$^{-1}$ ($v_{as}CH_3$ from lipids), 2854 cm$^{-1}$ ($v_sCH_2$ from lipids) and 1722 cm$^{-1}$ (C=O stretching from lipids). The observed increase in lipid content during hepatocyte differentiation differed from what was found by Thumanu et al. (2011) who had studied the hepatogenesis of mouse embryonic stem cells and observed higher lipid levels in the progenitor cells. These disparities could have been attributed to differences in lipid signalling between the two stem cells types. Further work that would provide a greater insight into these findings could involve inducing

hepatogenesis in both stem cell variants in parallel, and then comparing the derived hepatocyte like cells to each other.

The results from Ye et al. (2012) aligned with non-spectroscopic studies which report that hepatocytes routinely synthesise lipoproteins and synthesise and store triglycerides as a result of activities from carbohydrate metabolic pathways. Moreover, a significant increase in unsaturated fatty acid levels was seen in the late stage hepatocytes, indicated by an increase in the intensity of C–H stretching band associated with the *cis* double bond C=C at 3010 cm$^{-1}$. However, changes in band intensity in spectral regions representing proteins and nucleic acids are more complex and therefore more difficult to interpret compared to the studies carried out by Thumanu and colleagues.

In another study investigating this stem cell type, synchrotron FTIR microspectroscopy was employed for the first time to characterise the chondrogenic differentiation of human mesenchymal cells generated via the pellet culture method with and without stimulation by growth factors (TGF-β3 and BMP-6), over a period of 7, 14 and 21 days (Chonanant et al., 2011). Inspection of the average second derivative spectra revealed that the chondrocyte induced hMSCs were characterised by having higher absorbances at 1338 cm$^{-1}$, 1230 cm$^{-1}$ and 1203 cm$^{-1}$, assigned to the amide III, P–O stretching and C–O–C stretching modes of collagen type II respectively. Further, spectra of the chondrocyte induced group had higher absorbances at 1152 cm$^{-1}$, 1107 cm$^{-1}$, 1080 cm$^{-1}$, 1019 cm$^{-1}$, and 993 cm$^{-1}$, attributed to aggrecans, which were higher in chondrocyte-induced hMSCs than in the controls. The ability to detect chondrogenic differentiation was found to be high, with a clear separation of spectral clusters in scores plots and the regression loadings revealing differences that followed a similar pattern to what was observed in the average second derivative spectra. PLS-DA using the independent test spectra showed that the two spectral classes could be discriminated from each other with 100% accuracy.

Considering the high heterogeneity of pellet culture differentiated cells, compared to those derived from monolayer-based protocols, the exploration of the latter method in future work would be a worthy pursuit. Moreover, the researchers who examined paraffin sections need to be mindful that biochemical changes can arise from tissue processing, which will affect the spectral signatures of the sample. This was shown in experiments by Ó Faoláin et al. (2005) who saw shifts of 10 cm$^{-1}$ in the amide I and II bands, ascribed to the crosslinking of proteins, caused by formalin fixation. The authors advised against the diagnostic use of biological bands that are close to fixative peaks because even after careful washing, trace remnants may still persist.

### 1.7.2. Corneal stem cells

SR-FTIR microspectroscopy was undertaken by one research group in their experiments using bovine corneal epithelium (German et al., 2006). The cell types that they investigated were the lifelong proliferating putative adult stem cells (SC), the SC derived progenitor or transiently amplified (TA) cells which have only limited proliferative capacities, and the TA-derived, non-proliferative terminally differentiated (TD) cell populations. PCA of the spectra acquired from the corneal sections, showed the distinct clustering of spectra from the three different cell types, with only a slight overlapping of the spectral clusters of the SC and TA groups, possibly due to the presence of a small population of TA cells that had yet to migrate out of SC niche at the time. Statistically significant spectral differences ($P$ <0.001), as determined by the Mann Whitney test, were found to occur at the following peaks: 1714 cm$^{-1}$, 1600 cm$^{-1}$, 1450 cm$^{-1}$ to 1440 cm$^{-1}$, and 1225 cm$^{-1}$, all of which are known to be associated with nucleic acids. The spectra of the TD cells, whilst appearing distinctly clustered away from the other spectral groups in the PCA scores plot, were found to be more spectrally aligned with the TA cells than to the SCs, although the TA and TD spectra showed marked dissimilarities, mainly in the regions assigned to protein and RNA.

All of the observed differences were thought to be related to changes in chromatin structure, which is known to be linked to cell differentiation.

SR-FTIR microspectroscopy combined with spectral imaging has also been used successfully for the characterization and localisation of the biomarkers of SCs, TA cells, and TD cells in *human* derived corneal sections (Nakamura et al., 2010). PCA of the single point spectra acquired from the three putative cell type regions (SC versus TA cell versus TD cell) revealed an excellent separation between the spectral classes. Wavenumbers that were found to be differentially absorbed by the SC classes compared to the other two cell types were ascribed to DNA (1040 cm$^{-1}$, 1080 cm$^{-1}$, 1107 cm$^{-1}$ 1225 cm$^{-1}$), with some contributions from C–O stretching in carbohydrates derived from amino acid side chains or lipids at 1400 cm$^{-1}$, amide II absorption bands at 1525 cm$^{-1}$ and 1558 cm$^{-1}$, and the lipid associated band at 1728 cm$^{-1}$.Whereas protein/lipid biochemicals were the entities that most distinguished TA cells and TD cells. These two studies demonstrate the utility of SR-FTIR microspectroscopy to discriminate and segregate putative populations of SCs, TA and TD cells of the bovine and human corneal epithelium.

### 1.7.3. Gastrointestinal crypt stem cells

This spectroscopic approach has proved useful for studying the complex stem cell niche of the small and large human intestine. Both globar and synchrotron based FTIR microspectroscopy were employed by a leading group in an attempt to segregate and characterise the putative stem cells, transit amplifying cells and differentiated cells of human intestinal crypts (Walsh et al., 2008). These cells reside in different regions along the length of the gut, with locations differing between the small intestine, and large intestine originated crypts. PCA-LDA analysis of spectra obtained from the aforementioned cell types showed a clear segregation of scores related to the various classes. The most prominent separation of scores was attributed to the spectral region

assigned to DNA/RNA, with the symmetric $PO_2^-$ (1080 cm$^{-1}$) vibrations found to be a marker for the putative stem cell region possibly due to changes in chromatin. These same biomolecular signature differences between the different stem cell classes were found to be similar regardless of whether the cells were derived from small intestine or large intestinal crypts.

The same authors in another study used synchrotron FTIR microscopy to generate IR image maps of small intestinal and large intestinal crypts (Walsh et al., 2009) (Figure 1.7). The cell types that were interrogated were assigned step-wise positions along the length of the crypt. In the small intestine, the cell types and their positions were as follows: cells residing in the crypt base (Position 1), crypt base columnar (CBC)/Paneth cells (Position 2,3), label retaining cells (LRCs) (Position 4–6), transit amplifying (TA) cells (Position 7,8), and terminally differentiated (TD) cells (Position 9,10). Whereas the putative cell types and assigned positions that were probed in the large intestine crypts were the putative stem cells (Position 1-4), TA cells (Position 5-8), and TD cells (Position 9,10). PCA-LDA of spectra derived from both small intestine and large intestine crypts found that the spectral regions that most contributed to the variance segregating different putative cell types occurred at 970 cm$^{-1}$ (protein and nucleic acid phosphorylation), $v_sPO_2^-$ (1080 cm$^{-1}$) and $v_{as}PO_2^-$ (1225 cm$^{-1}$).

Analysis of individual and combined spectra acquired from the small intestine derived crypts showed that the greatest biochemical differences were between the spectra of the cells that resided in Position 1, and TD cells, and that the base of small intestine crypts was the most spectrally distinct region. Spectra from the intermingling CBC and Paneth cells were found to be spectroscopically disparate from the spectra of cells that resided in Position 1, whereas they shared biochemical similarities with LRCs and TA cells in spectral regions pertaining to DNA/RNA and protein, with the latter region causing the

**Figure 1.7:** Localization of the putative stem cell region in an IR spectral image (resolution=10μm×10μm) map of human GI crypts using synchrotron FTIR microspectroscopy. (A), A section of small intestine tissue stained with H&E post-interrogation with synchrotron FTIR microspectroscopy – the imaged area is designated by the black box; (B), a two-dimensional map of a small intestine crypt, smoothed at wavenumber 1080 cm$^{-1}$ and superimposed on the unstained region analysed – see inset (Walsh et al., 2009).
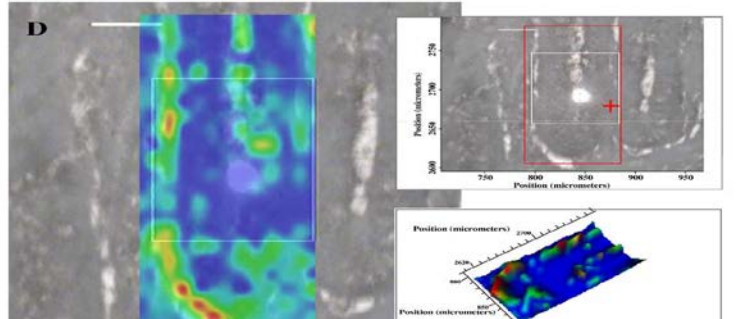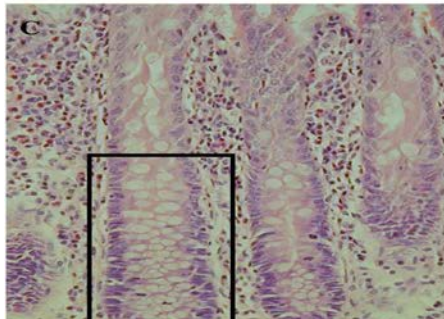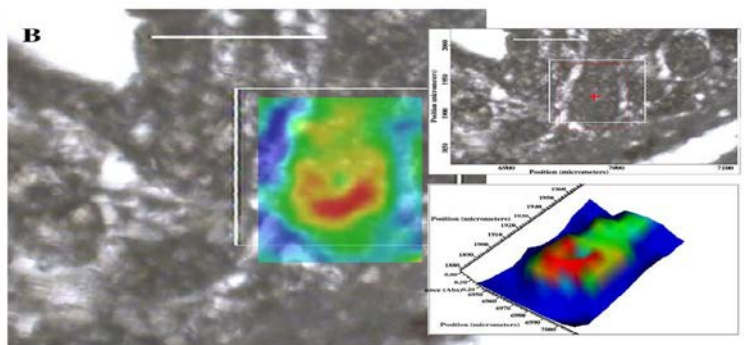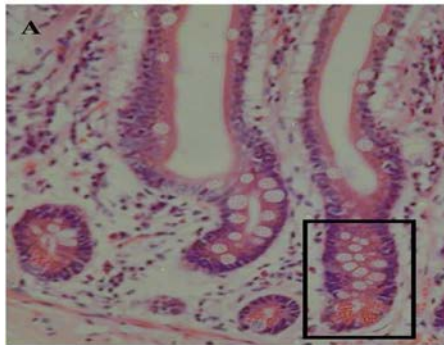
most prominent separation. However in the case of the spectra derived from large intestine crypts, the greatest segregation of spectral clusters occurred between the putative stem cells located in Positions 1-3 away from the closely clustered TA and TD spectra. Conversely, putative stem cells derived from Position 4 often clustered closely with TA derived spectra.

Upon comparing the spectra of these cell types from both the small and large intestine crypts, it was observed that relative to the spectra of the CBC/Paneth cells, the label retaining cells were spectroscopically similar to the large bowel derived putative stem cells. Whereas the small intestine derived spectra of the TA cells were spectroscopically similar to the spectra of the large intestine-derived TA cells. Interestingly, inspection of the PCA scores plot showed the Paneth spectra clustering between the clusters of LRC and TA clusters, suggesting that these cells possessed a biochemical 'fingerprint' that was intermediate between these two putative cell types.

## 1.8. Other applications of FTIR microspectroscopy to stem cell research

### 1.8.1. Haematopoietic stem cells

Globar FTIR microspectroscopy has been able to discriminate isolated murine haematopoietic stem cells (HSCs) from bone marrow (BM) cells (Zelig et al., 2010). PCA of spectra derived from these cells, revealed that they could be distinguished from each other as evidenced by the separation of their spectral clusters along PC1, with the biochemical entity which most contributed to the observed variance being DNA, assigned to the loadings bands at 966 cm$^{-1}$, 1088 cm$^{-1}$ and 1240 cm$^{-1}$. These bands were found to have higher IR intensities in the HSCs compared to the BM cells, possibly due to the former having a unique chromatin conformation. HSCs could also be identified from BM cells by having higher lipid absorbances, suggestive of the stem cells having a unique

membrane composition and structure. When determining the utility of this methodology for quantifying HSCs in the bone marrow by using the characteristic biomarkers of HSC which were previously identified to be wavenumbers associated with DNA, the group found that one HSC per thousand BM cells could be successfully identified.

## 1.8.2. Cancer stem cells

The lack of specific cancer stem cell biomarkers for the isolation of putative stem-like cells from either normal or malignant renal tissue, prompted one group to explore the ability of high brilliance synchrotron-FTIR spectroscopy to discern the unique spectral markers of three isolated cell types residing in this tissue (Hughes et al., 2010). The extraction of these cell types was achieved via employing the Hoechst 33342 dye efflux assay, from which the 3 sub-populations that were used for analysis were the non-SP (side population) containing differentiated cells and 2 SP (sub-populations), the proximal side population (PSP) and the distal side population (DSP). Of these groups, the latter two subtypes were known to consist of putative transit amplifying cells, and the most primitive (stem) cell types respectively. PCA of spectra acquired from the different cell types revealed a clear segregation of scores pertaining to the Non-SP group away from the DSP and PSP group, with the key biochemical differences explaining the separation being changes in lipids, phosphodiester groups and carbohydrates. Principal component–linear discriminant analysis (PC–LDA) revealed that the independent validation DSP, PSP and Non-SP dataset had at least 80% of their spectra correctly assigned.

When pairwise PCA comparisons were then performed using only the spectral datasets of the DSP and Non-DSP groups, a segregation of scores from the two cell types was witnessed, with spectra from all cell subtypes from the DSP group found to be more spectrally homogeneous, tightly clustering together compared to the more spectrally heterogeneous spectra from the Non-SP groups, which were found to form two groups.

The main biochemical moieties causing this spectral discrimination were the lipid components, denoted by loadings bands at 1468 cm$^{-1}$ and at 1380 cm$^{-1}$. Whereas along PC2, the PC where the non-SP spectra segregate into two clusters, the loadings bands were dominated by carbohydrates and symmetric and asymmetric phosphodiester signatures. Finally, the group compared the spectral datasets of the DSP and the PSP cells and found a clear separation between the two cell types along PC1 due to differences in the phosphodiester stretching bands vibrations. By using FTIR microspectroscopy, the group established that the lipid and phosphodiester vibrations were the most important markers for discriminating stem cell-like cells from the more differentiated variants.

## 1.9. Unique macromolecular chemical signature differences exist between undifferentiated and differentiated stem cells

A common observation in both of the pluripotent and multipotent stem cell studies was that stem cells undergoing differentiation possessed different lipid 'signatures' from their lineage committed progeny. However, there has been no consensus on a 'lipid profile' that is common to all stem cell types, although the possible effect of culturing environment on these spectral profiles cannot be dismissed. For instance, in the hepatocyte differentiation studies performed by Thumanu et al. (2011), and human embryonic stem cell differentiation work carried out by Heraud et al. (2010), lipid depletion was found to co-occur with a loss of pluripotency. However, the contrary was found to be the case by Tanthanuch et al. (2010) in their study of mouse embryonic stem cell derived neuronal cells, and by Ami et al. (2008), who did not report any significant lipid differences between their undifferentiated and spontaneously murine differentiated stem cells, but instead noticed differences in protein and nucleic acid spectral signatures between the stem cells and their derived cell types.

The interpretation of protein information, especially when the amide I band is involved, needs to be done carefully due to spectral signatures such as resonant Mie-scattering being particularly prominent in this spectral region. For example, Thumanu et al. (2011) were able to verify that the observed amide I band changes seen in their study were actually protein related, rather than artefactual, by noting that these alterations also coincided with the appearance of new band features, rather than simple wavenumber shifts in the second derivative spectra. Further, the authors observed that these protein changes coincided with albumin formation, a protein alteration known to occur during the differentiation of hepatocytes. Care must also be taken when interpreting changes in the nucleic acid region given that the intensity of these band features are affected by different levels of hydration. An example of this could be seen in work by Whelan et al. (2011) who showed that DNA specific bands in several eukaryotic cells were more intense in hydrated samples compared with dehydrated samples

## 1.10. FTIR spectroscopic signatures of stem cells are influenced by the growth environment

In light of the findings that FTIR microspectroscopy can be used to classify pluripotent and multipotent stem cells based on their spectral phenotype, there are preliminary studies appearing, addressing the question of how the environment might affect the results. The genotypic and phenotypic consequences of using different culturing conditions for the maintenance and differentiation of stem cells has been well established by non-spectroscopic methods (Akopian et al., 2010; Maitra et al., 2005; Sjögren-Jansson et al., 2005). By contrast, the influences of the growth milieu on stem cell spectral biomarkers have been the subject of only two studies to date (Cao et al., 2013a; Pijanka et al., 2010). The first group to explore this question tested the effect of two different oxygen

concentrations (2% and 21%) on the FTIR spectroscopic signatures of hESC and hMSC cells using synchrotron radiation FTIR microspectroscopy. The group found spectroscopic differences between the hESC and hMSC spectra when cultured in either oxygen concentration, with the former displaying more intense peaks in spectral regions assigned to lipid components (2850 cm$^{-1}$, 2920 cm$^{-1}$ and 1740 cm$^{-1}$). The findings that the spectra from the more differentiated hMSCs possessed a lower lipid content than the less differentiated hESC spectra, is in agreement with data reported by our colleagues, and by others. At present, the functional significance of these lipid differences is not yet known although it has been speculated that the eicosanoid pathway may play a role (Yanes et al., 2010). However, the observation of higher lipid levels in more differentiated stem cells by other groups suggests that this may be a cell-type specific phenomenon and will warrant further investigation.

## 1.11. The importance of correlative methodologies for the interpretation of stem cell spectral signatures

To facilitate the interpretation of spectroscopic data, particularly when differences reside in spectral regions containing band contributions from multiple macromolecular classes, it is critical to correlate these findings with those derived from non-spectroscopic methodologies. Common techniques that have been used in conjunction with FTIR microspectroscopy stem cell studies have included morphological analysis, flow cytometry, gene expression analysis, fluorescence studies, functional colony assays and histological staining. Despite its popular adoption in stem cell spectroscopy work, genotypic methods such as gene chip analysis may not be directly comparable to findings derived from spectroscopy, since the latter, being a phenotypic technique, is best compared to data derived from other phenotypic protocols. For example, a problem of using gene

chip analysis, as was demonstrated in a study by our laboratory (Heraud et al., 2010) is that the data may be uncoupled from the spectroscopic information due to the occurrence of events such as post-transcriptional modifications.

## 1.12. Aims of the study

FTIR microspectroscopy of stem cells is a nascent, but promising, research area which is providing an objective, non-destructive and holistic approach to the studying of these specialised cell types. The aims of this study were to determine its potential to elucidate the unique spectral 'signatures' of different strains of human embryonic and human induced pluripotent stem cells and their lineage committed progeny. Since a wealth of different tissue culturing methodologies exist, the effect of the environment on these spectral phenotypes were also compared.

| Band maxima (cm$^{-1}$) | Band assignments |
|---|---|
| ~2962 | C-H stretching peak of lipids and proteins |
| ~2923 | C-H asymmetrical stretching of lipid groups and protein |
| ~2850 | C-H symmetrical stretching of lipid groups |
| ~1743 | C=O stretching of lipid esters |
| ~1685 | β-turn protein secondary structure |
| ~1654 | α-helical protein secondary structure |
| ~1635 | β-pleated sheet protein secondary structure |
| ~1554 | Overall protein absorbance |
| ~1458 | Methyl and methylene groups from lipids and protein |
| ~1396 | COO$^-$ stretching vibrations of amino acid side chains |
| ~1238 | P=O asymmetrical stretching of $PO_2$ phosphodiester groups from phosphorylated molecules |
| ~1080 | P=O symmetrical stretching of $PO_2$ phosphodiester groups from phosphorylated molecules, and glycogen |
| ~1153 | C-O vibrations from glycogen and other carbohydrates |
| ~1050 | C-O vibrations from glycogen and other carbohydrates |
| ~1022 | C-O vibrations from glycogen and other carbohydrates |
| ~995 | C-O stretch from RNA ribose chain and other carbohydrates |
| ~950 | C-C vibrations from nucleic acids |

**Table 1.1**: Band assignments of mid-IR spectra common to biological samples

CHAPTER $2$

Fourier transform infrared microspectroscopy reveals macromolecular differences between human embryonic and induced pluripotent stem cells

# Monash University

# Declaration for Thesis Chapter 2

## Declaration by candidate

In the case of Chapter 2, the nature and extent of my contribution to the work was the following:

| Nature of contribution | Extent of contribution (%) |
| --- | --- |
| The work described in the paper by Cao et al. was predominantly performed by myself. I was also involved in the designing and analysis of the experiments performed, and did the majority of the writing. | 80 |
| The remainder of Chapter 2 was written by myself | 100 |

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

| Name | Nature of contribution | Extent of contribution (%) for student co-authors only |
| --- | --- | --- |
| Elizabeth Ng | Technical assistance, data interpretation, critical revision and final approval of the manuscript | 5% |
| Edouard Stanley | Data interpretation, critical revision and final approval of the manuscript | |
| Donald McNaughton | Supervision, data interpretation, critical revision and final approval of the manuscript | |
| Mark Tobin | Supervision, data interpretation, critical revision and final approval of the manuscript | |
| Andrew Elefanty | Supervision, data interpretation, critical revision and final approval of the manuscript | |
| Philip Heraud | Supervision, data interpretation, critical revision and final approval of the manuscript | |

| **Candidate's Signature** | ██████ | **Date** 04/06/13 |
|---|---|---|

**Declaration by co-authors**

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

| **Location(s)** | 1.Monash Immunology and Stem Cell Laboratories (MISCL) - now part of the Department of Anatomy and Developmental Biology, Monash University, 2.Centre for Biospectroscopy, Monash University |
|---|---|

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

| | Author name | Signature | Date |
|---|---|---|---|
| **Signature 1** | Elizabeth Ng | ██████ | 04/06/13 |
| **Signature 2** | Edouard Stanley | ██████ | 04/06/13 |
| **Signature 3** | Donald McNaughton | ██████ | 04/06/13 |
| **Signature 4** | Mark Tobin | ██████ | 04/06/13 |
| **Signature 5** | Andrew Elefanty | ██████ | 04/06/13 |
| **Signature 6** | Philip Heraud | ██████ | 04/06/13 |

## 2.1. Introduction

Chapter 1 described the emerging popularity of spectroscopic modalities such as FTIR microspectroscopy for the characterisation of various stem cell types. The Chapter provided a basic outline of the technique, gave a brief overview of the commonly used chemometric analysis tools used for the modelling and interpretation of the data, before describing some of its applications to both multipotent and pluripotent stem cells destined for use in regenerative therapies and also investigated its role in detecting the unique spectroscopic signatures of cancer stem cells.

This Chapter consists of a manuscript which has been published in *The Journal Of Biophotonics* and describes the use of this optical tool for discriminating between several lines of human embryonic (HES3, H9, MEL1, *MIXL1*$^{GFP/w}$, NKX 2.5$^{GFP/w}$, ENVY) and human induced pluripotent stem cell lines (IMR90C2, ES4CL1 and AUS) and their lineage committed progeny.

# Fourier transform infrared microspectroscopy reveals macromolecular differences between human embryonic and induced pluripotent stem cells

*Julie Cao [a,b], Elizabeth S. Ng [a,d], Don McNaughton [b], Edouard G. Stanley [a,d], Andrew G. Elefanty [a,d], Mark J. Tobin [c], Philip Heraud [a,b]*

Monash Immunology and Stem Cell Laboratories, Monash University, Building 75, STRIP 1, West Ring Road, Clayton, Victoria 3800, Australia [a]

Centre for Biospectroscopy and the School of Chemistry, Monash University, Clayton, Victoria 3800, Australia [b]

Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168 [c]

*Murdoch Childrens Research Institute, The Royal Children's Hospital, Parkville, Victoria, Australia, 3052 [d]*

* Corresponding author: e-mail ██████████████ Phone: ████████

**Key words: Fourier transform infrared microspectroscopy, human embryonic stem cells, human induced pluripotent stem cells, regenerative medicine**

## 2.2. Abstract

Fourier transform infrared (FTIR) microspectroscopy was employed to elucidate the macromolecular phenotype of human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) and their differentiated progeny. Undifferentiated hESCs and hiPSC lines were found to be not clearly distinguishable from each other. However, although both hESC and hiPSC variants appeared to undergo similar changes during differentiation in terms of cell surface antigens, the derived cell types from each class could be discriminated using FTIR spectroscopy. We foresee a possible future role for FTIR microspectroscopy as a powerful and objective investigative and quality control tool in regenerative medicine.

## 2.3. Introduction

Human embryonic (hESC) and induced pluripotent (hiPSC) stem cells can form all the foetally-derived cells and tissues of the body and have been advanced as a potential source of new cells for the emerging field of regenerative medicine. However, before their envisaged introduction into the clinic, the question of whether undifferentiated hESCs and hiPSCs and their differentiated progeny are equivalent must be answered (Narsinh et al., 2011) In regards to their pluripotency, ability to form the three embryonic germ layers, and overall global gene expression programs (Guenther et al., 2010; Takahashi et al., 2007; Yu et al., 2009), the two stem cell classes appear to be very similar (Takahashi et al., 2007; Yu et al., 2007). Nevertheless, others have observed subtle differences between their RNA levels, DNA methylation, protein expression and protein phosphorylation (Phanstiel et al., 2011). Some of these differences are attributed to the current limitations of the various reprogramming processes used in the generation of iPSCs (Chin et al., 2009; Marchetto et al., 2009) and may contribute to the variable efficiencies with which hiPSCs can undergo differentiation (Bock et al., 2011).

Our current knowledge concerning iPSCs and their embryonic stem cell (ESC) counterparts has been acquired via the use of conventional molecular biology techniques such as in vitro and in vivo assays, bisulfite sequencing, flow cytometry, and gene array analysis. Although these methods are highly informative, they cannot provide a global overview of the biochemical similarities and differences between these two cell types. An attractive modality that provides such information is Fourier transform infrared (FTIR) microspectroscopy.

FTIR microspectroscopy is a powerful, non-destructive, optical technique that is reliant on the intrinsic property of molecular systems to vibrate in resonance with different frequencies of infrared light. It is able to elucidate the macromolecular chemistry of a wide

spectrum of biological samples and unlike the other conventional methods, can readily attain quantitative and qualitative data without the use of additional reagents or contrast agents. For example, FTIR spectroscopy is sensitive to the relative concentrations of macromolecules such as proteins, lipids, carbohydrates and nucleic acids within biological cells. Vibrational modes in macromolecular functional groups within biological samples give rise to a series of identifiable bands in the FTIR spectrum, providing information about their relative concentrations. Spectroscopic information is therefore complementary to other methodologies used to characterise stem cells and their differentiated progeny in the sense that the differences in macromolecular composition revealed by the FTIR spectrum can provide a unique phenotypic chemical 'signature' for a particular cell or tissue structure (Heraud and Tobin, 2009).

At present, there is only a limited body of literature reporting the application of FTIR microspectroscopy in the study of embryonic and adult stem cells (Ami et al., 2008; Pijanka et al., 2010; Walsh et al., 2008). Furthermore, the only extant study using FTIR to classify human iPSCs (Sandt et al., 2012) has been restricted to undifferentiated cells, whereas iPSC derived lineage committed progenitors have not been examined.

We have previously demonstrated the success of this method in discriminating between undifferentiated human embryonic stem cell lines and their early differentiated mesendodermal and ectodermal progeny (Heraud et al., 2010). As a continuation of this earlier work, we have used laboratory-based FTIR microspectroscopy to elucidate the FTIR spectral profiles of several readily available hESC and hiPSC lines and their early differentiated progeny. To confirm that the differentiation had been successful, flow cytometry was used to verify the presence or absence of well-established cell surface markers indicative of differentiation state. Our study showed that FTIR microspectroscopy was able to successfully discriminate different human embryonic and human induced

pluripotent stem cells from each other and from their mesendodermal and ectodermal lineage committed progeny with high levels of sensitivity and specificity.

## 2.4. Materials and Methods

A schematic of the general methodology used in this study is shown in Figure 2.1.

### 2.4.1. FTIR microspectroscopy of mouse embryonic fibroblasts

A monolayer of inactivated mouse embryonic fibroblasts (MEFs) is commonly used as a substrate in the culture of hES and hiPS cells (Thomson et al., 1998). In our attempts to minimise any spectral contamination that may have arisen from their presence in samples measured by spectroscopy, all MEFs that were used in the experiments were seeded onto the gelatinized tissue culture flasks at half their normal density. Further, the hESCs and hiPSCs were always harvested from cultures that were greater than 80% confluent, to ensure a high stem cell to feeder cell ratio.

Focal Plane Array (FPA) FTIR images comprised of a $64 \times 64$ array of spectra (1024 spectra in total) after the binning of 4 adjacent pixels, were acquired from fields of cells. First, fields of MEF cells in the absence of hESCs and hiPSC were imaged to address the question of how spectroscopically similar the different cell types were to each other, and whether the presence of MEFs would confound the interpretation of their spectra., MEFs were maintained for 2 days on gelatin coated T25 flasks in KOSR-based hESC medium before being dissociated via enzymatic passaging (Costa, 2008) The harvested cells were then washed 3 times in phosphate buffered saline (PBS) before approximately 150,000 cells were cyto-centrifuged (Cytospin III, Thermo Fischer Scientific, Waltham, MA) onto infrared reflective slides (MirrIR slides, Tienta Technologies, OH, USA).

**Figure 2.1.Flow diagram of the experimental procedure employed during the study.** Human ESC lines (HES3, H9, *MIXL1*<sup>GFP/w</sup>, *NKX 2.5*<sup>GFP/w</sup> and ENVY) and hiPSC lines (IMR90C2, ES4CL1 and AUS) maintained in bulk culture were harvested and cytospun onto MirrIR slides for FPA-FTIR microspectroscopy. Some of these lines were also differentiated towards mesendodermal and ectodermal lineages before being cytospun onto the same slides for data acquisition. To verify that the differentiation process had been successful, Fluorescence Activated Cell Sorting (FACS) analysis was performed on the samples to check for the presence or absence of well-established stem cell and differentiation associated cell surface markers.

# Experimental procedure
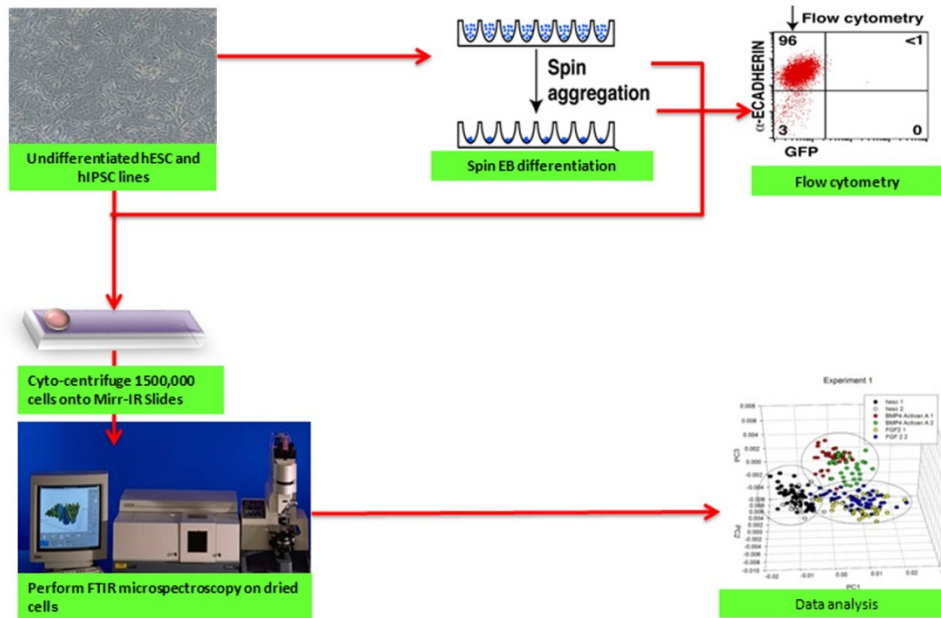
## 2.4.2. FTIR microspectroscopy of undifferentiated human embryonic and human induced pluripotent stem cells

Five human embryonic stem cell lines, HES3 (Richards et al., 2002), H9 obtained from WiCell, and, NKX 2.5$^{GFP/w}$ *MIXL1*$^{GFP/w}$, and ENVY (Costa et al., 2005; Davis et al., 2008; Elliott et al., 2011) in addition to 3 human induced stem cell lines, ES4CL1 derived from foreskin, and lung fibroblast derived IMR90C2 provided by the Australian Stem Cell Centre (Yu et al., 2007), and skin fibroblast derived AUS (Liu et al., 2011) from the Monash Institute of Medical Research, were cultured on top of a low density feeder monolayer on gelatinized T25 tissue culture flasks in KOSR-based hESC medium using methods previously described (Costa, 2008). The medium was changed daily and when the cells were at least 80% confluent, they were enzymatically passaged using TrypLE (Costa, 2008) and washed 3 times in PBS before 150,000 cells were cyto-centrifuged onto MirrIR slides for FTIR microspectroscopy using two replicate cultures.

## 2.4.3. FTIR microspectroscopy of human embryonic and human induced pluripotent stem cells differentiated towards a mesendodermal and ectodermal lineage

Two human embryonic stem cell lines, *MIXL1*$^{GFP/w}$ (a genetically modified hESC line expressing green fluorescent protein (GFP) from the MIXL1 locus) (Davis et al., 2008) and MEL1, in addition to one of the human induced pluripotent stem cell lines IMR90C2, were used for differentiation experiments. Differentiation towards mesendodermal lineages occurred in serum-free medium supplemented with bone morphogenetic protein (BMP4) and Activin A (BMP4/Act A medium) or toward ectodermal lineages in medium supplemented with fibroblast growth factor (FGF2), using the previously described spin embryoid body (EB) method (Ng et al., 2005; Ng et al., 2007). After 4 days of differentiation, the cells were immuno-stained for flow cytometry to confirm that the

differentiation process had been successful. Once differentiation had been confirmed by the expression of the appropriate cell biomarkers, 10 EBs were dissociated with TrypLE select, and washed 3 times with phosphate buffered saline (PBS) before being cytospun onto MirrIR slides.

## 2.4.4. Fluorescence Activated Cell Sorting

To verify that the differentiation process had been successful, flow cytometry was carried out using antibodies directed against human E-CADHERIN (Zymed, San Francisco, CA), TRA-1-60 (Millipore Corporation, MA), SSEA-4 (Millipore Corporation), and CD9 (BD Biosciences, CA) as previously described (Davis et al., 2008; Ng et al., 2007).

## 2.4.5. FPA FTIR imaging

FPA FTIR infrared images were acquired on a Agilent Technologies FTIR spectrometer (Model FTS 7000; Agilent Technologies Inc., Palo Alto, CA, USA) in transflectance mode, coupled to an infrared microscope (model 600 UMA; Agilent Technologies) using a $15\times$ objective and equipped with a $64 \times 64$ pixel MCT liquid nitrogen cooled FPA detector located at the Centre for Biospectroscopy at Monash University in Melbourne, Australia. FTIR spectra were acquired with 256 co-added background scans, at 8 cm$^{-1}$spectral resolution, with signals from groups of 4 adjacent pixels binned and co-added for 128 scans. The binning of 4 adjacent pixels on the FPA resulted in each spectrum corresponding to an area of 11µm $\times$ 11µm on the sample plane, which was the approximate dimensions of single cells in the dried monolayer. A Happ Genzel apodisation function was used to truncate the interferogram, and a zero-filling factor of 2 was employed.

## 2.4.6. Data preprocessing and multivariate data analysis

The infrared images were taken into Cytospec 1.03 IR imaging software (Cytospec Inc., NY, and USA) in order to extract the spectral data deemed suitable for preprocessing and subsequent analysis. Various analysis techniques are employed in the analysis of FTIR biospectroscopy spectra and the rationales for all these methods are explained extensively in Trevisan et al.'s review (Trevisan et al., 2012). However, the analysis methods which were used by our group are described as follows. First, to ensure the measurements had an appropriate signal to noise ratio and were in the linear range of the detector's response, the data first underwent a quality test whereby spectra with absorbance values outside the range of 0.2-0.8 (arising from sample regions devoid of cells or where cells were clumped and overlaid) were rejected. The spectra that passed the quality test underwent further preprocessing in The Unscrambler v 10.1 (Camo, Oslo, Norway) software package. Here, spectra were converted to their second derivatives using a Savitsky-Golay algorithm using 9 smoothing points in order to minimize baseline effects and to resolve spectral components that would otherwise be overlapping in the underivatised spectra. Normalisation was performed using Extended Multiplicative Signal Correction (Martens et al., 2003) using the following spectral ranges: 3050-2800 cm$^{-1}$, 1770-1730 cm$^{-1}$, 1470-900 cm$^{-1}$, as these ranges contain biological bands and therefore provided the best possibility of discrimination between different sample types. The amide I and amide II region were omitted from the analysis as this region is greatly affected by resonant Mie scattering, which may have confounded our interpretation of the spectra (Mohlenhoff et al., 2005). The data set was then randomly sorted into 2 new data sets comprising 2 thirds and 1 thirds of the spectra. These new sets were then used as the calibration and datasets respectively for Principal Component Analysis (PCA) and Soft Independent Modeling of Class Analogy (SIMCA) (Wold et al., 1987; Wold and Sjostrom, 1977) and Partial Least Squares

Discriminant Analysis (PLS-DA) (Geladi, 1988). To enable comparison between spectral averages and their standard deviation, mean and standard deviation spectra for each hESC and hiPSC line were calculated using the $2^{nd}$ derivative spectra that passed the quality test.

## 2.4.7. Principal Component Analysis, Partial Least Squares Discriminant Analysis, Soft Independent Modeling of Class Analogy

PCA was carried out on the calibration set using 5 PCs, and the score plots were used to visualize any clustering of the data, while the loading plots were used to determine which spectral regions contributed most to the variance in the data set, accounting for the observed clustering. SIMCA was then applied to the PCA models from the calibration set using an alpha level of 0.05 to determine the interclass distances. A SIMCA model distance of greater than 3 is generally indicative that the two clusters are classifiable and hence significantly different (Esbensen et al., 2002).

The data set was then randomly sorted into 2 new data sets comprising 2 thirds and 1 thirds of the spectra and used as the calibration and validation datasets respectively for classification of spectra using Partial Least Squares Discriminant Analysis (PLS-DA) (Geladi, 1988). The calibration data matrix employed for PLS-DA consisted of the spectral dataset (multivariate X) and two Y variables with integer values of 0 or 1 coding for the each of the two modelled spectral classes. Classification of the dataset was then carried out by predicting a Y value for each spectrum in the independent validation using PLSR models that had been generated from the calibration sets. Correct classification of each class was arbitrarily assigned to samples with predicted $Y>0.5$ for respective spectra. PLS loading plots were used to determine which spectral regions contributed most to the ability to classify spectra using PLS-DA.

## 2.5. Results and Discussion

### 2.5.1. Some independent human embryonic stem cell lines exhibit unique FTIR spectra

Previous work suggests that different origins and culture conditions can lead to different hESC lines having distinct gene expression signatures (Abeyta et al., 2004). Accordingly, our initial experiments involved determining the spectroscopic similarities between five independently derived hESC lines grown under standardized conditions.

Inspection of the average second derivative spectra and their standard deviations revealed line-specific FTIR absorbance differences in the spectral regions attributed to lipids, protein, amino acids, carbohydrates and nucleic acids (Figure 2.2C). PCA revealed limited clustering of spectra from the same cell lines in the scores plot along PC1- the principal component indicating the greatest source of variation in the dataset (Figure 2.2A). For example, some separation was observed between clusters of spectra from the H9, ENVY and *MIXL1*[GFP/w] groups, whereas considerable overlapping was evident between the other sample classes. Since the spectra had been converted to their 2nd derivatives, the negatively scored spectra corresponded to positive PC loadings, and positively scored spectra corresponded to negative PC loadings. That is, positively scored spectra would have higher absorbances for negatively loaded bands and vice versa (Figure 2.2B). The PC1 loadings bands that explained the variation in the dataset, were the bands at 2920 cm$^{-1}$ and 2850 cm$^{-1}$ and 1731 cm$^{-1}$, assigned to lipid groups, 2920 cm$^{-1}$ and 1450 cm$^{-1}$ assigned to proteins and lipids, at 1404 cm$^{-1}$ assigned to free amino acids, 1242 cm$^{-1}$, 1080cm$^{-1}$ and 995 cm$^{-1}$, from nucleic acids, and 1153 cm$^{-1}$, 1080 cm$^{-1}$ and 1022 cm$^{-1}$, assigned to C-O vibrations from glycogen and other carbohydrates (See Table 2.1 for details of the band assignments).

**Figure 2.2. Some, but not all hESCs and hiPSC lines can be readily discriminated from each other due to their unique FTIR spectroscopic profiles.** PCA scores plot (A) loadings plot (B) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (C) for spectra from 5 different hESC lines acquired using FPA-FTIR microspectroscopy. PCA scores plot (D) loadings plot (E) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (F) for spectra from 3 different hiPSC lines acquired using FPA-FTIR microspectroscopy.

Further analysis of the dataset employing SIMCA, revealed significant differences between some of the cell lines (Supplementary Figure 2.1A). The model distance plot showed that in relation to the H9 line, HES3, NKX2.5$^{GFP/w}$ and ENVY all had SIMCA model distances of greater than 3, indicating that these groups were more likely to be spectroscopically distinct from each other. In contrast, spectra from the *MIXL1*$^{GFP/w}$ line had a SIMCA model distance of less than 3, in relation to spectra from the H9 group, which suggested that they were very similar. PLS-DA modeling was also performed for the calibration and dataset shown in this experiment (Supplementary Figure 2.1D). The validation set was used to predict Y using the PLS-DA model and sensitivity and specificities for each of the lines was as follows: 83% and 100% for the *MIXL1*$^{GFP/w}$ line, 85% and 100% for NKX 2.5$^{GFP/w}$, 93% and 100% for the ENVY line, 55% and 96% for the H9 group and 48% and 100% for the HES3 line. The poorer classification outcome of some of these lines was not surprising given the observed overlapping of spectra from different classes in the scores plot, and the small SIMCA model distances as described above.

Considering that the NKX2.5$^{GFP/w}$, ENVY and *MIXL1*$^{GFP/w}$ hESC lines are sub-clones of HES3, it was surprising that some of them were spectroscopically distinguishable from each other. This result might suggest that, in addition to underlying genetics, the history of each line may impact on its overall biochemical composition. This study also highlights the power of FTIR microspectroscopy to provide invaluable information about a sample that may have otherwise gone undetected by conventional laboratory methods used in stem cell biology.

## 2.5.2. Different human induced pluripotent stem cell lines possessed unique FTIR 'spectral signatures'

Using the same rationale for performing the previous comparison, a similar multivariate data analysis was carried out using FTIR spectra collected from 3 different human iPSC

lines. Upon inspection of the PCA plots, the spectra were found to cluster into distinct groups according to cell type, and these groups appeared to be more disparate than was observed for the hESC lines (Figure 2.2D). Unlike the prior result, where there was considerable overlap between the lines, the PCA scores plot from this data showed distinctly negatively and positively scored clusters displaying a clear separation along PC1. The negatively scored spectra for the ES4CL1 line clustered away from the other two co-localising hiPSC lines, the IMR90C2 and AUS lines. In agreement with their second derivative spectra (Figure 2.2F), loadings plots indicated that the latter two cell lines had higher IR absorbances at 2920 cm$^{-1}$, 2850 cm$^{-1}$, and 1731 cm$^{-1}$, the spectral region that corresponded to lipids, compared to the ES4CL1 line. However, they had lower IR absorbances for bands ascribed to proteins and lipids ~1454 cm$^{-1}$, free amino acids ~1396 cm$^{-1}$, phosphorylated molecules (1080 cm$^{-1}$ and 1238 cm$^{-1}$), glycogen and other carbohydrates (1080 cm$^{-1}$, 1153 cm$^{-1}$ and 1018 cm$^{-1}$) (Figure 2.2E and Table 2.1).

Subsequent SIMCA modeling further verified that the groups were divergent in terms of macromolecular composition, as both the IMR90C2 and AUS cell lines had SIMCA interclass distances of greater than 3 in relation to the ES4CL1 group (Supplementary Figure 2.1B). The spectral data sets of these lines were found to classify very well with PLS-DA modeling, which was successful at discriminating between the three lines, achieving sensitivities and specificities of 100% for all three groups (Supplementary Figure 2.1E). We hypothesised that the spectroscopic disparities between the different cell lines were probably due to subtle differences between them resulting from the methods used to derive them. The protocols that are typically used to generate human induced pluripotent stem cells - viral transduction, DNA- based induction, mRNA transfection, and recombinant proteins- can be applied to a variety of somatic cell types. All of the hiPSC lines used in this study were 'reprogrammed' using viral vectors, but each had different

somatic cell origins. The ES4CL1 line was derived from foreskin, the IMR90C2 line, from lung fibroblast, and the AUS line from skin fibroblast. It has been reported that many human iPSC lines may retain an 'epigenetic memory' of the somatic cells from which they were created (Ghosh et al., 2010), and it is therefore plausible that the differences we observe between these cells lines could be related to this effect.

In the only other study that has investigated FTIR spectral 'signatures' of iPS cells. Sandt and colleagues found somatic amniotic fluid cells (AFC) derived iPSC lines to be spectrally indistinguishable from human ESC lines (Sandt et al., 2012). Conversely, small chemical differences were found to exist between hiPSC lines generated from H9 derived mesenchymal stem cells and hESC lines, indicated by their ability to be classified with PLS-DA.

## 2.5.3. Line dependent differences in the spectroscopic profiles of human embryonic stem cells and human induced pluripotent stem cells

Given the macromolecular variability that was seen amongst lines of the same stem cell type, we subsequently compared spectra from both stem cell types (hESC and hiPSC) together, to see if those differences were still detectable when compared to each other. From the average second derivative spectra, differences could certainly be seen across the spectral regions ascribed to lipids, protein, amino acids, carbohydrates, and nucleic acids but subsequent multivariate analyses did not reveal any consistent clustering based on whether the cell lines were hESC or hiPSC (data not shown).

Previous non-spectroscopic studies have identified a number of differences between the hiPSCs and hESCs, including their epigenetic state and transcriptional signature (Bock et al., 2011; Ghosh et al., 2010; Lister et al., 2011). How these molecular differences impact on the overall phenotype of the cells remains contentious. A recent investigation by Bock et al., for instance, found the effect of epigenetic memory on the DNA methylation and

gene expression of 12 fibroblast derived iPSC lines to be statistically insignificant (Bock et al., 2011). The group also found that most, but not all of the iPS cell lines could be readily distinguished from their ESC counterparts based on their DNA methylation and/or gene expression profiles. However, an epigenetic or transcriptional variation profile that was common to all the investigated iPS cell lines could not be detected, in line with the results presented here.

### 2.5.4. Undifferentiated human embryonic stem cells and human induced pluripotent stem cells can be discriminated from their differentiated progeny by their unique FTIR spectroscopic 'signatures'

We then wished to determine whether hESC and hiPSC undergoing differentiation towards mesendodermal or ectodermal lineages underwent similar macromolecular chemical changes. Firstly, in order to verify that the differentiation process had been successful in addition to enabling a correlation between cell differentiation and spectral signature, aliquots of the ectodermally or mesendodermally differentiated cultures that were interrogated by FTIR were analysed by flow cytometry. This analysis verified the assigned differentiation status of the samples and enabled a comparison of the differentiation obtained using the three pluripotent cell lines. As shown in Figure 2.3A, *MIXL1*$^{\text{GFP/w}}$ cells differentiated towards ectoderm using FGF2 (Heraud et al., 2010), did not express GFP from the MIXL1 locus, consistent with the lack of mesendoderm differentiation under these conditions. Similarly, these cells did not express PDGFRα or CXCR4, other markers associated with early mesoderm and endoderm, but retained expression of the epithelial marker EpCAM. Conversely, when these cells were differentiated in response to a combination of BMP4 and ACTIVIN A for four days, ~97% of the cells expressed MIXL1 in a graded fashion, consistent with mesendodermal commitment (Davis et al., 2008). Most

**Figure 2.3. Flow cytometry profiles of ectodermally and mesendodermally differentiated human pluripotent cell lines subsequently analysed by vibrational spectroscopy.** (A) *MIXL1*<sup>GFP/w</sup> cells differentiated for four days as spin EBs in APEL medium supplemented with FGF2 (Ectoderm differentiation) or with a combination of BMP4 and ACTIVIN A (Mesendoderm differentiation), were disaggregated and single cell suspensions analysed by flow cytometry for the expression of GFP from the MIXL1 locus (MIXL1-GFP) or for the expression of the indicated cell surface markers. (B, C) The human ESC lines *MIXL1*<sup>GFP/w</sup>, MEL1 or the human iPSC line IMR90C2 were differentiated as above towards (B) ectoderm of (C) mesendoderm and analysed by flow cytometry for the expression of the indicated cell surface markers. The quadrants were set using appropriate isotype control antibodies (data not shown).

cells also expressed PDGFRα and many were CXCR4 positive, but very few cells expressed EpCAM in the absence of MIXL1 expression. We have previously shown that during early mesendodermal differentiation, MIXL1 is co-expressed with epithelial markers such as E-CADHERIN or EpCAM (Davis et al., 2008; Heraud et al., 2010; Yu et al., 2012). Analyses examining the co-expression of EpCAM or PDGFRα with CXCR4 in the hESC lines *MIXL1*^GFP/w and MEL1 and the hiPSC line IMR90C2, demonstrated similar patterns of marker expression under ectodermal conditions, with retention of EpCAM and failure to induce significant proportions of the cells to express PDGFRα or CXCR4 (Figure 2.3B). In response to mesendoderm differentiation, very few cells expressed EpCAM alone (2-8%), although the proportion of PDGFRα and CXCR4 expressing cells varied a small amount between the three lines examined (Figure 2.4C). These data suggested that the concentrations of BMP4 and ACTIVIN A used for these experiments imparted a more mesendodermal bias to the *MIXL1*^GFP/w cells (leading to more cells expressing PDGFRα and fewer expressing CXCR4) and a more endodermal bias to the IMR90C2 line, in which a higher percentage of cells expressed CXCR4.

Spectral differences that were observed between undifferentiated hESC and hiPSC lines and their progeny resembled those previously described by our group (Heraud et al., 2010). In the present study, the chemical differences were characterised by higher signal intensities for lipid components at ~2920 cm$^{-1}$, ~2850 cm$^{-1}$ and ~1739 cm$^{-1}$ in the average second derivative spectra of all the undifferentiated cells across all three cell lines (Figures 2.4C, 2.4F and 2.4I). In contrast, higher FTIR absorbances were consistently seen in the spectral regions corresponding to glycogen and other carbohydrates at ~1153 cm$^{-1}$ and ~1022 cm$^{-1}$ in the spectra of the BMP4/ACTIVIN A differentiated progeny. Observations from the second derivative spectra were in agreement with the PCA results, which showed that in the PCA scores plots along PC1, spectra derived from undifferentiated cells

**Figure 2.4. hESC and hiPSC lines undergo the same macromolecular chemical changes upon differentiation towards mesendodermal and ectodermal lineages.**. PCA scores plot (A) loadings plot (B) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (C) for spectra from the hESC line *MIXL1*$^{\text{GFP/w}}$ differentiated towards mesendoderm using BMP4/ACTIVIN A or towards ectoderm with FGF2 acquired using FPA-FTIR microspectroscopy. PCA scores plot (D) loadings plot (E) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (F) for spectra from the hESC line MEL1 differentiated towards mesendoderm using BMP4/ACTIVIN A or towards ectoderm with FGF2 acquired using FPA-FTIR microspectroscopy. PCA scores plot (G) loadings plot (H) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (I) for spectra from the hiPSC line IMR90C2 differentiated towards mesendoderm using BMP4/ACTIVIN A or towards ectoderm with FGF2 acquired using FPA-FTIR microspectroscopy.

clustered separately from spectra acquired from their differentiated progeny (Figures 2.4A, 2.4D and 2.4G). In data derived from both stem cell types, the PC1 loadings plots (Figures 2.4B, 2.4H.and Table 2.2) were heavily loaded by bands attributable to lipids from C-H stretching vibrations from methylene groups at around 2920 cm$^{-1}$, 2850 cm$^{-1}$, and the ester carbonyl stretching band at 1739 cm$^{-1}$. Lipid absorbance was found to be higher in the undifferentiated cell lines, than in mesendoderm committed cells. Spectral clusters derived from BMP4/ACTIVIN A differentiated cells were found to be heavily loaded by bands ascribed to glycogen and other carbohydrates (~1149 cm$^{-1}$, ~1080 cm$^{-1}$ and 1018 cm$^{-1}$).

Overall, the SIMCA model distances were greater than 3 between the undifferentiated cells and the differentiated progeny, suggesting they should be highly classifiable (Supplementary Figures 2.2A, 2.2B and 2.2C). Indeed, for the *MIXL1*$^{GFP/w}$ differentiation experiment, PLS-DA classification could be applied with the following sensitivities and specificities: 100% and 100% for the BMP4/ACTIVIN A treated samples, 97% and 97% for the FGF2 treated samples, and 80% and 100% for the undifferentiated line (Supplementary Figure 2.2D). Classification of independent test spectra from the MEL1 differentiation experiment had sensitivities and specificities of 76% and 90% for the BMP4/ACTIVIN A treated group, 85% and 88% for the FGF2 treated group, and 100% and 96% for the undifferentiated MEL1 line (Supplementary Figure 2.2E). The validation set from the differentiation of the hiPSC line, IMR90C2 was found to classify at 100%, 93% and 80% for the BMP4/ACTIVIN A, FGF2, and undifferentiated groups respectively and their specificities that were obtained were 100%, 97% and 100% (Supplementary Figure 2.2F).

The finding that lipid stores diminished during the loss of pluripotency by hiPSCs corroborates hESC differentiation data previously published by our laboratory and other groups (Heraud et al., 2010; Pijanka et al., 2010). The reasons why lipid depletion occurs

during the differentiation process is unclear but one group has suggested that the eicosanoid pathway may play an important role (Yanes et al., 2010). They discovered that the inhibition of the eicosanoid signaling pathway of murine embryonic stem cells fostered pluripotency when the cells underwent neural differentiation. However, other groups have also observed either higher lipid levels in more differentiated stem cells (Tanthanuch et al., 2010) or non-significant lipid differences (Ami et al., 2008) between their undifferentiated and their progeny which suggests that this may be a cell-type specific phenomenon.

## 2.5.5. Differentiated progeny of human ESCs and iPSCs are spectroscopically distinguishable

Considering that FTIR microspectroscopy indicated that similar macromolecular changes were occurring during the differentiation of both the hESC and hiPSC lines, this prompted us to investigate whether these changes yielded progeny that were biochemically identical. PCA was performed using the spectra of BMP4/ACTIVIN A treated cells derived from the *MIXL1*[GFP/w] hESC line, MEL1 hESC line, and IMR90C2 iPSC line (Figure 2.5). Along PC1 of the PCA scores plot, a separation could be seen between clusters of spectra derived from the *MIXL1*[GFP/w] and IMR90C2 lines away from the MEL1 spectral clusters. In accordance with the average second derivative spectra, high FTIR absorbances were consistently seen in the spectral regions corresponding to lipids (at ~2916 $cm^{-1}$, 2842 $cm^{-1}$ and 1731 $cm^{-1}$), carbohydrates (at ~1153 $cm^{-1}$ and 1022 $cm^{-1}$) in the spectra of the BMP4/ACTIVIN A differentiated progeny derived from the *MIXL1*[GFP/w] and IMR90C2 lines. Whereas spectral clusters derived from MEL1 progeny were found to have high FTIR intensities from bands ascribed to proteins and lipids (1446 $cm^{-1}$), and free amino acids (~1392 $cm^{-1}$).

The SIMCA model distances were consistently greater than 3 when the MEL1 line was taken as the model centre, whereas the other two lines appeared to be more similar to each

**Figure 2.5. hESC and hiPSC differentiated towards mesendodermal and ectodermal lineages do not yield biochemically equivalent progeny.** PCA scores plot (A) loadings plot (B) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (C) for spectra from the hESC lines $MIXL1^{GFP/w}$, MEL1 and the hiPSC line IMR90C2 treated with BMP4/ACTIVIN A acquired using FPA-FTIR microspectroscopy. PCA scores plot (D) loadings plot (E) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (F) for spectra from the hESC lines $MIXL1^{GFP/w}$ and MEL1 and the hiPSC line IMR90C2 treated with FGF2 acquired using FPA-FTIR microspectroscopy.

other (Supplementary Figure 2.3). Further, PLS-DA classification could be applied with the following sensitivities and specificities, respectively: 53% and 100% for the $MIXL1^{GFP/w}$ line, 100% and 85% for the MEL1 line, and 100% and 82% for the IMR90C2 line.

Multivariate analysis of spectra derived from the FGF2 differentiated progeny of the same three lines showed the $MIXL1^{GFP/w}$ spectral cluster separating away from spectra of the two other cell lines along PC1 of the PCA scores plot (Figure 2.5D). The loadings bands explaining the variance contributing to the clustering pattern were the bands that corresponded to lipid groups (~2916 cm$^{-1}$, 2842 cm$^{-1}$ and 1731 cm$^{-1}$), in addition to those ascribed to glycogen and other carbohydrates (1149 cm$^{-1}$, 1080 cm$^{-1}$ and 1018 cm$^{-1}$). FTIR intensities from these aforementioned bands were higher in the $MIXL1^{GFP/w}$ spectral cluster compared to spectral clusters of the other two lines, which exhibited higher IR intensities for the spectral regions which corresponded to protein (~1446 cm$^{-1}$), free amino acids (1392 cm$^{-1}$), and phosphorylated molecules (~1242 cm$^{-1}$). SIMCA classification results showed the $MIXL1^{GFP/w}$ line to be highly classifiable compared to the other two lines, consistently having a model distance of greater than 3 when taken as the model centre. The spectral data sets of these lines were found to be classifiable with PLS-DA modeling, which was successful at discriminating between the three lines, achieving sensitivities and specificities of 100% and 77% for the MEL1 line, 86% and 96% for the IMR90C2 line, and 54% and 100% for the $MIXL1^{GFP/w}$ line.

This result is noteworthy because it shows that despite both stem cell variants undergoing the same macromolecular changes during the 4 days of differentiation, the resulting cells were spectroscopically distinguishable from each other. It is feasible that the aforementioned disparities between the two cohorts may reflect different responses to cytokine cues resulting in lineage committed progeny that appear phenotypically

equivalent in terms of the cell surface marker expression (Figure 2.3), but are still distinct in terms of their macromolecular composition.

This difference in differentiation potential between hESCs and iPSCs has also been observed by others using non-spectroscopic analyses. In one recent study, Bock and colleagues compared the capacity of 12 human iPSC lines generated via lentiviral, retroviral, or non-integrating episomal methods to form neural lineages with 5 hESC lines (Bock et al., 2011). The study found that the iPSC lines were able to differentiate into neuroepithelial (NE) cells and functional neurons or glia with the same kinetics as that observed for the hESC lines. Nevertheless, the differentiation outcome was found to proceed with reduced efficiency and higher variability when compared to human embryonic stem cells. Similarly, Narsinh's group described, poorer in vitro differentiation outcomes with hiPSC lines derived from lentiviral transgenesis and non-viral minicircle based reprogramming differentiating into beating cardiomyocytes and ECs, as a consequence of higher variability and lower expansion capabilities (Davis et al., 2008).

Certainly, to achieve greater success in hiPSC differentiation experiments, there needs to be an appropriate screening method, capable of selecting the iPSC lines with the greatest propensity for differentiating towards specific lineages. One such quality control technique was described in Bock et al.'s study whereby a 'lineage scorecard' assay, consisting of the profiles of 500 relevant genes was created to predict the hiPSC lines that would most effectively differentiate into motor neurons (Bock et al., 2011).

Nonetheless, the last finding in our study suggests that it may be insufficient to use parameters such as genetics and differentiation ability, as the sole criteria for determining the hiPSC lines that most closely resemble hESCs. Our data raise a critical point that despite hiPSC lines exhibiting similar differentiation kinetics to their hESC counterparts they may still be divergent, in terms of their macromolecular chemistry. At present the

exact implications of this finding are unclear, but one should remain mindful of these differences and their potential to affect experimental outcomes when designing future studies involving hESC and hiPSC lines.

Perhaps an alternative solution to Bock et al.'s strategy could be an integrative testing method, encompassing features of the lineage scorecard assays with other robust phenotypic-based modalities, including FTIR microspectroscopy. Not only is this chemical tool able to provide invaluable, complementary information alongside conventional molecular biology methods, but it is also able to do so rapidly, and non-invasively. In addition, its capacity for generating large quantities of data (up to 4096 spectra with the use of a 64 x 64 focal plane detector in a single experiment) would be advantageous in a clinical environment where rapid yet detailed assessment of a cells phenotype was required.

A recent exciting development in the field has been the utilisation of quantum cascade laser sources of mid infrared light (Haibach et al., 2011). These lasers provide sources of IR light more brilliant than synchrotron sources, allowing FTIR measurements with high signal to noise ratios to be achieved in the laboratory, even with single living cells. For future experiments we hope to use either laser based, or synchrotron IR systems to probe the macromolecular chemical signatures of these clinically anticipated cell lines *in vivo*. This would be an invaluable exercise as it would enable the investigation of IR spectral differences at the single cell level, and to more closely study the nucleic acid region of the IR spectrum, which can be ambiguous in dehydrated samples (Whelan et al., 2011).

Advances in instrumentation and analysis techniques will allow for the feasible implementation of FTIR microspectroscopy as an objective, label-free, non-destructive technique for the screening of clinically destined stem cells and their derivatives. This modality would not only be advantageous in the interrogation of pluripotent and/or

multipotent stem cells to be used for regenerative therapies, but also holds great promise for the detection of cancer stem cells (Hughes et al., 2010; Kelly et al., 2010).

## 2.6. Conclusions

The work presented here shows that some, but not all, human embryonic and human induced pluripotent stem cells have significant spectroscopic and hence, macromolecular differences from each other. Whilst both stem cell variants undergo similar macromolecular changes during differentiation, with the main change involving lipid depletion, they generate spectroscopically distinct, lineage committed cells. We have demonstrated the robust abilities of FTIR microspectroscopy for providing further insight into the contentious hESC versus hiPSC equivalency debate.

## 2.7. Acknowledgements

| Band maxima 2$^{nd}$derivative spectra (cm$^{-1}$) | PC1 loadings (cm$^{-1}$) | PC2 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|
| hESC, hiPSC | hESC, hiPSC | hESC, hiPSC | |
| ~2962 | ~2962 | ~2962 | C-H stretching peak of lipids and proteins |
| ~2923 | ~2920 | ~2920 | C-H asymmetrical stretching of lipid groups and protein |
| ~2850 | ~2850 | ~2850 | C-H symmetrical stretching of lipid groups |
| ~1743 | ~1731 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1454 | ~1446 | Methyl and methylene groups from lipids and protein |
| ~1396 | ~1396 | ~1392 | COO$^-$ stretching vibrations of amino acid side chains |
| ~1238 | ~1238 | ~1245 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| ~1080 | ~1080 | ~1091 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules, and glycogen |
| ~1153 | ~1153 | ~1164 | C-O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1018 | C-O vibrations from glycogen and other carbohydrates |
| ~995 | ~995 | ~991 | C-O stretch from RNA ribosechain and other carbohydrates |

**Table 2.1:** Band maxima of human embryonic and human induced pluripotent stem cells (replicate no. 1)

| Band maxima 2nd derivative spectra (cm⁻¹) | PC1 loadings (cm⁻¹) | PC2 loadings (cm⁻¹) | Band assignments |
|---|---|---|---|
| *MIXL1*GFP/w, MEL1, IMR90C2 +BMP4 ACTIVIN A/FGF2 | *MIXL1*GFP/w, MEL1, IMR90C2 +BMP4 ACTIVIN A/FGF2 | *MIXL1*GFP/w, MEL1, IMR90C2 +BMP4 ACTIVIN A/FGF2 | |
| ~2962 | ~2962 | ~2962 | C-H stretching peak of lipids and proteins |
| ~2923 | ~2920 | ~2923 | C-H asymmetrical stretching of lipid groups |
| ~2850 | ~2850 | ~2850 | C-H symmetrical stretching of lipid groups |
| ~1739 | ~1735 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1450 | ~1450 | Methyl and methylene groups from lipids and protein |
| ~1392 | ~1392 | ~1392 | $COO^-$ stretching vibrations of amino acid side chains |
| ~1238 | ~1245 | ~1249 | P=O asymmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| ~1080 | ~1080 | ~1091 | P=O symmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules and glycogen |
| ~1153 | ~1153 | ~1149 | C-O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1018 | C-O vibrations from glycogen and other carbohydrates |
| ~995 | ~991 | ~991 | C-O stretch from RNA ribosechain and other carbohydrates |

**Table 2.2:** Band maxima of human embryonic and human induced pluripotent stem cells differentiated with BMP4 ACTIVIN A and FGF2

| Band maxima 2$^{nd}$derivative spectra (cm$^{-1}$) | PC1 loadings (cm$^{-1}$) | PC2 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|
| *MIXL1*$^{GFP/w}$ v MEL1 v IMR90C2 +BMP4 ACTIVIN A, +FGF2 | *MIXL1*$^{GFP/w}$ v MEL1 v IMR90C2 +BMP4 ACTIVIN A, +FGF2 | *MIXL1*$^{GFP/w}$ v MEL1 v IMR90C2 +BMP4 ACTIVIN A, +FGF2 | |
| ~2962 | ~2962 | ~2962 | C-H stretching peak of lipids and proteins |
| ~2923 | ~2927 | ~2920 | C-H asymmetrical stretching of lipid groups |
| ~2850 | ~2854 | ~2850 | C-H symmetrical stretching of lipid groups |
| ~1739 | ~1739 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1446 | ~1446 | Methyl and methylene groups from lipids and protein |
| ~1392 | ~1392 | ~1392 | COO$^-$ stretching vibrations of amino acid side chains |
| ~1238 | ~1242 | ~1238 | P=O asymmetrical stretching of phosphodiesters from phosphorylated molecules |
| ~1080 | ~1091 | ~1080 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules and glycogen |
| ~1153 | ~1149 | ~1157 | C-O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1022 | C-O vibrations from glycogen and other carbohydrates |
| ~995 | ~991 | ~995 | C-O stretch from RNA ribosechain and other carbohydrates |

**Table 2.3:** Band maxima of comparisons between BMP4 ACTIVIN A and FGF2 treated progeny of hESC lines *MIXL1*$^{GFP/w}$ and MEL1 and hiPSC line IMR90C2

**Supplementary Figure 2.1. SIMCA and PLS-DA modelling indicate that some hES and hiPS lines are more spectroscopically distinct than others.** SIMCA model distance plots of spectra from (A) the five hESC lines and (B) three different hiPSC lines In relation to the H9 group, the HES3, NKX 2.5$^{GFP/w}$ and ENVY lines all had model distances of greater than 3, indicating that they were significantly different from each other. PLS-DA analysis of these groups showing the percentage correct classification of spectra drawn from an independent validation set (C), and (D).
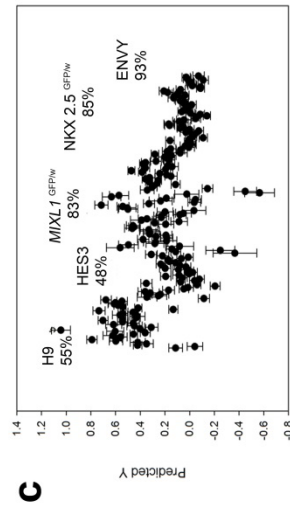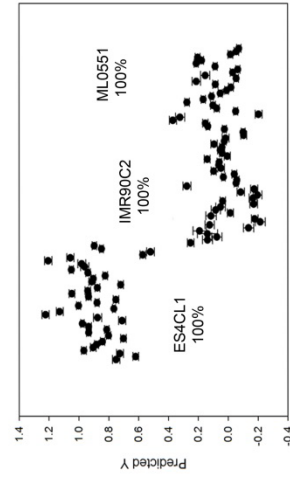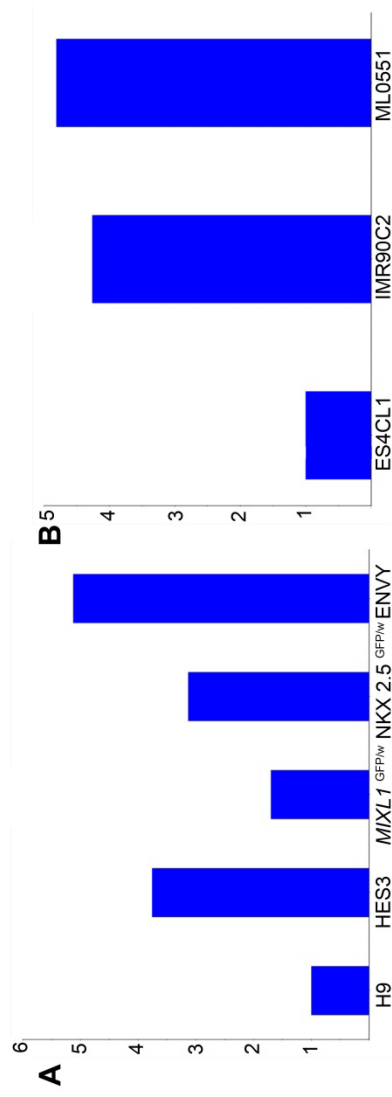
**Supplementary Figure 2.2. SIMCA and PLS-DA modelling indicate that cells differentiated under BMP4/ACTIVIN A are significantly different spectroscopically.** SIMCA model distance plots of spectra from (A) *MIXL1*[GFP/w,] (B) MEL1 and (C) IMR90C2 lines differentiated with BMPA/ACTIVIN A and FGF2 and their corresponding PLS-DA prediction plots shown in (D), (E) and (F), respectively.
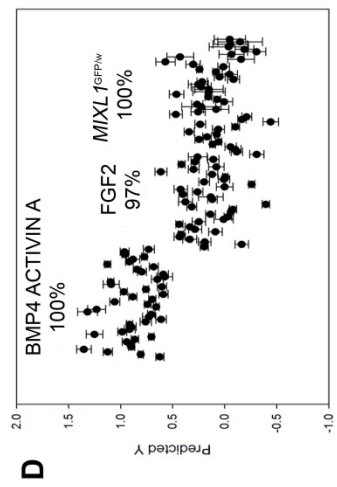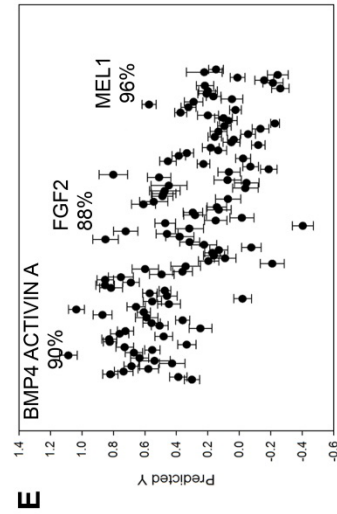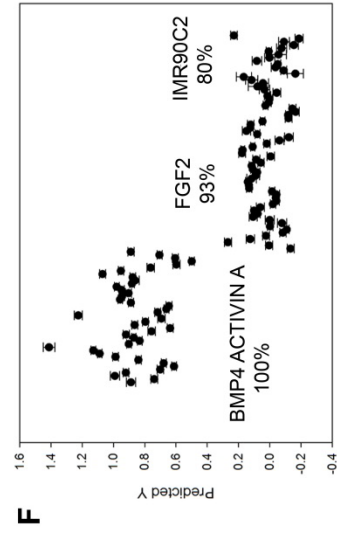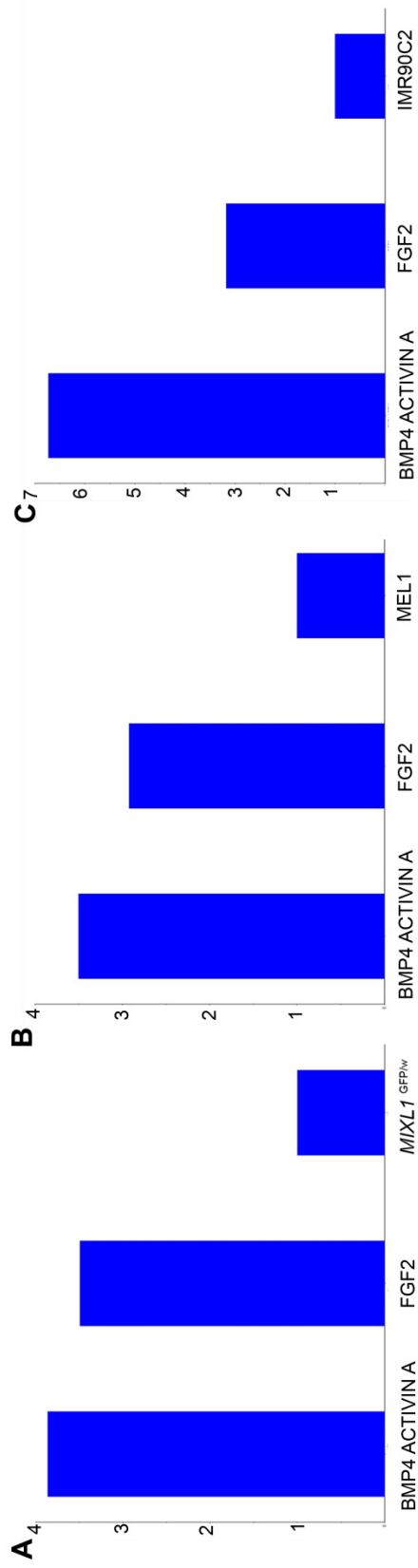
**Supplementary Figure 2.3.** SIMCA model distance plots of spectra from (A) BMP4/ACTIVIN A differentiated progeny of *MIXL1*<sup>GFP/w</sup>, MEL1 and IMR90C2 or (B) FGF2 differentiated progeny of *MIXL1*<sup>GFP/w</sup>, MEL1 and IMR90C2
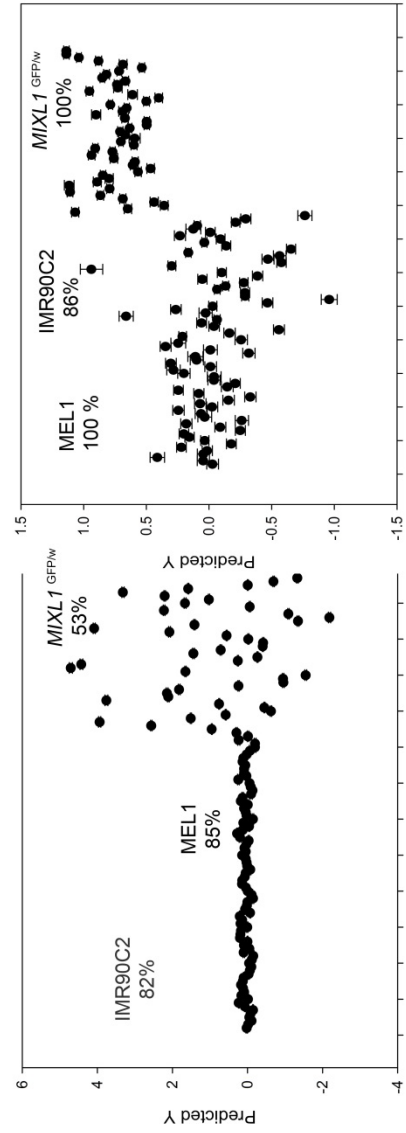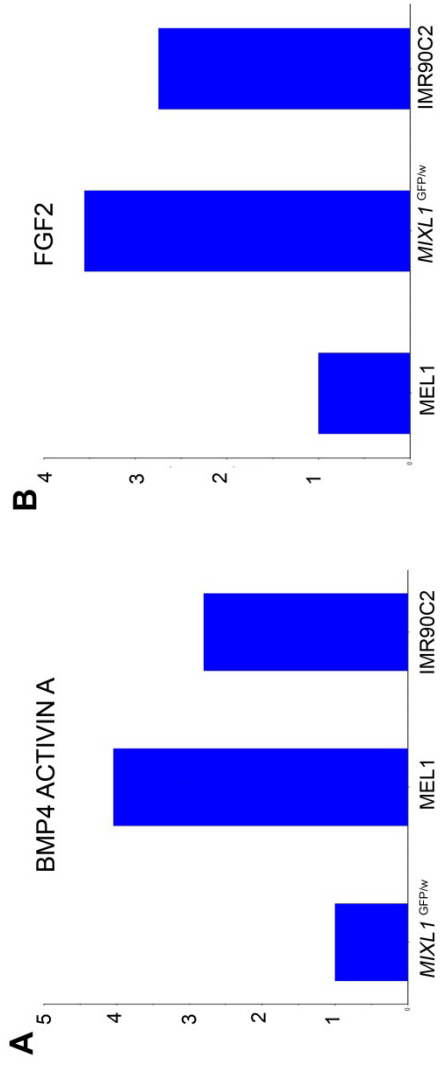
**Supplementary Figure 2.4.** PCA scores plot (A) loadings plot (B) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (C) for spectra from 5 different hESC lines acquired using FPA-FTIR microspectroscopy (replicate 2). PCA scores plot (D) loadings plot (E) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (F) for spectra from 3 different hiPSC lines acquired using FPA-FTIR microspectroscopy (replicate 2).
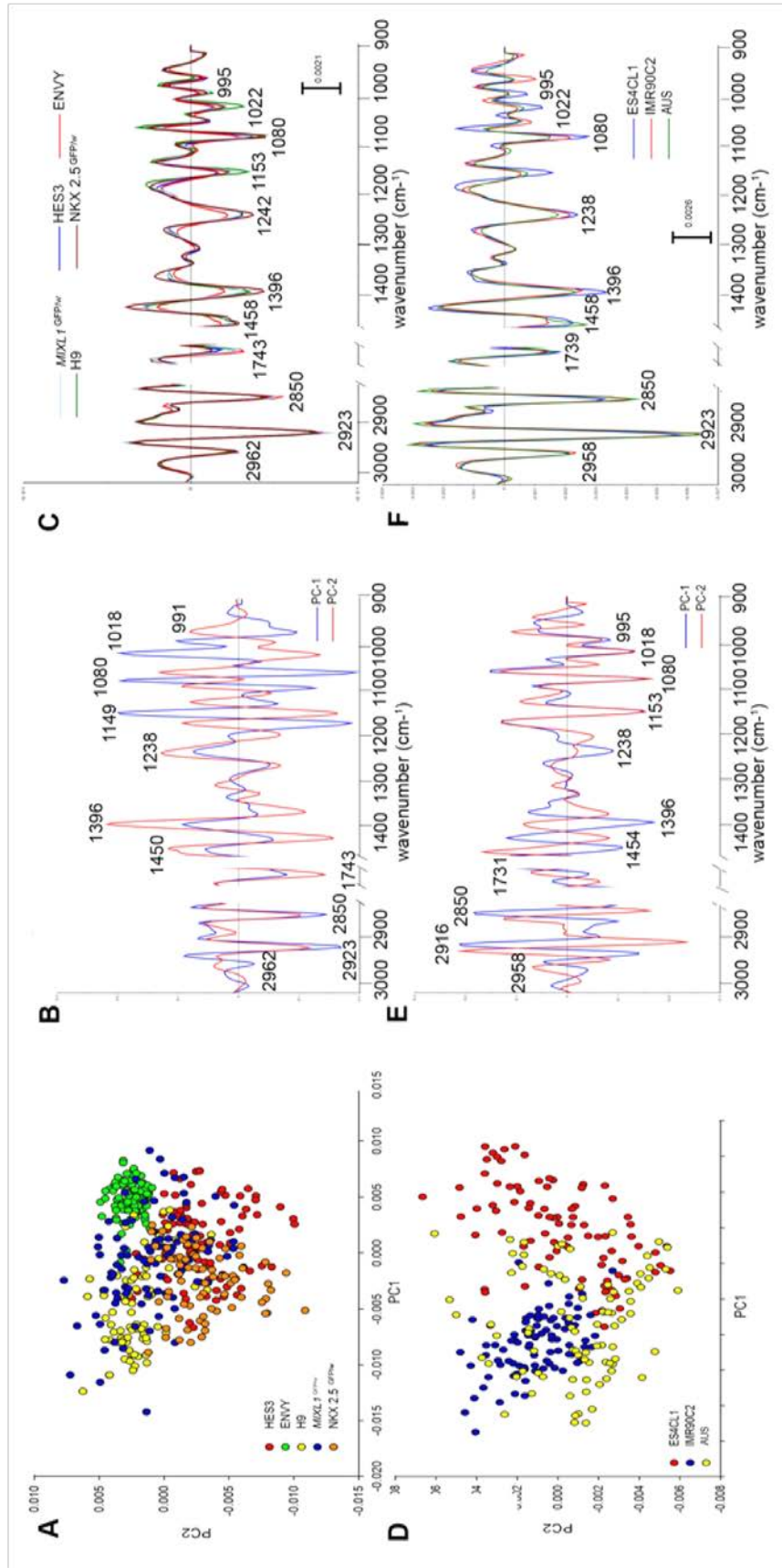
**Supplementary Figure 2.5.** SIMCA model distance plots of spectra from replicate no.2 of the same experiments performed and shown in Figure 2.3 and Supplementary Figure 2.1. SIMCA plots are shown of (A) five hESC lines and (B) three different hiPSC lines. In relation to the H9 group, the HES3, *NKX 2.5*<sup>GFP/w</sup> and ENVY lines all had model distances of greater than 3, indicating that they were significantly different from each other. PLS-DA analysis of these groups showing the percentage correct classification of spectra drawn from an independent validation set (C and D).

| Band maxima 2$^{nd}$derivative spectra (cm$^{-1}$) | PC1 loadings (cm$^{-1}$) | PC2 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|
| hESC, hiPSC | hESC, hiPSC | hESC, hiPSC | |
| ~2962 | ~2962 | ~2962 | C-H stretching peak of lipids and proteins |
| ~2923 | ~2923 | ~2923 | C-H asymmetrical stretching of lipid groups |
| ~2850 | ~2850 | ~2850 | C-H symmetrical stretching of lipid groups |
| ~1743 | ~1731 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1454 | ~1450 | Methyl and methylene groups from lipids and protein |
| ~1396 | ~1396 | ~1396 | COO$^-$ stretching vibrations of amino acid side chains |
| ~1242 | ~1238 | ~1238 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| ~1083 | ~1080 | ~1080 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules and glycogen |
| ~1153 | ~1153 | ~1153 | C-O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1022 | C-O vibrations from glycogen and other carbohydrates |
| ~991 | ~995 | ~995 | C-O stretch from RNA ribosechain and other carbohydrates |

**Supplementary Table 2.1:** Band maxima of human embryonic and human induced pluripotent stem cells (replicate no. 2)

| REPLICATE 1, AND REPLICATE TWO OF hESC LINES | | | | | |
|---|---|---|---|---|---|
| **REP 1, REP 2** | **ENVY** | **H9** | **HES3** | *MIXL1*GFP/w | **NKX 2.5**GFP/w |
| **ENVY** | 1.00,1.00 | 4.61,3.75 | 3.13,1.70 | 3.02,3.13 | 3.03,5.12 |
| **H9** | 4.61,3.75 | 1.00,1.00 | 3.07,2.37 | 2.47,4.35 | 2.91,3.61 |
| **HES3** | 3.13,1.70 | 3.07,2.37 | 1.00,1.00 | 3.11,1.59 | 3.70,1.90 |
| *MIXL1*GFP/w | 3.02,3.13 | 2.47,4.35 | 3.11,1.59 | 1.00,1.00 | 2.65,2.77 |
| **NKX 2.5**GFP/w | 3.03,5.12 | 2.91,3.61 | 3.70,1.90 | 2.65,2.77 | 1.00,1.00 |
| REPLICATE 1, AND REPLICATE 2 OF hiPSC LINES | | | | | |
| **REP 1, REP 2** | **ES4CL1** | | **IMR90C2** | **AUS** | |
| **ES4CL1** | 1.00,1.00 | | 4.26,3.88 | 4.80,3.48 | |
| **IMR90C2** | 4.26,3.88 | | 1.00,1.00 | 4.86,4.07 | |
| **AUS** | 4.80,3.48 | | 4.86,4.07 | 1.00,1.00 | |
| *MIXL1*GFP/w, MEL1, AND IM90C2 DIFFERENTIATED WITH BMP4 ACTIVIN A AND FGF2 | | | | | |
| *MIXL1*GFP/w, MEL1, IM90C2 | **BMP4 ACTIVIN A** | | **FGF2** | **UNDIFFERENTIATED** | |
| **BMP4 ACTIVIN A** | 1.00,1.00,1.00 | | 1.90,2.70,5.20 | 3.90,3.50,6.70 | |
| **FGF2** | 1.90,2.70,5.20 | | 1.00,1.00,1.00 | 3.50,2.90,3.20 | |
| **UNDIFFERENTIATED** | 3.90,3.50,6.70 | | 3.50,2.90,3.20 | 1.00,1.00,1.00 | |
| BMP4 ACTIVIN A /FGF TREATED PROGENY OF *MIXL1*GFP/w v MEL1 v IMR90C2 | | | | | |
| **BMP4 ACTIVIN A/FGF2** | *MIXL1*GFP/w | | **MEL1** | **IMR90C2** | |
| *MIXL1*GFP/w | 1.00,1.00 | | 4.04,3.56 | 2.80,2.74 | |
| **MEL1** | 4.04,3.55 | | 1.00,1.00 | 3.71,4.15 | |
| **IMR90C2** | 2.80,2.74 | | 3.70,4.15 | 1.00,1.00 | |

**Supplementary Table 2.2:** SIMCA classification results showing the various model distances of the previously described analyses

CHAPTER **3**

Fourier transform infrared microspectroscopy reveals that tissue culture conditions affect the macromolecular phenotype of human embryonic stem cells

# Monash University

# Declaration for Thesis Chapter 3

## Declaration by candidate

In the case of Chapter 3, the nature and extent of my contribution to the work was the following:

| Nature of contribution | Extent of contribution (%) |
|---|---|
| The work described in the paper by Cao et al. was predominantly performed by myself. I was also involved in the designing and analysis of the experiments performed, and did the majority of the writing. | 80 |
| The remainder of Chapter 3 was written by myself | 100 |

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

| Name | Nature of contribution | Extent of contribution (%) for student co-authors only |
|---|---|---|
| Elizabeth Ng | Technical assistance, data interpretation, critical revision and final approval of the manuscript | 5% |
| Edouard Stanley | Data interpretation, critical revision and final approval of the manuscript | |
| Donald McNaughton | Supervision, data interpretation, critical revision and final approval of the manuscript | |
| Mark Tobin | Supervision, data interpretation, critical revision and final approval of the manuscript | |
| Andrew Elefanty | Supervision, data interpretation, critical revision and final approval of the manuscript | |
| Philip Heraud | Supervision, data interpretation, critical revision and final approval of the manuscript | |

| Candidate's Signature | <span style="background:black">          </span> | **Date** 04/06/13 |
|---|---|---|

## Declaration by co-authors

The undersigned hereby certify that:

(7) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(8) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(9) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

(10)   there are no other authors of the publication according to these criteria;

(11)   potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(12)   the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

| Location(s) | 1.Monash Immunology and Stem Cell Laboratories (MISCL) - now part of the Department of Anatomy and Developmental Biology, Monash University, 2.Centre for Biospectroscopy, Monash University |
|---|---|

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

| | Author name | Signature | Date |
|---|---|---|---|
| **Signature 1** | Elizabeth Ng | <span style="background:black">████</span> | 04/06/13 |
| **Signature 2** | Edouard Stanley | <span style="background:black">████</span> | 04/06/13 |
| **Signature 3** | Donald McNaughton | <span style="background:black">████</span> | 04/06/13 |
| **Signature 4** | Mark Tobin | <span style="background:black">████</span> | 04/06/13 |
| **Signature 5** | Andrew Elefanty | <span style="background:black">████</span> | 04/06/13 |
| **Signature 6** | Philip Heraud | <span style="background:black">████</span> | 04/06/13 |

## 3.1. Introduction

This Chapter consists of a manuscript which has been published in the *Analyst* and describes the use of this optical tool for investigating the effects of different tissue culturing methodologies on the unique molecular 'signatures' of human embryonic stem cells and their mesendodermal and ectodermal committed progeny.

# Fourier transform infrared microspectroscopy reveals that tissue culture conditions affect the macromolecular phenotype of human embryonic stem cells

*Julie Cao [a,b], Elizabeth S.Ng[a,d], Don McNaughton [b], Edouard G. Stanley[a,d], Andrew G. Elefanty[a,d], Mark J. Tobin[c], Philip Heraud[a,b]*

Monash Immunology and Stem Cell Laboratories [a], Monash University, Building 75, STRIP 1, West Ring Road, Clayton, Victoria 3800, Australia [a]

Centre for Biospectroscopy and the School of Chemistry [b], Monash University, Clayton, Victoria 3800, Australia [b]

Australian Synchrotron c, 800 Blackburn Road, Clayton, Victoria 3168 c

Murdoch Childrens Research Institute, The Royal Children's Hospital, Parkville, Victoria, Australia, 3052[d]

* Corresponding author: e-mail █████████████████

**Key words: Fourier transform infrared microspectroscopy, human embryonic stem cells, tissue culture conditions, regenerative medicine**

## 3.2. Abstract

We employed Fourier transform infrared (FTIR) microspectroscopy to investigate the effects of different tissue culture environments on the FTIR spectra of undifferentiated human embryonic stem cells (hESCs) and their differentiated progeny. First we tested whether there were any possible spectral artifacts resulting from the use of transflectance measurements by comparing them with transmission measurements and found no evidence of these concluding that the lack of any differences resulted from the homogeneity of the dried cytospun cellular monolayers. We found that hESCs that were enzymatically passaged onto mouse embryonic fibroblasts (MEFs) in KOSR based hESC medium, hESCs enzymatically passaged onto Matrigel in mTESR medium and hESCs mechanically passaged onto MEFs in KOSR-based hESC medium, possessed unique FTIR spectroscopic signatures that reflect differences in their macromolecular chemistry. Further, these spectroscopic differences persisted even upon differentiation towards mesendodermal lineages. Our results suggest that FTIR microspectroscopy is a powerful, objective, measurement modality that complements existing methods for studying the phenotype of hESCs and their progeny, particularly changes induced by the cellular environment.

## 3.3. Introduction

Owing to their unique ability to differentiate into the three embryonic germ layers (Thomson et al., 1998), human embryonic stem cells (hESCs) have been advanced as a potential source of cells for the replacement of disease-compromised cells and tissue. Since the initial derivation of hESCs from blastocysts (Thomson et al., 1998), cell culturing conditions have undergone multiple refinements. Today, hESC lines are either co-cultured with embryonic fibroblast feeder cells (feeder-dependent systems) (Amit et al., 2003; Hovatta et al., 2003; Richards et al., 2002; Thomson et al., 1998), or in the absence of feeders (feeder-free systems) (Braam et al., 2008; Rosler et al., 2004; Sjögren-Jansson et al., 2005). An example of the latter system involves culturing onto Matrigel (Becton Dickinson, Bedford, MA), a basement membrane preparation extracted from a murine Englebreth-Holm-Swarm sarcoma. Stem cells grown using either platform are maintained in a range of growth media formulations and routinely passaged via either mechanical dissociation (Carpenter et al., 2003; Mitalipova et al., 2005; Thomson et al., 1998), or disaggregation methods that include both non-enzymatic (cell dissociation buffer, CDB) and enzymatic treatments (collagenase/trypsin, CT).

The diversity of culture methodologies has contributed to a range of morphological (Sjögren-Jansson et al., 2005) and genotypic (Maitra et al., 2005) disparities which may affect the success of stem cell maintenance (Akopian et al., 2010) and differentiation outcomes. There is currently no data documenting how different methodologies affect the macromolecular phenotype of hESCs, as traditional stem cell biology techniques cannot provide this information. However, with the recent emergence of spectroscopic modalities such as Fourier transform infrared (FTIR) microspectroscopy in stem cell research (Heraud

and Tobin, 2009) such information can now be easily attained and modeled using a variety of chemometric-based approaches.

FTIR microspectroscopy has provided new insights into a variety of different stem cells, including hESC, adult (tissue derived) and induced pluripotent stem cells (iPSC) (Ami et al., 2008; German et al., 2006; Heraud et al., 2010; Pijanka et al., 2010; Walsh et al., 2008) and their lineage committed cells. Our laboratory has employed it to discriminate undifferentiated hESCs from their mesendodermal progeny (Heraud et al., 2010). However, no studies to our knowledge have applied it to investigate the influence of the culturing environment on the macromolecular chemistry of hESCs. Since hESCs can be both genotypically and phenotypically affected by their growth conditions, we hypothesised that hESCs and their progeny, maintained under different conditions, will possess unique FTIR "spectroscopic signatures.''

## 3.4. Materials and Methods

### 3.4.1. Atomic force microscopy methodology

It has been reported recently that transflectance measurements may be affected by the electric field standing wave (EFSW) (Bassan et al., 2013; Filik et al., 2012), which is claimed to give rise to distorted spectra and absorbance changes and thereby influencing the classification outcomes. The artifact has been reported to be most pronounced in transflectance measurements compared with transmission measurements. Since the phenomena is thought to be most problematic when there are differences in the thickness of compared samples, we measured the height of day 4 mesendodermal differentiated $MIXL1^{GFP/w}$ cells that were initially cultured on Matrigel in mTESR medium, and $MIXL1^{GFP/w}$ cells that were initially cultured on MEFs in KOSR-based hESC medium, on MirrIR slides using Atomic Force Microscopy (AFM). This comparison was chosen

because this constituted the most significant comparison in the study in terms of the findings, and the differences in thickness between cells derived from the two conditions would allow us to determine whether the spectroscopic differences between the cells in each condition were of a physical (attributable to the EFSW phenomenon) or biochemical origin.

AFM measurements of cell height were made in tapping mode using a JPK Nanowizard 3 AFM. This instrument is equipped with capacitative sensors to ensure accurate reporting of height, $z$, and $x$-$y$ lateral distances. Cantilevers used were Bruker NCHV model tapping mode levers, with nominal resonant frequencies of 340 kHz and spring constants of 20-80 N/m respectively. Imaging was performed with a set-point force of <1 nN. Sectional heights of cells were measured by taking line profiles (effectively a single $x$-$z$ fast-axis scan line from an image) that had clear areas of substrate on either side of a cell, and calculating the cell's highest point with respect to the substrate. Repeated imaging at different scan angles was performed to ensure that no sample damage or deformation occurred during measurement.

### 3.4.2. Preparation of stem cells for FTIR microspectroscopy

#### 3.4.2.1. Testing the effects of different tissue culture methodologies on the FTIR spectroscopic signatures of human embryonic stem cells

Four karyotypically normal hESC lines, HES3 (Richards et al., 2002) and H9 (Thomson et al., 1998), MISCES-01 (Tecirlioglu et al., 2010) and *MIXL1*[GFP/w] (Davis et al., 2008) were used for the various experiments in this study.

Three replicate cultures of the hESC lines HES3 (Richards et al., 2002) and H9 (Thomson et al., 1998), and MISCES-01 (Tecirlioglu et al., 2010) were either enzymatically passaged onto mouse embryonic fibroblasts (MEFs) using TrypLE Select (Invitrogen) in Knock Out Serum Replacer (KOSR)-based hESC medium (Costa et al., 2007), mechanically passaged

onto MEFs in KOSR medium or enzymatically passaged onto Matrigel (Becton Dickinson, Bedford, MA), in mTESR medium (Stem Cell Technologies, Vancouver, BC, Canada), using methods previously described (Costa, 2008).

### 3.4.2.2. Testing the persistence of culture condition induced spectroscopic effects upon differentiation towards mesendodermal lineages

Three replicate cultures of the hESC line $MIXL1^{GFP/w}$, a genetically modified hESC line expressing green fluorescent protein (GFP) from the $MIXL1$ locus (Davis et al., 2008), were differentiated toward mesendoderm or ectoderm as monolayers, in 6 well tissue culture plates. Differentiation towards mesendoderm was performed in serum-free APEL medium (Ng et al., 2008) supplemented with 20 ng/ml of BMP4, 50 ng/ml of Activin A, 10 ng/ml of FGF2 and 10 ng/ml of VEGF. After 3 and 4 days of differentiation, cells were analysed for induction of GFP from the MIXL1 locus (denoted MIXL1-GFP) by flow cytometry (Davis et al., 2008; Ng et al., 2007) to confirm that differentiation had occurred, before being prepared as previously described for FTIR microspectroscopy (Heraud et al., 2010).

### 3.4.2.3. Testing the effects of changing culture conditions on the FTIR spectroscopic signatures of hES cells

For the adaptation experiments, $MIXL1^{GFP/w}$ cells were enzymatically passaged three times onto Matrigel in mTESR medium before being enzymatically passaged onto MEFs in KOSR medium. hESCs were maintained for 5 passages under the same conditions, and during each passage, 120,000 cells were harvested and prepared as previously described for FTIR microspectroscopy above.

### *3.4.2.4. Testing the effects of the growth medium on the FTIR spectroscopic signatures of hESC cells*

The effect of different growth media on cell spectra was tested by co-culturing hESCs with MEFs in either KOSR or mTESR medium. *MIXL1*$^{\text{GFP/w}}$ cells were cultured in 6 well plates, on top of low density MEFs in either KOSR or mTESR medium. hESCs were maintained for 5 passages in the same conditions, and during each passage, 120,000 of the cells were collected and prepared as previously described for FTIR microspectroscopy above.

### 3.4.3. Transmission versus transflectance comparison

To test whether the recently reported spectral artifact caused by the EFSW effect may have affected our data acquired from cells deposited on the transflectance substrate, we compared the spectral datasets of two replicates of *MIXL1*$^{\text{GFP/w}}$ cells that had been co-cultured with MEFs in KOSR based hESC medium acquired in both experimental modes. For the former studies, 120,000 cells were collected and cytospun onto MirrIR slides as previously described. For the transmission experiments, the same number of cells were cytospun onto 0.5 mm thick CaF$_2$ windows.

### 3.4.4. Focal Plane Array FTIR imaging

Focal Plane Array (FPA) FTIR infrared images were acquired at the Centre for Biospectroscopy at Monash University in Melbourne, Australia, in transflectance mode using an Agilent Technologies FTIR spectrometer (Model FTS 7000; Agilent Technologies Inc., Palo Alto, CA, USA) coupled to an infrared microscope (model 600 UMA; Agilent Technologies) using a 15× cassegrain objective coupled with a 64×64 pixel MCT liquid nitrogen cooled FPA detector. The areas of the deposits chosen for spectral acquisition were selected via visual examination by light microscopy where cells were found to be in a true monolayer, and were generally around the edges of the deposits. FTIR

spectra were acquired with 128 co-added scans, at 8 cm$^{-1}$ spectral resolution, with the binning of signals from groups of 4 adjacent pixels on the FPA resulting in each spectrum corresponding to an area of 11 μm × 11 μm on the sample plane, the approximate dimensions of single cells in the dried monolayer A Happ-Genzel apodisation function was used to truncate the interferogram, and a zero-filling factor of 2 was employed.

### 3.4.5. Data preprocessing for multivariate data analysis

The infrared images were taken into Cytospec 1.03 FTIR imaging software (Cytospec Inc., NY, and USA) in order to extract the spectral data that were deemed suitable for further analysis. The spectra first underwent a quality test whereby spectra that had maximum absorbance values outside the range of 0.6 - 0.8, arising from sample regions devoid of cells or where cells were clumped and overlaid, were rejected. This was to ensure that measurements with a good signal to noise ratio and spectral signal within the linear range of the detector's response were used in further steps. The spectra which passed the quality test underwent further preprocessing in The Unscrambler v 10.1 (Camo, Oslo, Norway) software package. Here, they were converted to their second derivatives with a Savitsky-Golay algorithm using 9 smoothing points in order to minimize baseline effects and to resolve spectral components that would otherwise be overlapping in the underivatised spectra. This procedure results in the inversion of the peaks in the underivatised spectra and so appear as minima bands. Extended Multiplicative Signal Correction (EMSC) (Martens et al., 2003), which also normalizes the data, was carried out using the following spectral ranges: 3050-2800 cm$^{-1}$, 1770-1730 cm$^{-1}$, and 1470-900 cm$^{-1}$, These ranges were selected because they contain key biological absorption bands and therefore provide the best discrimination of different sample types. The amide I and amide II regions were omitted from the analysis as these regions are thought to be influenced by the effects of resonant Mie scattering, which may have confounded our interpretation of the spectra.

### 3.4.6. Partial Least Squares Discriminant Analysis

The data set was randomly sorted into 2 new data sets comprising 2 thirds and 1 third of the spectra and used as the calibration and independent test datasets respectively for classification of spectra using Partial Least Squares Discriminant Analysis (PLS-DA) (Geladi, 1988). The calibration data matrix employed for PLS-DA consisted of the spectral dataset (multivariate X) and two Y variables with integer values of 0 or 1 coding for the each of the two modelled spectral classes. Data outliers in the calibration set were identified and removed by the examination of residual influence plots. The number of outliers constituted less than 1% of the original number of spectra for all data sets. Classification of the dataset was then carried out by predicting a Y value for each spectrum in the independent validation using PLSR models that had been generated from the calibration sets. Correct classification of each class was arbitrarily assigned to samples with predicted *Y*>0.5 for respective spectra. PLS scores and loading plots were used to determine which spectral regions contributed most to the ability to classify spectra using PLS-DA.

## 3.5. Results and Discussion

### 3.5.1. Testing for possible biases introduced by the transflectance measurement

To determine whether there were differences between spectra acquired in transflectance and transmission, which were attributable to EFSW effects, spectra were acquired from *MIXL1*[GFP/w] cells in both measurement modes and analysed using PLS-DA. No clear clustering was seen in the PLS-DA scores plots of FTIR spectral clusters derived from both replicates of *MIXL1*[GFP/w] cells that were grown on MEFs in KOSR-based medium, suggesting that spectra from the two types of measurement were highly similar

(Supplementary Figure 3.1A). Along PLS Factor 1, the loadings bands that explained the greatest variance in the dataset were the bands at ~1149 cm$^{-1}$, ~1076 cm$^{-1}$, ~1014 cm$^{-1}$, and ~991 cm$^{-1}$ (Supplementary Figure 3.1B). This was the type of variation typically seen between spectra from cells drawn from the same populations.

PLS-DA classification of the independent test spectra was poor, particularly with the two transflectance experiment derived datasets which had sensitivities of 11% and ~34% at for replicates 1 and 2 respectively. Discrimination of spectra derived from the transmission experiments was better, but still very low, with sensitivities of 54% for replicate 1 and 43% for replicate 2.

The day 4 mesendodermal differentiated MIXL1 $^{GFP/w}$ cells very thin, with thicknesses of 0.93 μm ± 0.05, and 0.91 μm ± 0.07 for the cells cultured on MEFs in KOSR-based medium, and on Matrigel in mTESR based medium respectively. Due to their thinness, the position of the nuclei in these cells would have been approximately in the same position for each experiment, and therefore any EFSW that was probing the nucleus would have been of the same intensity. This was supported by the AFM imaging of our cytospun cell populations where populations of cells from both feeder grown and Matrigel grown cells had flat cylindrical morphologies of uniform thickness, without evidence of protruding nuclei (Supplementary Figure 3.1).The p value from the two sample independent t-test was 0.80. This result showed that the cells in the monolayer deposits derived from both conditions had no statistically significant difference in thickness. In any case, differences were well below the 0.5 μm, which was the sample height difference were at which EFSW effects have been reported to be detectable (Bassan et al., 2013). Further, any remaining contributions of the EFSW would be largely nullified by our earlier spectral pre-processing. For instance, the computation of spectral derivatives and vector normalization were shown to significantly decrease the effects of the EFSW in a study by Miljković and

colleagues. (Miljković et al., 2013) Certainly the prominent spectral differences observed in second derivative spectra between the experimental conditions reported here appear unlikely to have been due to EFSW effects, a view supported by the replication of these observations across different experiments, stem cell strains and the range of measurement modalities (FPA, SR-FTIR, transflection and transmission, dried and living cells). A good example is the high lipid absorbance (C-H stretching bands 3050-2800 cm-1) observed consistently in undifferentiated human embryonic and reprogrammed stems cells in both dried transflection measurements and SR-FTIR transmission measurements on living cells, compared to differentiated progeny cell spectra (See Chapter 2, Figure 2.4 &Chapter 4, Figure 4.2).

### 3.5.2. Both cellular morphology and spectroscopic signature of hESCs are influenced by culture conditions

Mechanically passaged hESCs formed discrete colonies of tightly packed cells with scant cytoplasm and prominent nucleoli (Figures 3.1A and 3.1D). The morphology of enzymatically passaged cells was generally similar, although cells passaged on Matrigel in mTESR medium appeared smaller than those passaged on MEFs as the cultures became denser (compare Figure 3.1E with 3.1F). Also, for hESCs passaged by either enzymatic protocol, cultures became confluent with time, blurring any separation between colonies.

We compared the spectral profiles of hESCs grown under the three different culture conditions. Examination of results from this experiment clearly showed that the different culture conditions affected the spectro-phenotype of all the hESC lines tested, with PLS discriminant analysis (PLS-DA) allowing classification of independent test spectra with very high levels of accuracy (>92%, Figures 3.2A, 3.2B and 3.2C). This was in accordance with both the average second derivative spectra (Figures 3.3A, 3.3B, and 3.3C) and distinct

separation of clusters in the PLS scores plots (Figures 3.4A, 3. 4C and 3.4E), with HES3

and H9 showing very similar responses compared with the MISCES-01 line.

**Figure 3.1.** Brightfield images of human embryonic stem cells propagated using (A, D) manual passaging, (B, E) enzymatic passaging on mouse embryonic fibroblast feeder cell layers and (C, F) enzymatic passaging on Matrigel coated plates using mTESR medium. (A) Low power view of part of a manual passaged colony. (B, C) Subconfluent and (C, F) confluent cultures of enzymatically passaged cultures imaged at the same magnification. Original magnifications (A) x50, (B-F) x100.
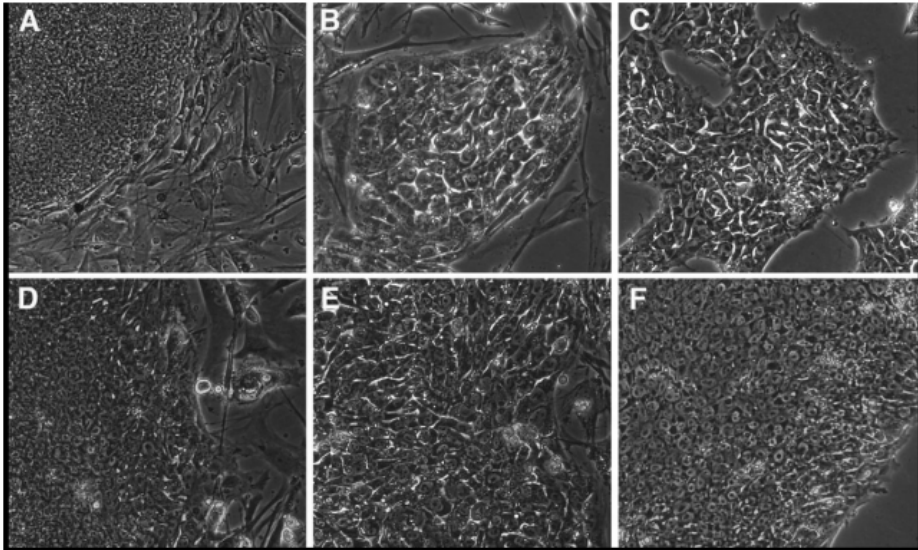
**Figure 3.2.** Predicted Y values from PLS-DA modelling for the calibration and validation set resulting from the PLS-DA discriminant analyses of all of the previously described experiments in Figures 3-8.
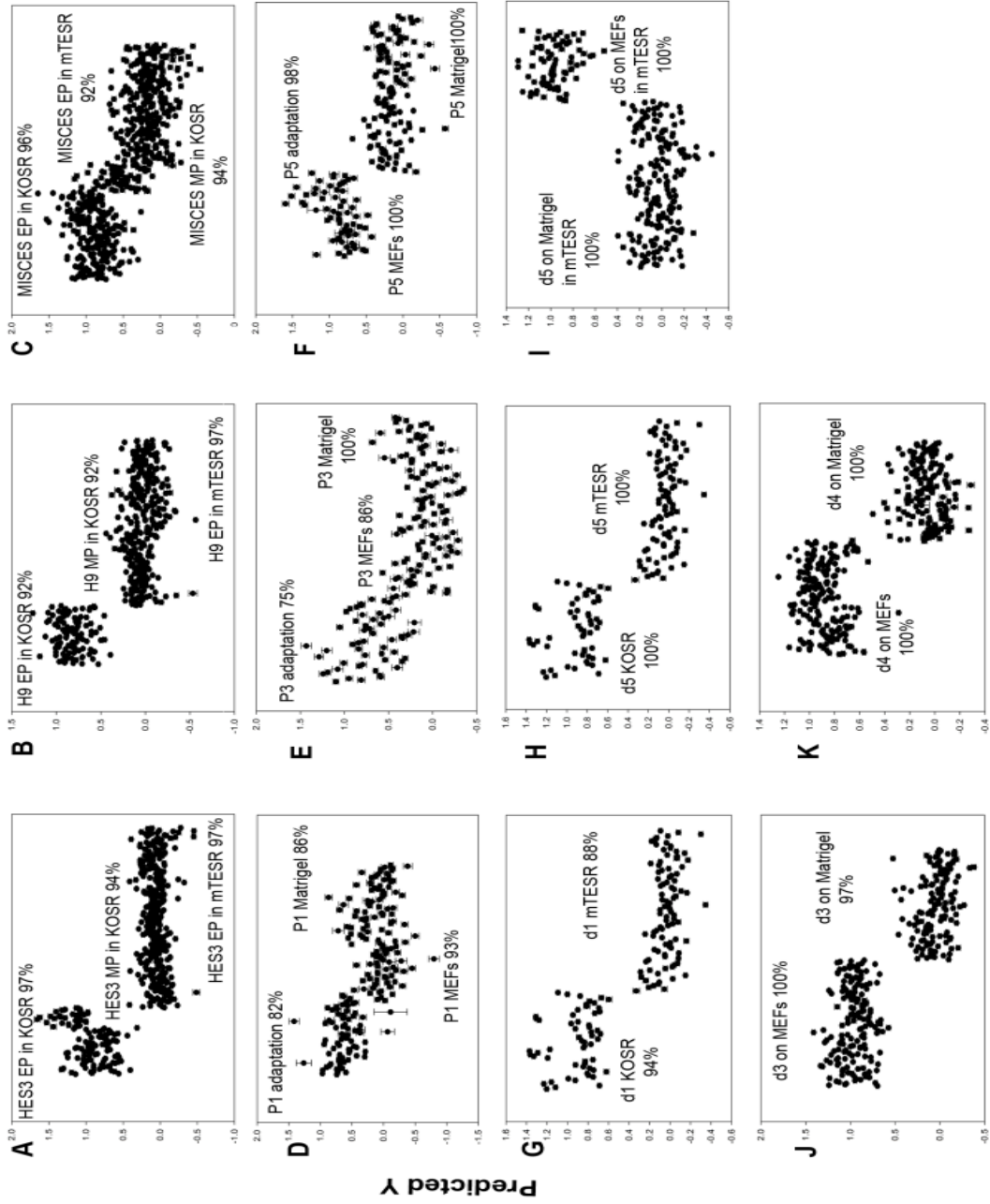
**Figure 3.3.** Average second derivative spectra for the (A) HES3, (B) H9 and (C) MISCES-01 lines enzymatically passaged onto MEFs in KOSR medium, mechanically passaged onto MEFs in KOSR medium and enzymatically passaged onto Matrigel in mTESR medium. Average second derivative spectra for $MIXL1^{GFP/w}$ lines previously cultured on Matrigel in mTESR media, and transferred onto MEFs in KOSR media (adaptation group), continuously cultured on MEFs in KOSR medium (MEFs control) and continuously cultured on Matrigel in mTESR medium, after 1 passage (D), 3 passages (E) and 5 passages (F) back into their assigned conditions. Average second derivative spectra for $MIXL1^{GFP/w}$ cells in co culture with MEFs and supplemented with either KOSR medium or mTESR medium after 1 passage (G) and 5 passages (H). Average second derivative spectra of $MIXL1^{GFP/w}$ co-cultured with MEFs and maintained for 5 passages in mTESR medium versus the same line, after 5 passages on Matrigel, in mTESR medium.

**Figure 3.4.** (A) PLS-DA scores plots from the analysis of the spectral data from the human embryonic stem cell line, HES3, enzymatically passaged onto MEFs in KOSR medium, mechanically passaged, and enzymatically passaged onto Matrigel in mTESR medium, and (B) PLS-DA loadings plots used to explain the spectral differences explaining the clustering observed in the previous scores plot. (C) PLS-DA scores plot from the analysis of the spectral data from the human embryonic stem cell line, H9, enzymatically passaged onto MEFs in KOSR medium, mechanically passaged onto MEFs in KOSR medium, and enzymatically passaged onto Matrigel in mTESR medium, and (D) PLS-DA loadings plots. (E) PLS-DA scores plot from the analysis of the spectral data from the human embryonic stem cell line, MISCES-01, enzymatically passaged onto MEFs in KOSR medium, mechanically passaged onto MEFs in KOSR medium, and enzymatically passaged onto Matrigel in mTESR medium and (F) PLS-DA loadings plots

In general, cells enzymatically passaged onto Matrigel in mTESR medium had spectra consistent with higher concentrations of amino acids and phosphorylated molecules, with an intense carboxylate band at ~1392 cm$^{-1}$, and bands from PO$_2^-$ stretching vibrations at ~1238 cm$^{-1}$ and ~1080 cm$^{-1}$, respectively, seen in the average second derivative spectra (Figures 3.3A, 3.3B and 3.3C). This was also evident by the clustering patterns along Factor 1 (Figures 3.4C and 3.4E) or Factor 2 (Figure 3.4A) in PLS scores plots associated with prominent loadings for these bands (Figures 3.4B, 3.4D and 3.4F and Table 3.1). By contrast, HES3 and H9 hESCs enzymatically passaged onto MEFs in KOSR medium contained higher levels of carbohydrates, as shown by intense negative peaks at ~1153 cm$^{-1}$ and ~1022 cm$^{-1}$ (Figures 3.3A, 3.3B and 3.3C). Spectral clusters from cells that were enzymatically passaged onto MEFs in KOSR medium clearly separated along Factor 1 (Figures 3.4A and 3.4C), explained by strong positive loadings for these bands ascribed to carbohydrates (Figures 3.4B and 3.4D). For all three cell lines, mechanically dissociated cells had higher lipid concentrations, as indicated by more intense $\nu_s$CH$_2$, $\nu_{as}$CH$_2$ stretching modes in the average second derivative spectra at ~2850 cm$^{-1}$ and 2920 cm$^{-1}$ respectively (Figure 3.3), and separation of clusters of spectra from mechanically dissociated cells along Factor 1 with the MISCES-01 hESC line (Figure 3.4E) or along Factor 2 with the HES3 and H9 hESC lines (Figures 3.4A and 3.4C).

The effect of the culture conditions appeared to be least pronounced with the MISCES-01 line, as demonstrated by the less prominent separation of spectral clusters from the three treatment cohorts in scores plots. This contrasted with the excellent separation of spectra observed between clusters of spectra from similarly treated samples in the HES3 and H9 groups. The data from this initial study convincingly demonstrated that different tissue culture platforms affected the IR spectroscopic phenotypes of human embryonic stem cells,

and that the extent of these effects appeared to be dependent on the individual hESC line used.

Our findings are in agreement with other studies which have reported alterations in hESCs caused by the culturing environment (Carpenter et al., 2003; Costa, 2008; Maitra et al., 2005; Mitalipova et al., 2005; Sjögren-Jansson et al., 2005). However, those studies focused on changes in morphology, gene expression, or cell surface antigen expression, rather than macromolecular information. Here we present the first study showing environmentally-induced changes in hESCs occurring at the biochemical level. For instance, different media and substrates have both been found to influence proliferation rates and cellular attachment abilities (Sjögren-Jansson et al., 2005). Subtle morphological differences between cells cultured via different methods have also been consistently observed in our laboratory (Figure 3.1), with Matrigel-based culture systems producing smaller cells than feeder-based systems. Further, the choice of passaging technique has also been found to significantly influence cellular genotype and phenotype. Bulk culture passaging methods (Carpenter et al., 2003; Mitalipova et al., 2005; Thomson et al., 1998), which include non-enzymatic (cell dissociation buffer, CDB) and enzymatic passaging (collagenase/trypsin, CT) promote the development of karyotypic abnormalities in some late passage embryonic stem cell lines (Maitra et al., 2005). In contrast, chromosomal alterations have been found to either not occur or occur less frequently (Costa, 2008; Maitra et al., 2005), after extended periods of mechanical passaging.

### 3.5.3. Dependence of phenotype on growth media and substrate

To determine the effect on phenotype of the growth medium alone, we cultured hESCs on the same substrate (MEFs) but in different growth media (KOSR based or mTESR). PLS-DA classification showed the $MIXL1^{GFP/w}$ cells that were grown in KOSR-based medium could be correctly discriminated from those grown in mTESR medium with 94% accuracy.

Furthermore, the group that was grown in mTESR medium could also be discriminated from the KOSR treated group with 88% accuracy after one passage (Figure 3.2G). In addition, after 5 passages, these spectroscopic differences became more distinct as verified by the improved classification outcomes with greater separation of the spectral clusters on the PLS scores plots. Average spectra (Figures 3.3G and 3.3H) and PLS scores and loadings plots (Figures 3.5A and 3.5C and Table 3.3) indicated that cells grown in mTESR medium for 1 and 5 passages had higher levels of amino acids, phosphorylated molecules and carbohydrates compared to the cells cultured in KOSR-based medium.

Nevertheless, when we then compared the spectra of the hESCs grown in mTESR medium on MEFs with previously acquired spectra of the same cells grown in mTESR medium on a Matrigel substrate, we found significant differences between the two treatment classes, which suggested that the growth substrate also contributed to the observed phenotypic differences between hESCs grown in the different culture systems (Figures 3.5E and 3.5F). This finding that the substrate type affected the cell phenotype was verified in further experiments whereby spectra from $MIXL1^{GFP/w}$ hESCs grown in mTESR medium and cultured either on MEFs or Matrigel were compared. Classification of independent test spectra by PLS-DA could distinguish spectra from each treatment group with 100% accuracy (Figure 3.2). This was confirmed by PLS scores plots that revealed an excellent separation of clusters of spectra representing cells cultured on feeder layers and on Matrigel along Factor 1 (Figure 3.5E). Loadings plots showed that carbohydrate (bands at ~1149 cm$^{-1}$, ~1080 cm$^{-1}$, ~1018 cm$^{-1}$ and ~991 cm$^{-1}$) levels were higher in cells grown on mouse embryonic fibroblast feeder cells (MEFs) compared to those grown on Matrigel (Figure 3.5F and Table 3.5). In contrast to cells grown on MEFs in KOSR medium, those grown on MEFs in mTESR medium had lower lipid levels compared to hESCs grown on Matrigel (Figure 3.3 compared with Figures 3.5E and 3.5F). This data correlated with

**Figure 3.5.** PLS-DA scores plots (A) and (C) and loadings plots (B), (D) from the analysis of spectra acquired from *MIXL1*<sup>GFP/w</sup> in co culture with MEFs and supplemented with either KOSR medium or mTESR medium. PLS-DA analysis was performed using FTIR spectra acquired of cells following 1 passage (A) and (B), and 5 passages (C) and (D), in the their assigned growth conditions. PLS-DA scores plots (E) and corresponding loadings plot (F) from the comparison made between the day 5 *MIXL1*<sup>GFP/w</sup> co-cultured with MEFs, in mTESR as shown in (C) and (D) and *MIXL1*<sup>GFP/w</sup> cells in mTESR medium and maintained on a Matrigel substrate after 5 days.
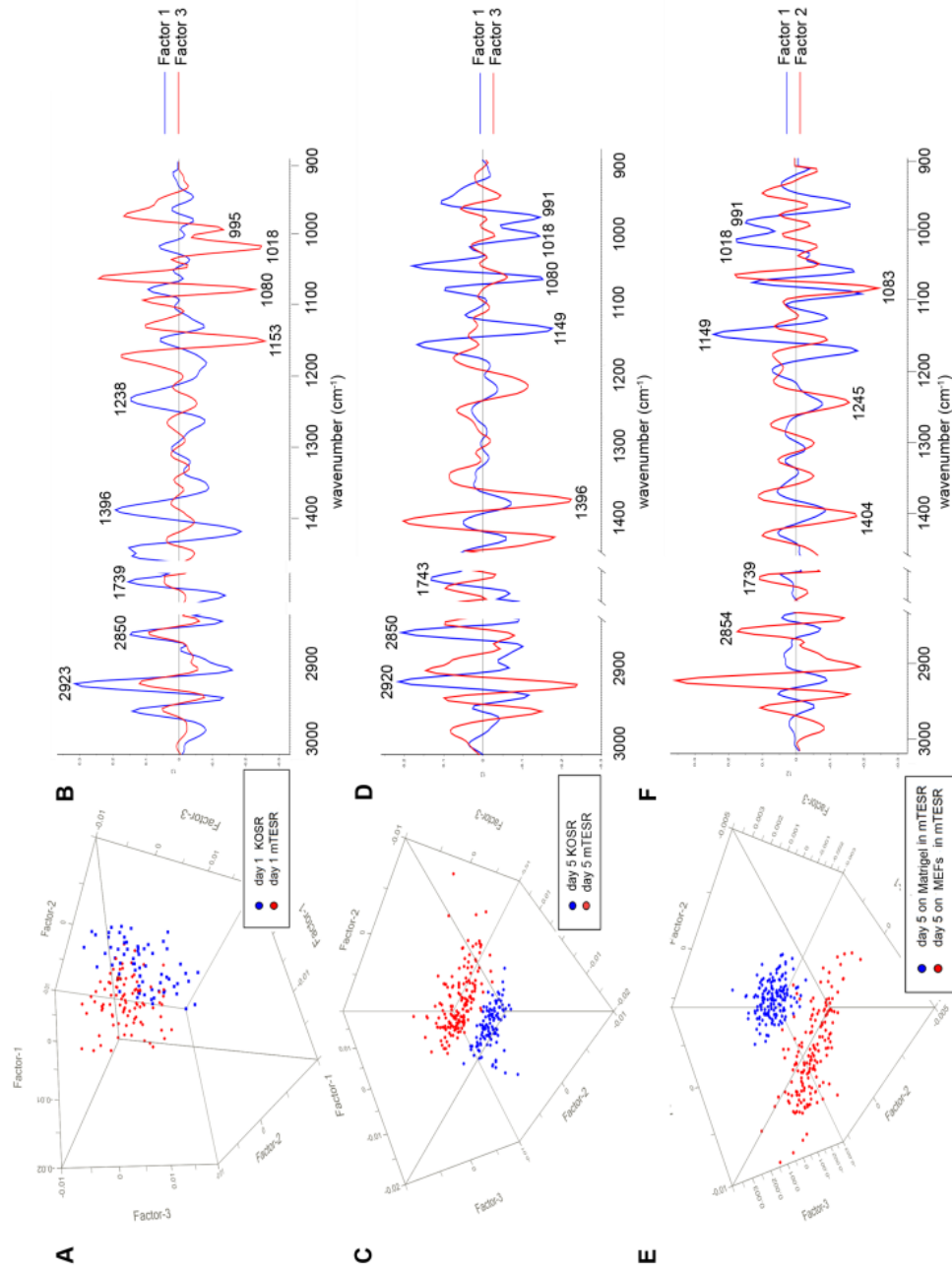
differences seen in the average second derivative spectra (Figures 3.3A, 3.3B and 3.3C compared with Figure 3.3H).

Since the growth medium, substrate and dissociation methods had such a profound influence on cell phenotype, we then wanted to determine if this influence persisted following several passages onto different tissue culture platforms. We therefore compared the spectra of *MIXL1*^GFP/w hESCs cultured on Matrigel and then enzymatically passaged onto MEFs in KOSR medium, to spectra of hESCs continuously cultured on MEFs in KOSR medium, and to those continuously cultured on Matrigel in mTESR medium. Results from the PLS-DA analysis on the independent validation set showed that cells originally enzymatically passaged onto Matrigel in mTESR medium and subsequently co-cultured with MEFs in KOSR medium, maintained a phenotypic distinctness from cells that had been continuously cultured on MEFs, even after five days in the new environment. PLS-DA (Figures 3.2D, 3.2E and 3.2F): showed that these cells were classified as distinct from the continuously passaged cells with 97% correct discrimination after day 1 of growth in the new conditions; 75% correct after 3 days; and 98% correct after 5 days. However, the phenotype of the cells that were continuously passaged onto Matrigel in mTESR medium could be distinguished from both hESCs continuously passaged onto feeder layers in KOSR medium and from hESCs which had been transferred from a Matrigel-based to MEF dependent conditions with 100% accuracy at the two later time points (Figure 3.2). This suggested that although some adaptation to the new environment had occurred, the original Matrigel-based growth environment had some continuing effect on phenotype even after several passages.

Average second derivative spectra (Figures 3.3D, 3.3E and 3.3F) and PLS scores plots (Figures 3.6B, 3.6D and 3.6F) showed that the major spectroscopic differences between cells enzymatically passaged continuously onto Matrigel in mTESR medium and the other

**Figure 3.6.** PLS-DA scores plots (A), (C) and (E) and loadings plots (B), (D) and (F) from the analysis of spectra acquired from *MIXL1*<sup>GFP/w</sup> cells previously enzymatically passaged onto Matrigel in mTESR medium and readapted to co culture with MEFs in KOSR medium, compared with spectra of *MIXL1*<sup>GFP/w</sup> cells continuously maintained on MEFs in KOSR medium, and *MIXL1*<sup>GFP/w</sup> cells continuously Matrigel. PLS-DA analysis was performed using FTIR spectra acquired of the cells following 1 passage (A) and (B), 3 passages (C) and (D), and 5 passages (E) and (F) back onto the same growth conditions.

treatment groups was lower absorbance for lipid components at ~2920 cm$^{-1}$, ~2850 cm$^{-1}$ and ~1739 cm$^{-1}$ at all measured time points. PLS-DA indicated that cells passaged onto Matrigel in mTESR medium and subsequently passaged onto MEFs in KOSR medium were consistently different from hESCs continuously grown on MEFs in terms of carbohydrate absorbance, with loadings plots indicating lower levels in the cells grown continuously on feeder layers at all time points (C-O stretching bands at 1150 cm$^{-1}$ and 1022 cm$^{-1}$; Figure 3.6). Further, spectra from the hESCs grown on MEFs initially had less lipid than those recently adapted to these conditions (days 1 and 3) as indicated by PLS loadings for bands at 2920 cm$^{-1}$ and 2850 cm$^{-1}$, and then higher lipid levels at the last time point (day 5). These differences are summarised in Table 3.2.

### 3.5.4. The effect of culture conditions on the phenotype of differentiated cells

Upon observing culture condition influenced "spectroscopic signatures" in undifferentiated hESCs, we wished to determine the heritability of these spectroscopic differences in lineage committed progeny. To verify that the cells had differentiated to mesendodermal lineages, flow cytometry was performed on the *MIXL1*$^{GFP/w}$ cells after 3 and 4 days of differentiation in BMP4/Act A containing medium (Figures 3.7A-3.7F). Undifferentiated hESCs, which were used as the control for these experiments, are known to express a variety of stem cell surface antigens such as EpCAM and do not express GFP from the *MIXL1* locus. By 4 days of differentiation, some of the *MIXL1*$^{GFP/w}$ hESCs had down-regulated cell surface marker EpCAM and the majority expressed MIXL1-GFP and platelet derived growth factor PDGFRα, an early mesendodermal marker, whereas undifferentiated cells were EpCAM positive but negative for MIXL1 and PDGFRα expression.

**Figure 3.7.** Flow cytometric analysis results of undifferentiated *MIXL1*<sup>GFP/w</sup> cells maintained on MEFs in KOSR medium (A) and differentiated in BMP4 ACTIVIN A towards mesendodermal lineages and stained for expression of EPCAM and PDGFRα at day 3 (B) and at day 4 (C). Percentages of cells within the indicated quadrants or regions are shown. Panels (D-F) show the flow cytometric analysis results of undifferentiated *MIXL1*<sup>GFP/w</sup> cells maintained on Matrigel in mTESR medium (D) and differentiated in BMP4 ACTIVIN A and stained for expression of EPCAM and PDGFRα at day 3 (E) and at day 4 (F). Percentages of cells within the indicated quadrants or regions are shown. Panels (G-H) show the average second derivative spectra for *MIXL1*<sup>GFP/w</sup> cells after 3 days (G) and 4 days (H) of differentiation with BMP4 ACTIVIN A.

Surprisingly, even after 4 days of differentiation, and despite the flow cytometry data showing highly comparable differentiation kinetics, the phenotypic effects of the initial growth conditions were not entirely eliminated. PLS-DA classification could discriminate between spectra from day 3 progeny derived from hESCs maintained under the two growth conditions with greater than 97% accuracy (Figure 3.2J). After 4 days of differentiation, both groups could be discriminated with 100% accuracy (Figure 3.2K). PLS-DA scores plots at day 3 and day 4 (Figure 3.8A and Figure 3.8C) showed excellent separation along Factor 1 between clusters of spectra from the differentiated cells. Loadings values (Figures 3.8B and 3.8D and Table 3.4) revealed that spectra of progeny derived from $MIXL1^{GFP/w}$ hESCs maintained in mTESR medium on a Matrigel substrate, from both time points, exhibited more intense IR absorbances at ~2923 cm$^{-1}$, ~2850 cm$^{-1}$ and ~1739 cm$^{-1}$ from lipids, and at ~1396 cm$^{-1}$ from the COO$^{-}$ stretching vibrations of amino acid side chains, and at ~1238 cm$^{-1}$ from phosphorylated molecules. Conversely, spectra from hESC progeny grown on MEFs had higher absorbances from glycogen and other carbohydrates (bands at ~1153 cm$^{-1}$, ~1080 cm$^{-1}$and ~1018 cm$^{-1}$). This result was noteworthy, as we had previously found that undifferentiated MEF-derived $MIXL1^{GFP/w}$ cells had higher lipid absorbances than the Matrigel-derived hESCs. Average second derivative spectra of the day 3 and day 4 progeny (Figure 3.7G and Figure 3.7H) corroborated the spectroscopic differences indicated by the PLS-DA.

Interestingly, differences between the progeny from both culture conditions were different from those that had initially allowed their discrimination from the undifferentiated hESCs from which they were derived. Prior to differentiation, spectra from MEF-derived hESCs had higher lipid absorbances compared to spectra from Matrigel-derived hESCs. However the converse was true following differentiation. The finding that lipid stores diminished during the loss of pluripotency corroborates data previously published by our laboratory

**Figure 3.8.** PLS-DA scores plots (A), and loadings plots (B) of spectra acquired from *MIXL1*<sup>GFP/w</sup> after 3 days of differentiation towards mesendodermal lineages using BMP4 ACTIVIN A. PLS-DA scores plots (C), and loadings plots (D) of spectra acquired from *MIXL1*<sup>GFP/w</sup> after 4 days of differentiation towards mesendodermal lineages
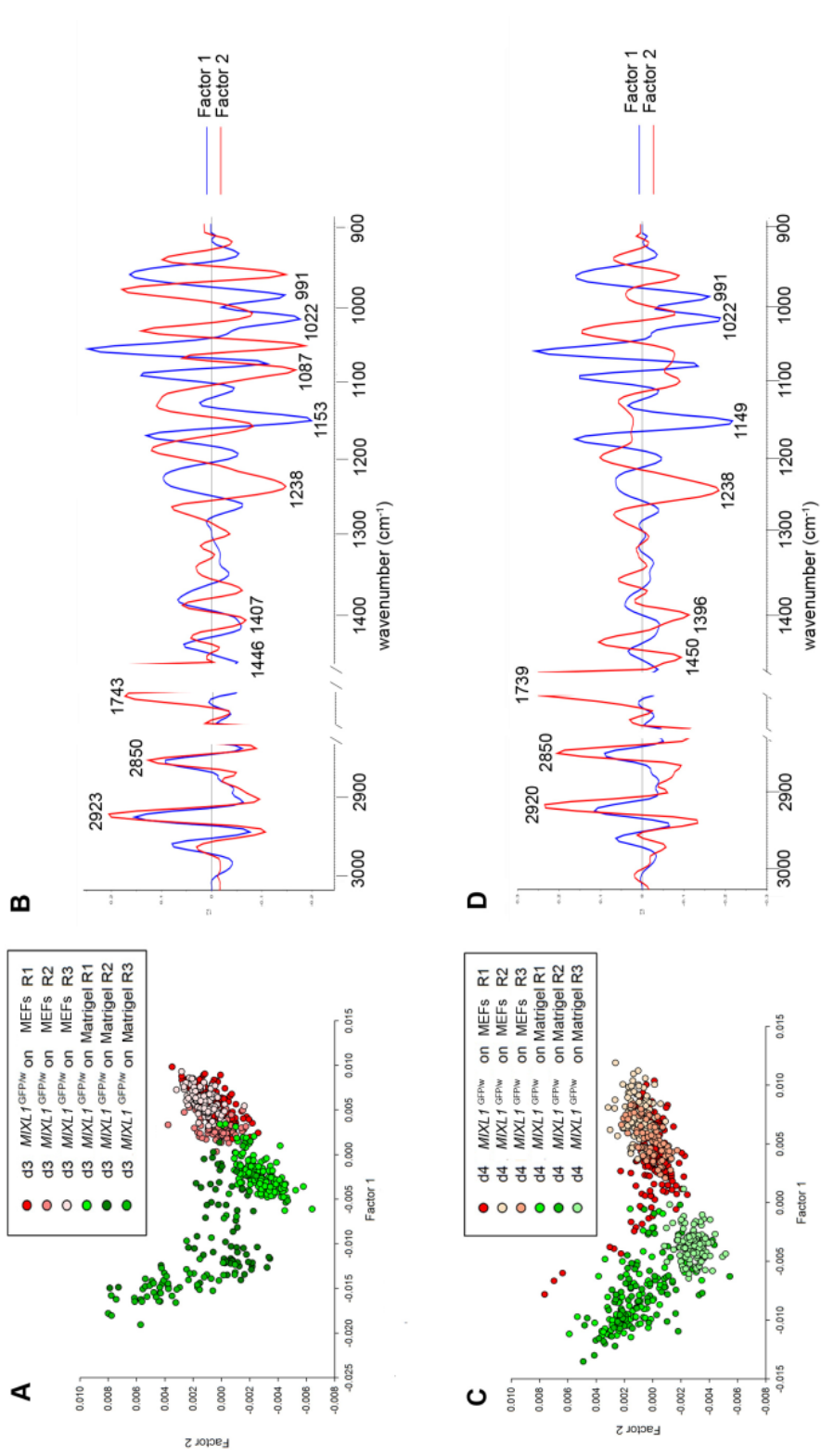
and other groups (Heraud et al., 2010; Pijanka et al., 2010). A starting hypothesis may be that the eicosanoid signaling pathway (Yanes et al., 2010), thought to be involved in lipid depletion during differentiation, may be impaired in the Matrigel cultured cells, causing less efficient lipid down-regulation than the MEF-derived hESCs.

A possible explanation for the hESCs' "memory" of their original conditions may be due to different culturing environments affecting the endogenous production of growth factors which impacted differentiation. FGF and nodal (Ginis et al., 2004; Sperger et al., 2003), activators of the SMAD1/5/8 signaling pathway such as BMPs (Sato et al., 2003), as well as nodal activin signaling antagonists (Brandenberger et al., 2004) are all possible candidates. An alternative explanation may involve culture-induced epigenetic mechanisms, which have been well documented in the literature (Allegrucci et al., 2007; Downing et al., 2008). For example, one group described various culture-dependent methylation changes that were heritable even after differentiation of their hESC lines (Allegrucci et al., 2007). Another study reported transcriptional and epigenetic modifications, thought to reduce transcriptional heterogeneity, in normoxic cultured (2% $O_2$) hESCs compared to 21% $O_2$ cultured hESCs (Downing et al., 2008). Transcriptional heterogeneity within hESCs has been known to give rise to suboptimal differentiation efficiencies although no evidence of this was found in our study.

## 3.6. Conclusion

This study emphasises the view that stem cell biologists need to be aware that hESCs and their differentiated progeny, maintained by different means, are not spectroscopically equivalent and this knowledge underscores the need to develop standardised protocols for the production of cells destined for use in the clinic.

## 3.7. Acknowledgements

| Band maxima 2nd derivative spectra (cm-1) | | | PC1 loadings (cm-1) | PC2 loadings (cm-1) | Band assignments |
|---|---|---|---|---|---|
| HES3, H9 and MISCES lines | | | HES3, H9 and MISCES lines | HES3, H9 and MISCES lines | HES3, H9 and MISCES lines |
| Enzymatically passaged onto MEFs in KOSR | Mechanically passaged onto MEFs in KOSR | Enzymatically passaged onto Matrigel in mTESR | | | |
| 2923,2927,2923 | 2923,2923,2923 | 2923,2923,2923 | 2920,2920,2920 | 2920,2920,2916 | C-H asymmetrical stretching of lipid groups |
| 2850,2850,2850 | 2850,2850,2850 | 2850,2850,2850 | 2850,2850,2850 | 2850,2850,2846 | C-H symmetrical stretching of lipid groups |
| 1743,1739,1739 | 1739,1739,1739 | 1739,1739,1739 | 1747,1739,1731 | 1739,1735,1739 | C=O stretching of lipid esters |
| 1396,1396,1396 | 1396,1396,1392 | 1392,1396,1392 | 1392,1392,1396 | 1388,1404,1484 | COO- stretching vibrations of amino acid side chains |
| 1242,1224,1238 | 1238,1238,1238 | 1238,1238,1238 | 1234,1238,1234 | 1238,1245,1253 | P=O asymmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1080,1080,1083 | 1083,1080,1080 | 1080,1083,1080 | 1091,1091,1091 | 1080,1080,1091 | P=O symmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1153,1153,1153 | 1157,1153,1153 | 1153,1153,1153 | 1153,1153,1153 | 1149,1153,1149 | C-O vibrations from glycogen and other carbohydrates |
| 1022,1022,1022 | 1022,1022,1022 | 1022,1022,1022 | 1022,1018,1022 | 1018,1022,1018 | C-O vibrations from glycogen and other carbohydrates |
| 995,995,995 | 995,995,995 | 995,995,995 | 995,995,995 | 991,995,991 | C-O stretch from RNA ribose-chain |

**Table 3.1:** Band maxima of the human embryonic stem cell lines HES3, H9 and MISCES maintained on MEFs in KOSR medium and enzymatically passaged, maintained on MEFs in KOSR medium and mechanically passaged, and maintained on Matrigel and enzymatically passaged.

| Band maxima 2nd derivative spectra (cm^-1) | | | Factor 1 loadings (cm^-1) | Factor 2 loadings (cm^-1) | Band assignments |
|---|---|---|---|---|---|
| P1, P3, P5 adaptation | P1, P3, P5 on MEFs | P1, P3, P5 on Matrigel | | | |
| 2923,2923,2923 | 2923,2923,2923 | 2923,2923,2923 | 2920,2920,2920 | 2927,2923,2920 | C-H asymmetrical stretching of lipid groups |
| 2850,2850,2850 | 2850,2850,2850 | 2850,2850,2850 | 2850,2850,2850 | 2854,2850,2850 | C-H symmetrical stretching of lipid groups |
| 1739,1739,1739 | 1739,1739,1739 | 1739,1739,1739 | 1739,1731,1735 | 1747,1739,1739 | C=O stretching of lipid esters |
| 1396,1396,1396 | 1396,1396,1396 | 1396,1396,1396 | 1400,1396,1392 | 1388,1396,1404 | COO^- stretching vibrations of amino acid side chains |
| 1238,1238,1238 | 1238,1238,1242 | 1238,1238,1238 | 1242,1238,1249 | 1249,1238,1245 | P=O asymmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1080,1080,1080 | 1080,1080,1083 | 1083,1080,1080 | 1083,1080,1087 | 1091,1080,1087 | P=O symmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1153,1153,1153 | 1153,1153,1153 | 1153,1153,1153 | 1157,1149,1145 | 1145,1153,1145 | C-O vibrations from glycogen and other carbohydrates |
| 1022,1022,1022 | 1022,1022,1022 | 1022,1022,1018 | 1026,1018,1014 | 1014,1022,1018 | C-O vibrations from glycogen and other carbohydrates |
| 995,995,995 | 995,995,995 | 995,995,995 | 991,991,991 | 987,995,987 | C-O stretch from RNA ribose-chain |

**Table 3.2:** Band maxima of the human embryonic stem cell line *MIXL1*[GFP/w] previously cultured on Matrigel in mTESR media and transferred over to culture with feeders, after 1, 3 and 5 passages. Band maxima are also shown of the two control groups, which consisted of *MIXL1*[GFP/w] cells that had been continuously co-cultured with feeders in KOSR medium (feeder control), and *MIXL1*[GFP/w] cells that had been continuously cultured on Matrigel in mTESR medium (Matrigel control).

| Band maxima 2$^{nd}$derivative spectra (cm$^{-1}$) | | Factor 1 loadings (cm$^{-1}$) | Factor 3 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|---|
| Day 1 and Day 5 in KOSR based medium | Day 1 and Day 5 in mTESR medium | | | |
| 2923,2923 | 2923,2923 | 2923,2920 | 2923,2923 | C-H asymmetrical stretching of lipid groups |
| 2850,2850 | 2850,2850 | 2854,2850 | 2850,2854 | C-H symmetrical stretching of lipid groups |
| 1739,1739 | 1739,1739 | 1739,1743 | 1739,1739 | C=O stretching of lipid esters |
| 1396,1396 | 1396,1396 | 1392,1396 | 1384,1396 | COO$^-$ stretching vibrations of amino acid side chains |
| 1238,1238 | 1238,1238 | 1234,1238 | 1242,1234 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1080,1080 | 1080,1080 | 1080,1080 | 1080,1080 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1153,1153 | 1153,1153 | 1149,1149 | 1153,1157 | C-O vibrations from glycogen and other carbohydrates |
| 1018,1022 | 1018,1018 | 1018,1018 | 1018,1018 | C-O vibrations from glycogen and other carbohydrates |
| 991,995 | 991,995 | 983,991 | 995,995 | C-O stretch from RNA ribose-chain |

**Table 3.3:** Band maxima of the human embryonic stem cell line *MIXL1*[GFP/w] co-cultured with feeders and maintained in either KOSR medium or mTESR medium, after 1 and 5 passaging events.

| Band maxima 2nd derivative spectra (cm-1) | | PC1 loadings (cm-1) | PC2 loadings (cm-1) | Band assignments |
|---|---|---|---|---|
| Day 3 and Day 4 *MIXL1*GFP/w on MEFs in KOSR based medium | Day 3 and Day 4 *MIXL1*GFP/w on Matrigel in mTESR medium | Day 3 and Day 4 *MIXL1*GFP/w | Day 3 and Day 4 *MIXL1*GFP/w | Day 3 and Day 4 *MIXL1*GFP/w |
| 2920,2923 | 2923,2923 | 2923,2923 | 2920,2920 | C-H asymmetrical stretching of lipid groups |
| 2850,2850 | 2850,2850 | 2854,2850 | 2850,2850 | C-H symmetrical stretching of lipid groups |
| 1739,1739 | 1739,1739 | 1743,1739 | 1735,1731 | C=O stretching of lipid esters |
| 1396,1396 | 1392,1392 | 1388,1407 | 1407,1396 | COO- stretching vibrations of amino acid side chains |
| 1238,1238 | 1238,1234 | 1234,1234 | 1238,1238 | P=O asymmetrical stretching of PO2 phosphodiesters from phosphorylated molecules |
| 1080,1080 | 1080,1080 | 1080,1091 | 1087,1072 | P=O symmetrical stretching of PO2 phosphodiesters from phosphorylated molecules |
| 1153,1153 | 1153,1157 | 1153,1149 | 1157,1157 | C-O vibrations from glycogen and other carbohydrates |
| 1022,1018 | 1022,1018 | 1022,1018 | 1018,1014 | C-O vibrations from glycogen and other carbohydrates |
| 995,995 | 995,983 | 991,991 | 987,991 | C-O stretch from RNA ribose-chain |

**Table 3.4:** Band maxima of the human embryonic stem cell line *MIXL1*GFP/w for spectra from the hESC line *MIXL1*GFP/w, differentiated towards mesendodermal lineages in BMP4 ACTIVIN A supplemented differentiation medium.

| Band maxima 2$^{nd}$ derivative spectra (cm$^{-1}$) | | Factor 1 loadings (cm$^{-1}$) | Factor 3 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|---|
| Day 5 on Matrigel in mTESR medium | Day 5 on feeders in mTESR medium | | | |
| 2923 | 2923 | 2923 | 2923 | C-H asymmetrical stretching of lipid groups |
| 2850 | 2850 | 2850 | 2854 | C-H symmetrical stretching of lipid groups |
| 1739 | 1739 | 1728 | 1739 | C=O stretching of lipid esters |
| 1396 | 1396 | 1400 | 1404 | COO$^-$ stretching vibrations of amino acid side chains |
| 1242 | 1238 | 1245 | 1245 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1083 | 1080 | 1091 | 1083 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1153 | 1153 | 1149 | 1157 | C-O vibrations from glycogen and other carbohydrates |
| 1022 | 1018 | 1018 | 1022 | C-O vibrations from glycogen and other carbohydrates |
| 995 | 991 | 991 | 995 | C-O stretch from RNA ribose-chain |

**Table 3.5:** Band maxima of the human embryonic stem cell line *MIXL1*[GFP/w] co-cultured with feeders and maintained for 5 passages in mTESR medium versus the same line, after 5 passages on Matrigel, in mTESR medium.

**Supplementary Figure 3.1.** AFM image of a monolayer of human embryonic stem cells cytospun onto a MirrIR slide.
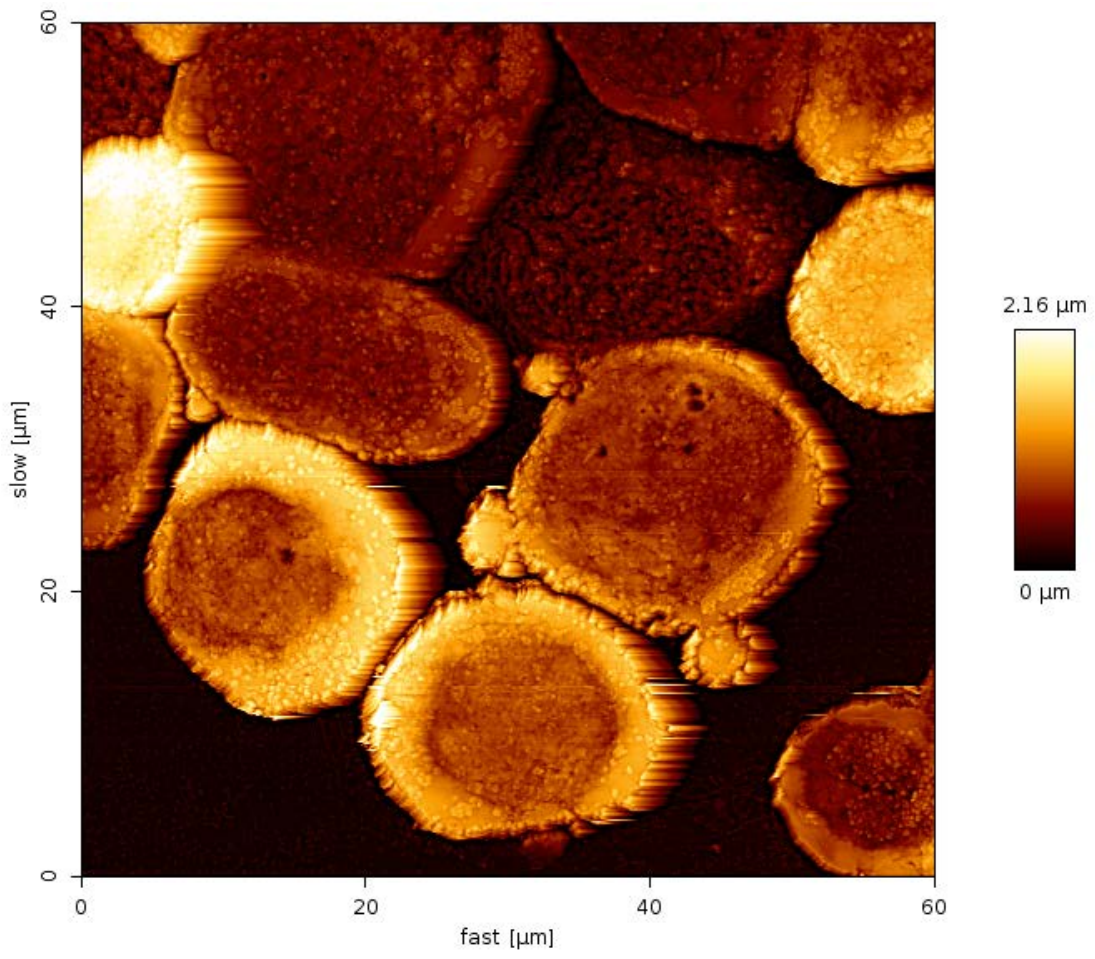
**Supplementary Figure 3.2.** (A) PLS-DA scores plots from the analysis of the spectral data from two replicate cultures of cytospun *MIXL1*$^{\text{GFP/w}}$ deposits acquired in both transflectance and transmission modes (B) PLS-DA loadings plots used to explain the spectral differences explaining the clustering observed in the previous scores plot.
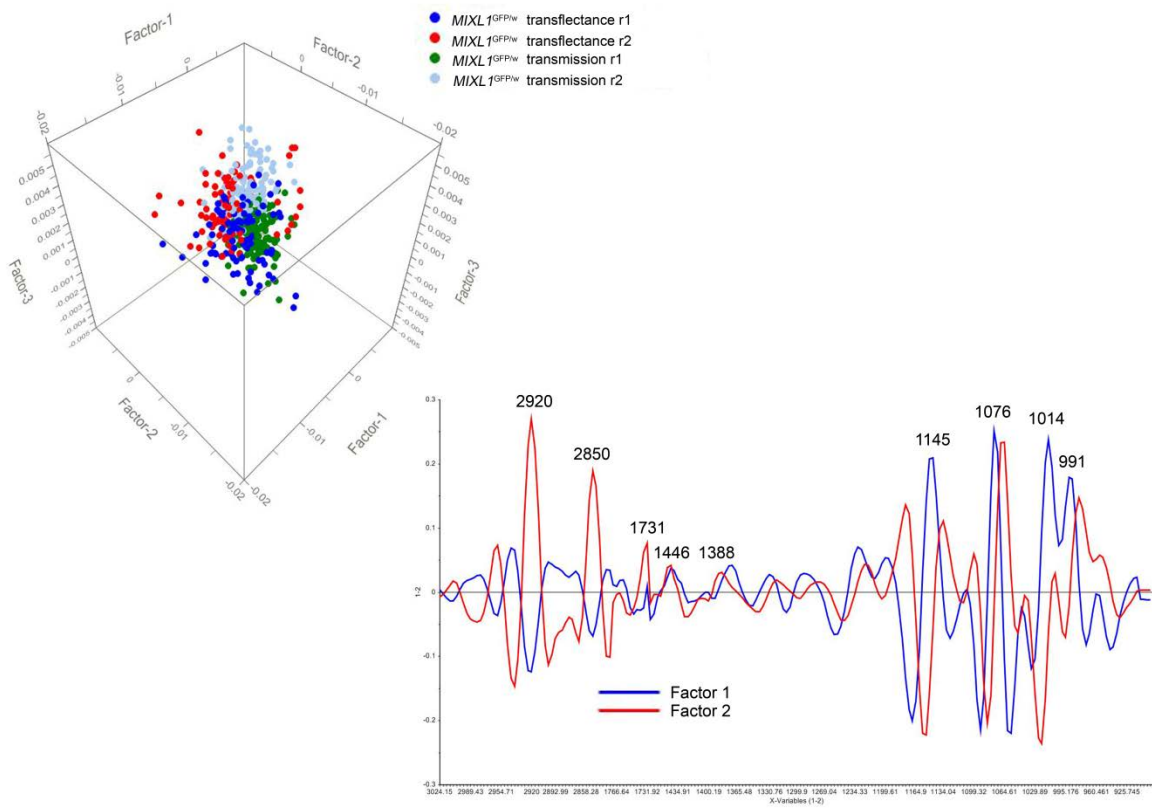
CHAPTER **4**

Preliminary investigation of live stem cells using Fourier transform infrared synchrotron microspectroscopy

## 4.1. Introduction

This Chapter describes FTIR transmission studies that were performed in parallel with the transflectance studies that have been described in Chapters 2 and 3. However, the spectra which were recorded were not of dried cell monolayers, but of live stem cells, and using synchrotron mid infrared radiation rather than a globar source of IR. This data is currently being prepared for publication.

The advantages of acquiring biological spectra in their live, rather than dehydrated state are manifold. First, it is advantageous to acquire spectra of living cells rather than dehydrated ones since these experimental conditions would result in the data more accurately reflecting their *in vivo* macromolecular composition. This is particularly important in light of recent work, which has shown that histological fixation can significantly alter cellular biochemistry (Ó Faoláin et al., 2005; Whelan et al., 2011). Another benefit is that live cell spectra are less affected by resonant Mie scattering, known to distort the spectra of fixed cells (Bassan et al., 2009), due to minimal refractive index differences between the cellular interior and its surrounding aqueous environment (Vaccari et al., 2012). Further, by using synchrotron coupled sources of IR radiation, spectra can even be attained of *individual* cells, due to the higher brilliance and spatial resolution that can be provided.

### 4.1.1. Sampling chambers used in live cell FTIR microspectroscopy

A number of sample holders for acquiring infrared spectra from live cells have been reported in the literature. Disadvantages with these sample chambers that have been identified include the difficulty in maintaining consistent optical path lengths from one experiment to another, particularly in the flow through designs that are subject to pressure pulsing. These chambers also risk the loss of the cell from the field of view in the event of

high medium exchange rates. However, the disadvantage of low exchange rates means that the cells remain in a suboptimal milieu that includes metabolites and extracellular matrix (ECM) components such as fibrous proteins and glycosaminoglycans. This is particularly problematic in the designs that do not incorporate a flow component. Nonetheless, these issues can be ameliorated with the utilisation of more elaborate liquid cell devices, such as the one used by Vaccari et al. (2012), whereby a porous septum enables the medium to slowly diffuse through to the separate cellular compartment.

### 4.1.2. Overcoming the 'water absorption barrier' in living cell studies

Whilst live cell spectra are not as affected by light scattering compared to those from dried cells, they possess their own technical and spectroscopic limitations, largely concerning water's strong mid infrared absorbing properties. Spectra derived from cells maintained under physiological conditions are heavily dominated by water bands, which have been reported to account for greater than 80% of the sampled volume (~48% extracellular and ~36% intracellular, when using a knife edge aperture to target a single cell) (Vaccari et al., 2012), and inconveniently coexist in spectral regions corresponding to key biological bands. At present, the exact structure of intracellular water remains a contentious issue (Jasnin et al., 2008), but at this time, the contribution of water to the live cell spectrum is considered to consist of : the symmetric ($v_1$) and asymmetric ($v_3$) stretching bands of the O-H groups at ~3450 cm$^{-1}$ and ~3600 cm$^{-1}$ respectively, in addition to the H-O-H ($v_2$) water bending band at ~1643 cm$^{-1}$ and water combination band at ~2120 cm$^{-1}$. With the exception of the combination band, all of these bands obscure other biological spectral features, such as the predominantly protein and carbohydrate derived amine and hydroxyl stretching bands moieties, as well any protein information from the amide I region.

### 4.1.3. Water correction methods for live cell spectra

There are various protocols for subtracting the water contribution from these live cell spectra. A straightforward approach that has been used by the Monash laboratory involves acquiring spectra through the medium that is adjacent to the interrogated cells and using that as a background reference spectrum (Tobin et al., 2010). This approach results in spectra where the amide I to amide II intensity ratio has been distorted due to the effect of the water deformation band, and a distorted amide A band due the symmetric ($\nu_1$) and asymmetric ($\nu_3$) O-H stretching groups at ~3450 cm$^{-1}$ and ~3600 cm$^{-1}$. Since this effect is confined to the amide A and I bands, these spectral regions (>3100 cm$^{-1}$ and 1715-1580 cm$^{-1}$) are removed from the data set before any further analysis. The method employed by our group is the equivalent to the first of two water correction procedures reported in the literature (Vaccari et al., 2012) as Methods 1 and 2. Method 1 involves calculating the ratio between the aqueous cell sample single beam spectrum and an aqueous medium reference spectrum, and is essentially the same method employed by the Monash laboratory. Its reliability is based on the assumption that any path length differences between the aqueous background and the sample solution are insignificant, and that any water contributions are equal for both the sample solution and the aqueous medium. We believe that the design of the wet chamber and the taking of background measurements very close to cell measurements satisfies the first assumption adequately. The second assumption will never be satisfied given the composition of live cells will not equate to aqueous media containing them, however, we would argue that removal of spectral regions contaminated by liquid water contributions from spectral analysis as discussed above, obviates this problem.

Method number 2, which is more time consuming, involves acquiring an air reference spectrum inside the microfluidic chamber prior to measuring the spectrum of the cell or aqueous medium, and then calculating the ratio between the latter two single channel

spectra against the reference spectrum. At this stage it is then possible to subtract the aqueous medium spectra and obtain what is hopefully a 'pure' cellular spectrum. However, this approach is not devoid of its own shortcomings. One major problem with this technique is the subjective nature of the process associated with choosing which scaling factor to use. Currently, many investigators employing this approach, flatten the largely biological band free spectral region between 1800-2500 $cm^{-1}$, which contains the 2125 $cm^{-1}$ water band (Rahmelow and Hubner, 1997; Vaccari et al., 2012; Venyaminov and Prendergast, 1997). No other water reference band can be used for this procedure as they fall in regions of the spectra which contain biological information. Since the water combination band is broad and weak, it can be difficult to judge when it has been effectively eliminated, hence making the process quite subjective. Whilst Mie scattering effects on the spectral baseline shape are greatly reduced in aqueous live samples, they can still limit the effectiveness of using the combination band for full water corrections.

Out of Methods 1 and 2, the latter was argued by Vaccari et al. (2012) to be the most advantageous as it enables the researcher to monitor the purity of the aqueous medium, achieved by calculating the ratio between the aqueous medium spectrum with the air spectrum. This is important because this medium acts as a repository for cellular metabolic by-products that may create a suboptimal growth environment for the cells. In the work reported here, the rapidity of the measurements (minutes) minimised the impact of this problem. It was also argued by Vaccari et al. (2012) that method 2 could correct for biological bands outside of spectral regions associated with water. Apart from being counter-intuitive, these claims are questionable because they are based on comparisons of live cell spectra with dried cell spectra, with the latter being used as a 'gold standard.' These claims need to be reconsidered in light of work by Whelan et al. (2011) that reports changes in DNA bands resulting from conformational changes in DNA between the

dehydrated and hydrated state (Figure 4.1), suggesting that at least for DNA bands, differences between live and dried cell spectra, are not due to inadequate correction of water absorbance, but result from the effects of the *in vivo* environment itself.

## 4.2. Materials and Methods

### 4.2.1. Stem cell biology

Undifferentiated HES3 human embryonic stem cells (hESCs) (Richards et al., 2002) were maintained on mouse embryonic fibroblasts (MEFs) in Knock Out Serum Replacer (KOSR)-based hESC medium (Costa et al., 2007), until they were ~80% confluent. The cells were then enzymatically harvested (Costa, 2008) and washed twice in PBS, before being re-suspended in APEL differentiation medium (Ng et al., 2008). To maintain the viability of the cells, they were placed on ice immediately after being harvested, and kept in these conditions for their transport to the Australian Synchrotron, and between spectral acquisitions.

### 4.2.2. Preparation of cells for spectroscopy

The sampling chamber that was used to interrogate live cells in our study was an example of a demountable liquid cell, in which the cells of interest were contained within physiological solution and placed between two infrared transparent windows by using a gasket (Marcsisin, 2011; Nasse et al., 2009; Tobin et al., 2010) (Figure 4.2). To mitigate against the metabolic consequences of the lack of flow within the chamber that we designed, a fresh aliquot of cells was placed into the sampling compartment every ten minutes to avoid apoptotic-induced biochemical changes influencing the spectra. Therefore, measurements were only made from a small number of cells at a time. Also, to

**Figure 4.1:** Average raw (top), and second derivative (bottom) spectra of single dried and living hES cells (n=122 dried cells; n=94 live cells). The important band differences between spectra acquired from live and dried cells are indicated with the approximate band positions given.
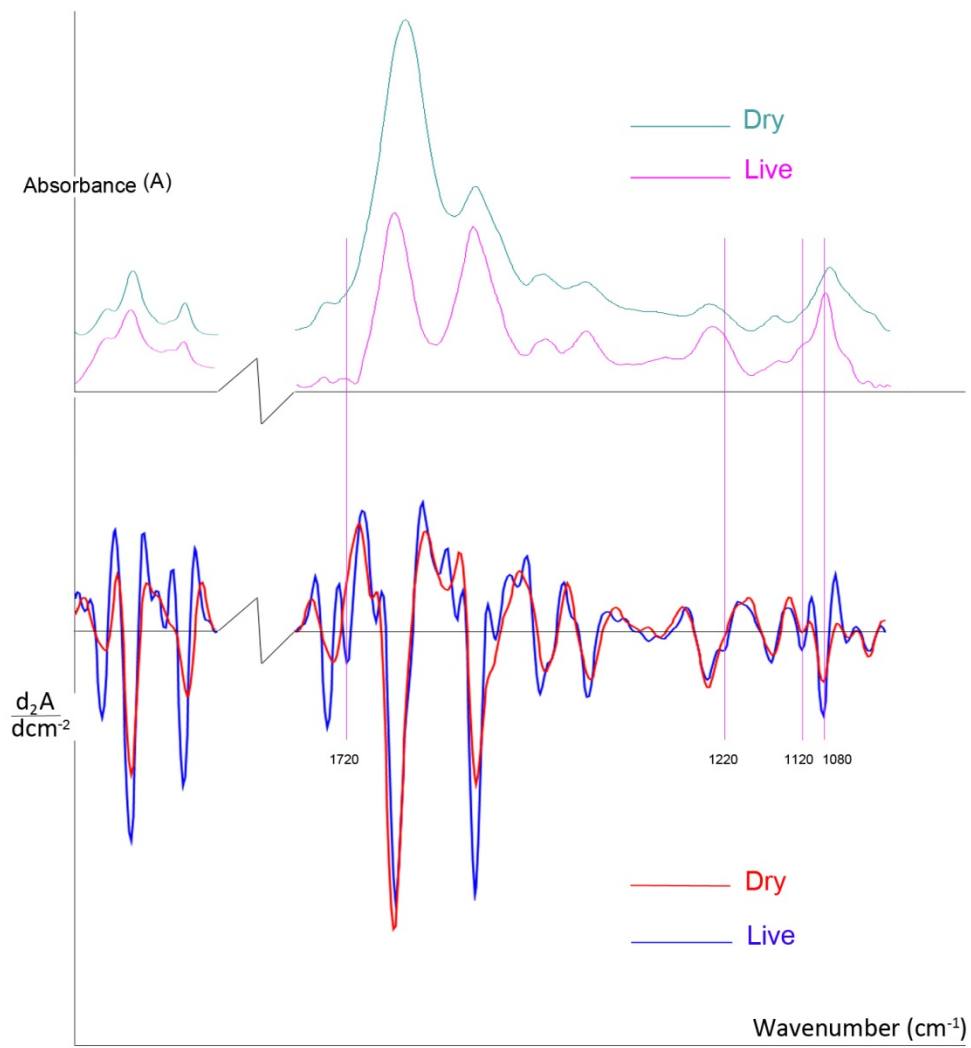
**Figure 4.2.** Acquisition of spectra from living human embryonic stem cells at the Australian Synchrotron. A 10 μl aliquot of cell suspension was placed at the centre of a $CaF_2$ IR that has been previously modified by the addition of a photo resist gasket allowing for the establishment of a path length of approximately 7.6 μm after the placement of the top window (Tobin et al., 2010). Cells were targeted with a 10 μm by 10 μm aperture using OPUS 6.5 control software .
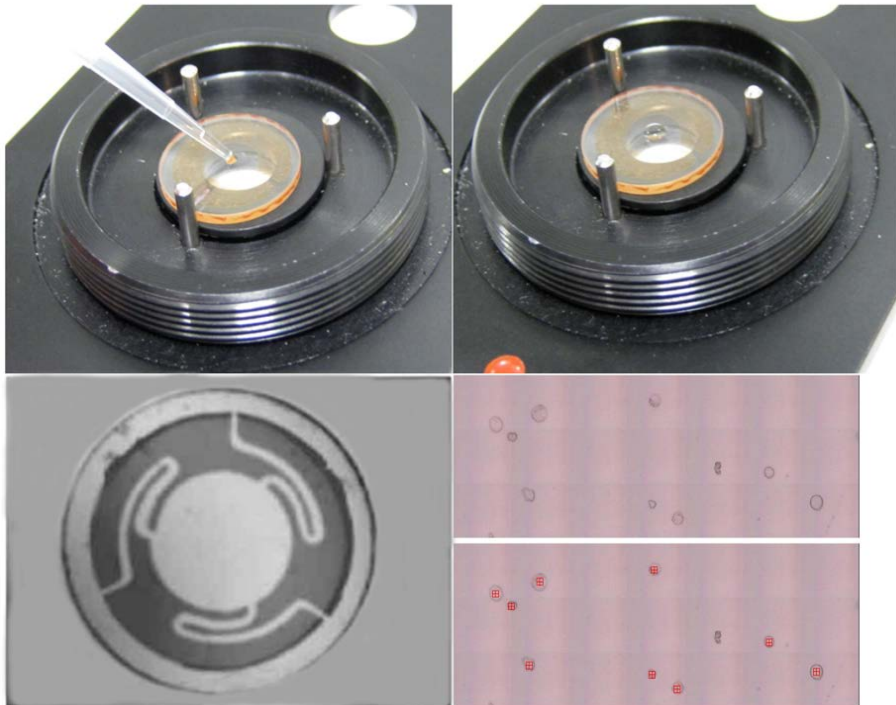
minimise the effects of path length variations, background measurements were taken close to the interrogated cells.

Spectra of single living hESCs were measured using an adapted Thermofisher compression cell which consisted of a pair of thin (1 mm thick lower window and 0.5 mm top window) 13 mm diameter $CaF_2$ windows, one of which was lithographically patterned with a polymeric spacer consisting of a photoresistive polymer (Microchemicals GmbH) fixed to the windows by UV photolithography, using a metal mask (MiniFAB Australia Pty. Ltd., Scoresby, Australia; Figure 4.2). The sampling chamber allowed for living cells to be studied with spectroscopic measurements acquired in transmission through a path length of 7.6 µm, after the placement of the top window. The three serpentine channels incorporated in the gasket spacer design allowed for the equilibration of capillary forces, enabling the aqueous medium to be retained for a sufficient duration of time (tens of minutes) whilst the cells were interrogated.

10 µl of cell suspension was placed on the patterned window which was previously rinsed in deionised water and air dried. An unpatterned $CaF_2$ window was then placed on top and the two windows were placed in the micro-compression cell which was fastened by the lid of the chamber. Cells (n=94) were targeted with a 10 µm by 10 µm focal plane aperture in the FTIR microscope using OPUS 6.5 control software at the IR microspectroscopy beamline, Australian Synchrotron. FTIR spectra were acquired with 128 co-added scans of the background culture medium, 128 co-added scans of the living cell at 8 $cm^{-1}$ spectral resolution. A Happ-Genzel apodisation function was used to truncate the interferogram, and a zero-filling factor of 2 was employed.

## 4.3. Results and Discussion

Comparisons of average spectra obtained from dried and live hESCs revealed clear band differences between the two conditions (Figure 4.1). These differences were first evident in the raw spectra, and became clearer upon their conversion to the second derivative, which allows for bands that are superimposed in the underivatised spectra to be more clearly identified. A band that was present in the live cell spectra but not in the dried cells occurred at ~1720 cm$^{-1}$, attributed to base pairing carbonyl vibrations from duplexed nucleic acids (Whelan et al., 2011). In the live spectra, the phosphate antisymmetric stretching vibrations of phosphodiester groups, was red shifted from ~1250 cm$^{-1}$ to 1220 cm$^{-1}$ in live cells compared to dried equivalents, whereas the ~1120 cm$^{-1}$ C-O stretching vibration from RNA, and ~1080 cm$^{-1}$ symmetric phosphodiester stretching vibration, were found to be of a higher intensity in live cells. These band changes corroborated observations made by Whelan et al. (2011) with human fibroblast and lymphocytes (Figures 4.3C and 4.3D), who investigated the effect of dehydration and rehydration of these and other eukaryotic cell types, with similar spectroscopic differences explained by a B to A DNA conformational change upon dehydration of cells.

The discovery that DNA bands could be assigned unambiguously in live cell spectra has far reaching diagnostic implications. For instance, live cell FTIR microspectroscopy could be employed to probe the hypermethylation induced-conformational changes in DNA that occurs in cancer and epigenetic events. In relation to our previously discussed work, these spectral markers could also be used to screen therapy-based differentiated stem cells for the purposes of isolation, identification, enrichment and purification, and may help to shed light on the environmentally elicited epigenetic effects that were observed in experiments presented in Chapter 3.

**Figure 4.3** Second derivative spectra of fully hydrated (blue) and dehydrated (red) (A) extracted chicken erythrocyte nuclei undergoing dehydration (ATR-FTIR), (B) extracted chicken erythrocyte nuclei undergoing rehydration (ATR-FTIR), (C) mammalian cancerous fibroblasts (L-929) undergoing rehydration (ATR-FTIR) and (D) mammalian lymphocytes undergoing rehydration (T-FTIR) (Whelan et al. 2011).
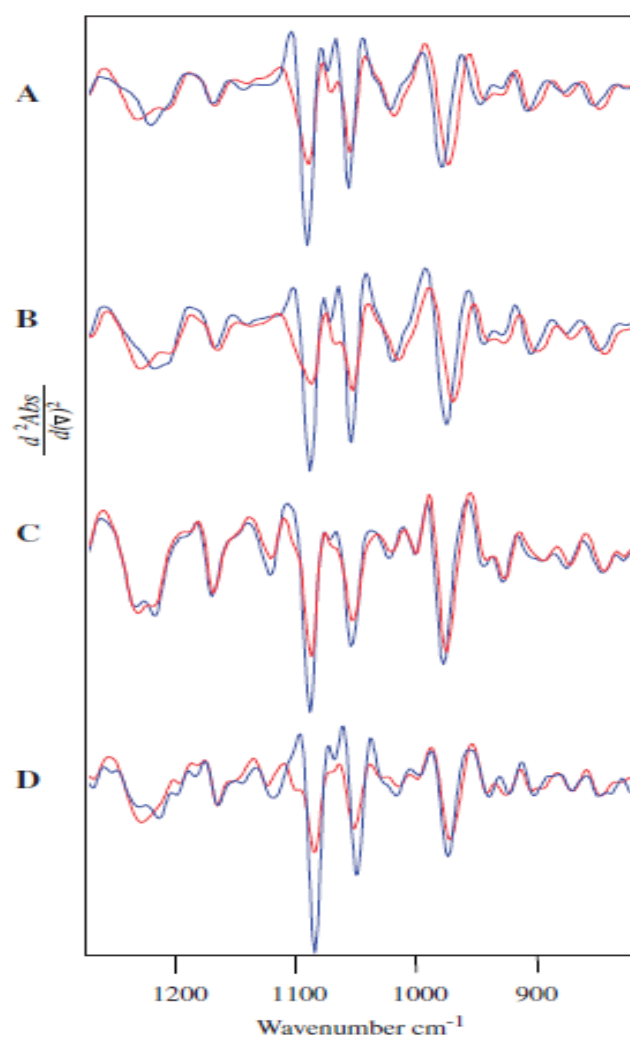
## 4.4. Future directions

Fixed cell studies employing spectroscopy-based methods have provided the foundations for our current understanding of macromolecular composition and structure, but unlike the real time experiments they don't allow us to investigate these samples under *in vivo* conditions. Live cell spectroscopy experiments have enabled the identification of fundamental molecular 'signature' differences between hydrated and dehydrated spectra. It is envisaged that with the improvement of microfluidic device design, advances in data processing procedures, and a greater understanding of the structure of intracellular water, the capabilities of real time synchrotron FTIR microspectroscopy will be fully realised.

CHAPTER **5**

General discussion

## 5.1. Introduction

This study described the utility of FTIR microspectroscopy, as a powerful optical modality for elucidating the unique macromolecular chemical profiles of human derived multipotent and pluripotent stem cells, and their lineage committed cell types. Chapter 2 described the employment of this method to investigate the similarities and/or differences between the spectroscopic 'signatures' of these stem cell variants and their progeny. Upon modelling the datasets using various chemometric analysis techniques, it was found that overall, each stem cell class had distinct spectral 'fingerprints', but these differences were not defined by whether the cells were hESC or hiPSC. Interestingly, both hESC and hiPSC lines were observed to undergo similar macromolecular chemical changes during their differentiation towards mesendodermal and ectodermal cell types, however, their progeny were not biochemically equivalent. To determine whether these phenotypic disparities could be attributed to the culture conditions, Chapter 3 described experiments employing this system to assess the effect of various tissue culturing platforms on the stem cells' IR spectral profiles. Indeed, the choice of growth matrix, growth medium, and dissociation method appeared to affect the spectra of cells cultured using these protocols, despite the flow cytometric data indicating they were similar. Interestingly, there was evidence of the cells appearing to 'remember' their original conditions, despite being transferred to another cellular environment. It was hypothesised that this phenomena could involve epigenetic processes, however this would require further investigation.

Collectively, the tools and protocols described in this thesis have contributed to efforts aimed at establishing FTIR microspectroscopy as a feasible optical modality to complement the existing repertoire of methods currently used to study and screen stem cell differentiation for clinical therapies.

## 5.2. Transflectance mode FTIR microspectroscopy

In the experiments that have been described in previous chapters, FTIR spectra of dried cell monolayers were acquired in transflectance mode. For these studies, samples were thinly deposited onto a highly infrared reflecting substrate, such as on a gold mirror or more recently, microscope slides coated with tin(iv) oxide ($SnO_2$) and silver (Ag) coating, known as low-e microscope slides. During the measurement, the mid-IR radiation effectively produced a 'double-pass' transmission spectrum as the beam passed through the sample and was reflected back by the mirror surface before being transmitted through the sample again.

The rationale for acquiring IR spectra in transflectance mode is largely due to the following: the low cost of reflective substrates compared to transmission windows; being glass slides of standard size, the substrates can be easily handled by histologists and are amenable to techniques designed to handle normal microscope slides; the optically transparent properties of these substrates also enables simultaneous light microscopy, post-staining and histology, and greater absorbance signals that can be obtained due to the double pass through the sample. All of this work was performed before the publication of recent work that indicates that there may be intrinsic problems with transflectance studies due to spectral contamination by optical effects such as RMieS ,and electric field standing wave (EFSW) produced signatures (Bassan et al., 2012; Filik et al., 2012). Nevertheless we believe that the results are still valid for the reasons argued below.

Recent work by Gardiner's group have identified spectral signatures in FTIR data ascribed to resonant Mie-scattering (RMieS). Modelling studies (Bassan et al., 2009; Bassan et al., 2012) have shown that it causes sinusoidal oscillations in the baseline, affecting band positions and their intensities. In biological samples, it is believed that major organelles such as the nucleus, or the cells themselves, give rise to the scattering artefacts observed in

biological spectra. The scattering has effects throughout the entire spectrum but is most prominent on the Amide I band shape and position, whereby a derivative-like distortion can be seen at its higher wavenumber end, called a 'dispersion artefact'. This signature can cause red or blue shifts of band positions in the spectra depending on the gradient of the Mie scattering efficiency curve. Although resonant Mie scattering occurs in both transflectance and transmission modes, there are claims that it is more conspicuous in transflectance measurements.

An algorithm has been developed which claims to address the problems of these signatures in biological spectra (Bassan et al., 2012), however its effectiveness has not been fully established and there has been no comprehensive empirical testing of this approach. A major criticism that could be directed against the RMieS algorithm is that it has some subjective aspects to it, such as the need to choose an initial reference spectrum and the number of processing iterations. In this study, we employed a simpler approach that has been advocated in the literature for the pre-processing of biological spectra with good classification of various biological and medical states (Kohler et al., 2009). This involved first calculating the second derivative spectrum, which has the effect of removing additive effects such as spectral offset, and flattening the spectral baseline to reveal bands that would normally be superimposed in the raw spectrum. Following this, the spectra were normalised accounting for multiplicative effects, using extended multiplicative signal correction (EMSC). This controls for changes in path length such as those arising from cell size differences. We then used Principal Component Analysis to observe any initial clustering in the data and removed outliers based on their residual variance and model leverage. Visual examination of the outlier spectra showed that this group contained spectra that were heavily affected by light scattering artefacts, as evidenced by their distorted baselines, and derivative features around the Amide I band. In summary, we

argue that this pre-processing approach constituted a straight forward method of eliminating spectra that had pronounced light scattering signatures, and correcting the remaining spectra for additive and multiplicative effects.

In addition to RMieS, another optical phenomena that has been recently discussed called the electric field standing wave (EFSW), claims to affect transflectance measurements (Bassan et al., 2013; Filik et al., 2012). Interference events between the incident and reflected light at metallic surfaces is believed to produce the EFSW, an optical signature which is argued to give rise to spectra that are distorted non linearly with respect to wavenumber, with absorbance changes that don't obey the Beer-Lambert law. This recent study purports that this standing wave may influence the classification outcomes, with the discriminations between spectral clusters being due to differences that have a physical origin i.e. cell thickness, rather than those of a macromolecular origin. Further, owing to the wavelength-dependent nature of this signature, it is claimed to occur along different parts of the IR spectrum and manifest in the loadings vectors. If found to be true, this would result in inaccurate interpretation of data, due to some of these loadings values being attributable to these unwanted physical effects rather than real biochemical differences. To test whether the EFSW had affected the data derived from our transflectance studies, we compared spectra acquired of hESC monolayers deposited on both transmission and transflectance substrates in experiments described in Chapter 3. The spectra of cells recorded from both studies, across all replicates, were found to be highly similar, as indicated by the colocalisation of spectral clusters from both classes in the PLS scores plots, and their poor PLS-DA classification values. Due to these results, in addition to the finding that thickness differences across the cell deposits were found to be well below 0.5 µm, the sample height difference within which the EFSW effect is thought to be detectable, we do not believe that this artefact was giving rise to any variances within our

spectra. These results suggested that because cyto-centrifugation results in very thin and consistent dried cellular monolayers, MirrIR slides may still be a valid option as an infrared measurement substrate, at least in the case of human embryonic stem cells.

## 5.3. Other spectroscopic modalities used for interrogating live cells

FTIR microspectroscopy using fluidic devices in transmission mode is not the only optical technique that can be used to screen live stem cells and their lineage committed progeny. Two other commonly used spectroscopic modalities, ATR-FTIR spectroscopy (attenuated total reflectance) and Raman microspectroscopy have been utilised in studies of stem cells and are described below.

ATR-FTIR spectroscopy relies on a property of light called total internal reflection which occurs at the interface between a high refractive internal reflection element (IRE), such as a diamond, and a low refractive index, infrared absorbing sample (Kazarian and Chan, 2013). The beam penetrates the sample as an evanescent wave, which interacts with the adsorbed sample and is limited to a given 'depth of penetration', that is dependent on the wavelength and angle of the incoming radiation, refractive index of the IRE, and the refractive indices of the interacting layers at the IRE interface. Typical depth of penetration values for ATR-FTIR measurements range between 0.2 µm and 5 µm. This small depth of penetration, achieved by the intimate contact of the sample with the ATR element, and thereby small effective path length, enables live cell studies to be carried out, as there will be no saturation of the detector.

ATR-FTIR microspectroscopy requires little to no sample preparation due to the sampling path not being thickness dependent, and because spectra derived from this technique are not subject to resonant Mie-scattering or EFSW signatures. Further, the coupling of ATR elements such as germanium to microscopes results in improved spatial resolutions since

their high refractive indices gives rise to high numerical apertures, and an increase in the outer ray angle of incidence as the light is refracted inwards.

However, a major disadvantage of this system is that due to the wavelength dependent, limited penetration depth, it is difficult to determine which parts of the investigated sample the signal originates from. Further, compared to transflection or transmission studies, the smaller effective path lengths results in spectra with lower signal to noise ratios. Although yet to be demonstrated, it is anticipated that ATR-FTIR microspectroscopy will eventually be carried out using synchrotron and/or quantum cascade laser sources of infrared radiation to improve this.

Raman spectroscopy of biological samples (Chan et al., 2009; Konorov et al., 2007; Schulze et al., 2010) is a two-photon inelastic light-scattering event, concerning an incident photon that is of much greater energy than the vibrational quantum energy, losing or gaining energy during the interaction with the molecular vibration. In Raman spectroscopy, the interaction between light and matter is an off-resonance condition involving the Raman polarisability of the molecule. This process yields a Raman spectrum, a molecular 'signature' of the interrogated sample, which provides us with biochemical information which is complementary to that attained via FTIR microspectroscopy.

Some positive points about this technique include its ability to operate at sub-micrometer spatial resolutions, and the weak Raman scattering properties of water. Whereas some negative points about the method include its slower spectral acquisition times when measuring macromolecular components in intact cells, and lower sensitivity, although the latter can be improved with the use of surface enhanced techniques. Moreover, it has been shown to compromise the structural and thereby biochemical integrity of the samples if the laser power is not strictly controlled and monitored because high laser power can give rise to heating and or/ photoproduct generation (Puppels et al., 1991).

In contrast to ATR-FTIR microspectroscopy, which has been applied to a very limited number of stem cell studies (Aksoy et al., 2012), Raman microspectroscopy has been extensively utilised to interrogate a variety of stem cell types that include embryonic stem cells (Chan et al., 2009; Konorov et al., 2007; Notingher et al., 2004), induced pluripotent stem cells (Tan et al., 2012), and mesenchymal stem cells (Azrad et al., 2006; Kim et al., 2008).

Interestingly, in Raman microspectroscopy studies performed by Tan et al. (2012), similar comparisons were made between hESC and hIPS cells to those performed by our laboratory using FTIR microspectroscopy. The group found the two stem cell variants to share highly similar spectral signatures, and that both these classes were spectroscopically distinct from hES cells that had undergone 20 days of non-specific differentiation. In our work, we could not discern consistent spectral differences when comparing FTIR spectra between these two classes, but instead found significant biochemical variation within these classes. By contrast, Tan et al. (2012) only compared one hESC line and one hIPSC line, and therefore any intraclass spectral variations could not be properly assessed. Their finding that spectra from differentiated hES cells and undifferentiated hIPS cells were distinct is unremarkable in comparison with our study, in which progeny cells derived from both stem cell classes were compared and found to be spectroscopically distinguishable, despite expressing the same immunohistochemical markers.

Despite their shortcomings, FTIR microspectroscopy, ATR-FTIR spectroscopy, and Raman microspectroscopy each have their merits as robust analytical tools for living studies of pluripotent and multipotent stem cells.

## 5.4. Conclusion

Stem cells, like many biological specimens, are inherently complex, and no single research tool is all-encompassing enough to discern their complete biological, chemical and physical information. Depending on the parameter under investigation, researchers will employ a wealth of molecular biology techniques that include flow cytometry, in vivo and in vitro assays and RT-PCR to attain data related to cell surface marker expression, functionality, and gene expression respectively.

FTIR microspectroscopy offers a new experimental approach to understanding stem cell biology. It provides a holistic measurement of macromolecular composition such that a signature representing internal cellular phenotype is obtained. This contributes information that is complimentary to that acquired by conventional genetic and immunohistochemical modalities.

However, before the benefits of spectroscopy in stem cell research can be fully realised, several important developments must be made. For example, advances in instrumentation to speed up data acquisition and improve signal to noise ratios would help to significantly further the field. The most recent progress in this area, as earlier described, have been the coupling of FPA detectors to synchrotron light sources, and the use of quantum cascade laser sources of infrared radiation. In addition to instrumental development, there is a critical and growing need for a focus on metrology to further the diagnostic potential of FTIR microspectroscopy. Standardising measurement parameters such as spectral resolution, the number of co-added scans, and the number of binned pixels will allow for results between studies to be more readily comparable. In this context, instrument intercomparisons are cogent to gauge the effects of instrument type on spectral classification. Also imperative is a better understanding of the effects of confounding

variables on spectroscopic data and ways to minimise these or to correct for the resulting spectral signatures by data pre-processing.

Once these developments are made, we anticipate the successful employment of FTIR microspectroscopy in future stem cell research due to its ability to provide phenotypic information, in a rapidly, non-invasively, and objectively.

# BIBLIOGRAPHY

1.  Abeyta, M.J., Clark, A.T., Rodriguez, R.T., Bodnar, M.S., Pera, R.A.R., and Firpo, M.T. (2004). Unique gene expression signatures of independently-derived human embryonic stem cell lines. Human Molecular Genetics *13*, 601-608.

2.  Agatonovic-Kustrin, S., and Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modelling and its application in pharmaceutical research. Journal of Pharmaceutical and Biomedical Analysis *22*, 717-727.

3.  Akopian, V.; Andrews, P.; Beil, S.; Benvenisty, N.; Brehm, J.; Christie, M.; Ford, A.; Fox, V.; Gokhale, P.; Healy, L.; Holm, F.; Hovatta, O.; Knowles, B.; Ludwig, T.; McKay, R.; Miyazaki, T.; Nakatsuji, N.; Oh, S.; Pera, M.; Rossant, J.; Stacey, G.; Suemori, H. (2010). Comparison of defined culture systems for feeder cell free propagation of human embryonic stem cells. In Vitro Cellular *&* Developmental Biology. *46*, 247-258

4.  Aksoy, C., Guliyev, A., Kilic, E., Uckan, D., and Severcan, F. (2012). Bone marrow mesenchymal stem cells in patients with beta thalassemia major: molecular analysis with Attenuated Total Reflection-Fourier Transform Infrared spectroscopy study as a novel method. Stem Cells and Development *21*, 2000-2011.

5.  Allegrucci, C.; Wu, Y.-Z.; Thurston, A.; Denning, C. N.; Priddle, H.; Mummery, C. L.; Ward-van Oostwaard, D.; Andrews, P. W.; Stojkovic, M.; Smith, N.; Parkin, T.; Jones, M. E.; Warren, G.; Yu, L.; Brena, R. M.; Plass, C.; Young, L. E. (2007). Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome. Human Molecular Genetics. *16*, 1253-1268.

6.  Ami, D., Neri, T., Natalello, A., Mereghetti, P., Doglia, S.M., Zanoni, M., Zuccotti, M., Garagna, S., and Redi, C.A. (2008). Embryonic stem cell

differentiation studied by FT-IR spectroscopy. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1783,* 98-106.

7.   Amit, M.; Margulets, V.; Segev, H.; Shariki, K.; Laevsky, I.; Coleman, R.; Itskovitz-Eldor, J. (2003). Human feeder layers for human embryonic stem cells. Biology of Reproduction. *68*, 2150-2156.

8.   Azrad, E., Zahor, D., Vago, R., Nevo, Z., Doron, R., Robinson, D., Gheber, L.A., Rosenwaks, S., and Bar, I. (2006). Probing the effect of an extract of elk velvet antler powder on mesenchymal stem cells using Raman microspectroscopy: enhanced differentiation toward osteogenic fate. Journal of Raman Spectroscopy *37*, 480-486.

9.   Bassan, P., Byrne, H.J., Bonnier, F., Lee, J., Dumas, P., and Gardner, P. (2009). Resonant Mie scattering in infrared spectroscopy of biological materials– understanding the 'dispersion artefact'. Analyst *134*, 1586-1593.

10.  Bassan, P., Sachdeva, A., Kohler, A., Hughes, C., Henderson, A., Boyle, J., Shanks, J.H., Brown, M., Clarke, N.W., and Gardner, P. (2012). FTIR microscopy of biological cells and tissue: data analysis using resonant Mie scattering (RMieS) EMSC algorithm. Analyst *137*, 1370-1377.

11.  Bassan, P.; Lee, J.; Sachdeva, A.; Pissardini, J.; Dorling, K. M.; Fletcher, J. S.; Henderson, A.; Gardner, P. (2013).The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications. Analyst *138*, 144-157.

12.  Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439-452.

13. Braam, S. R.; Zeinstra, L.; Litjens, S.; Ward-van Oostwaard, D.; van den Brink, S.; van Laake, L.; Lebrin, F.; Kats, P.; Hochstenbach, R.; Passier, R.; Sonnenberg, A.; Mummery, C. L. (2008). Recombinant vitronectin is a functionally defined substrate that supports human embryonic stem cell self-renewal via αVβ5 integrin. Stem Cells *26,* 2257-2265.

14. Brandenberger, R.; Wei, H.; Zhang, S.; Lei, S.; Murage, J.; Fisk, G. J.; Li, Y.; Xu, C.; Fang, R.; Guegler, K.; Rao, M. S.; Mandalam, R.; Lebkowski, J.; Stanton, L. W. (2004). Transcriptome characterization elucidates signalling networks that control human ES cell growth and differentiation. Nature Biotechnology *22*, 707-716.

15. Byler, D.M., and Susi, H. (1986). Examination of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers *25*, 469-487.

16. Cao, J., Ng, E., McNaughton, D., Stanley, E., Elefanty, A., Tobin, M.J., and Heraud, P. (2013a). Fourier transform infrared microspectroscopy reveals that tissue culture conditions affect the macromolecular phenotype of human embryonic stem cells. Analyst. DOI**:** 10.1039/C3AN00321C

17. Cao, J., Ng, E.S., McNaughton, D., Stanley, E.G., Elefanty, A.G., Tobin, M.J., and Heraud, P. (2013b). Fourier transform infrared microspectroscopy reveals unique phenotypes for human embryonic and induced pluripotent stem cell lines and their progeny. Journal of Biophotonics. DOI:10.1002/JBIO.201200217

18. Carpenter, M. K.; Rosler, E.; Rao, M. S. (2003) Characterization and differentiation of human embryonic stem cells. Cloning Stem Cells *5*, 79-88.

19. Chan, J.W., Lieu, D.K., Huser, T., and Li, R.A. (2009). Label-free separation of human embryonic stem cells (hESCs) and their cardiac derivatives using Raman spectroscopy. Analytical Chemistry *81*, 1324.

20. Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., et al. (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell *5*, 111-123.

21. Chonanant, C., Jearanaikoon, N., Leelayuwat, C., Limpaiboon, T., Tobin, M.J., Jearanaikoon, P., and Heraud, P. (2011). Characterisation of chondrogenic differentiation of human mesenchymal stem cells using synchrotron FTIR microspectroscopy. Analyst *136*, 2542-2551.

22. Costa, M. S., Koula ; Hatzistavrou, T.; Elefanty, AG.; and Stanley, EG, Expansion of human embryonic stem cells in vitro (2008). In Current Protocols in Stem Cell Biology, 1C.1.1-1C.1.7

23. Costa, M., Dottori, M., Ng, E., Hawes, S.M., Sourris, K., Jamshidi, P., Pera, M.F., Elefanty, A.G., and Stanley, E.G. (2005). The hESC line Envy expresses high levels of GFP in all differentiated progeny. Nature Methods *2*, 259-260.

24. Davis, R. P.; Ng, E. S.; Costa, M.; Mossman, A. K.; Sourris, K.; Elefanty, A. G.; Stanley, E. G. (2008). Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak–like cells and enables isolation of primitive hematopoietic precursors. Blood *111*, 1876-1884.

25. Downes, A., Mouras, R., and Elfick, A. (2010). Optical spectroscopy for non-invasive monitoring of stem cell differentiation. Journal of Biomedicine and Biotechnology.

26. Downing, A., Forsyth, N.R., Hampson, K., Kay, A., McWhir, J., and Talbot, R. (2008). Transcriptome alterations due to physiological normoxic (2% $O_2$) culture of human embryonic stem cells. Regenerative Medicine *3*, 817-833

27. Elliott, D.A., Braam, S.R., Koutsis, K., Ng, E.S., Jenny, R., Lagerqvist, E.L., Biben, C., Hatzistavrou, T., Hirst, C.E., Yu, Q.C., et al. (2011). *NKX 2-5$^{eGFP/w}$* hESCs for isolation of human cardiac progenitors and cardiomyocytes. Nature Methods *8*, 1037-1040.

28. Esbensen, K.H., Guyot, D., Westad, F., and Houmøller, L.P. (2002). Multivariate data analysis-in practice: An introduction to multivariate data analysis and experimental design (Multivariate Data Analysis).

29. Filik, J.; Frogley, M. D.; Pijanka, J. K.; Wehbe, K.; Cinque, G. (2012). Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials. Analyst *137*, 853-861.

30. Geladi, P. (1988). Notes on the history and nature of partial least squares (PLS) modelling. Journal of Chemometrics *2*, 231-246.

31. Geladi, P., and Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. Analytica Chimica Acta *185*, 1-17.

32. German, M.J., Pollock, H.M., Zhao, B., Tobin, M.J., Hammiche, A., Bentley, A., Cooper, L.J., Martin, F.L., and Fullwood, N.J. (2006). Characterization of putative stem cell populations in the cornea using synchrotron infrared microspectroscopy. Investigative Ophthalmology & Visual Science *47*, 2417-2421.

33. Ghosh, Z., Wilson, K.D., Wu, Y., Hu, S., Quertermous, T., and Wu, J.C. (2010). Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PloS ONE *5*, e8975.

34. Ginis, I.; Luo, Y.; Miura, T.; Thies, S.; Brandenberger, R.; Gerecht-Nir, S.; Amit, M.; Hoke, A.; Carpenter, M. K.; Itskovitz-Eldor, J.; Rao, M. S. (2004). Differences between human and mouse embryonic stem cells. Developmental Biology *269*, 360-380.

35. Griffith, L.G., and Naughton, G. (2002). Tissue engineering--current challenges and expanding opportunities. Science *295*, 1009-1014.

36. Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249-257.

37. Haibach, F., Erlich, A., & Deutsch, E. (2011). Mid-infrared absorption spectroscopy using quantum cascade lasers. SPIE Defense, Security, and Sensing, 803208-803208.

38. Heraud, P., and Tobin, M.J. (2009). The emergence of biospectroscopy in stem cell research. Stem Cell Research *3*, 12-14.

39. Heraud, P., Ng, E.S., Caine, S., Yu, Q.C., Hirst, C., Mayberry, R., Bruce, A., Wood, B.R., McNaughton, D., and Stanley, E.G. (2010). Fourier transform infrared microspectroscopy identifies early lineage commitment in differentiating human embryonic stem cells. Stem Cell Research *4*, 140-147.

40. Heraud, P.; Tobin, M. J. (2009). The emergence of biospectroscopy in stem cell research. Stem Cell Research *3*, 12-14.

41. Hoffman, L.M., and Carpenter, M.K. (2005). Characterization and culture of human embryonic stem cells. Nature Biotechnology *23*, 699-708.

42. Hovatta, O. O.; Mikkola, M. M.; Gertow, K. K.; Strömberg, A.-M. A. M.; Inzunza, J. J.; Hreinsson, J. J.; Rozell, B. B.; Blennow, E. E.; Andäng, M. M.; Ahrlund-Richter, L. L. (2003). A culture system using human foreskin fibroblasts as feeder cells allows production of human embryonic stem cells. Human Reproduction *18*, 1404-1409.

43. Hughes, C., Liew, M., Sachdeva, A., Bassan, P., Dumas, P., Hart, C.A., Brown, M.D., Clarke, N.W., and Gardner, P. (2010). SR-FTIR spectroscopy of renal epithelial carcinoma side population cells displaying stem cell-like characteristics. Analyst *135*, 3133-3141.

44. Jackson, M., and Mantsch, H.H. (1995). The use and misuse of FTIR spectroscopy in the determination of protein structure. Critical Reviews in Biochemistry and Molecular Biology *30*, 95-120.

45. Jasnin, M., Moulin, M., Haertlein, M., Zaccai, G., and Tehei, M. (2008). Down to atomic-scale intracellular water dynamics. EMBO reports *9*, 543-547.

46. Kansiz, M., Christian Schuster, K., McNaughton, D., and Lendl, B. (2005). Sequential injection/mid-infrared spectroscopic analysis of an acetone-butanol-ethanol fermentation: analyte cross-correlation effects. Spectroscopy Letters *38*, 677-702.

47. Kazarian, S.G., and Chan, K.L.A. (2013). ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems. Analyst *138*, 1940-1951.

48. Kelly, J.G., Nakamura, T., Kinoshita, S., Fullwood, N.J., and Martin, F.L. (2010). Evidence for a stem-cell lineage in corneal squamous cell carcinoma using synchrotron-based Fourier-transform infrared microspectroscopy and multivariate analysis. Analyst *135*, 3120-3125.

49. Kim, B.S., Lee, C.C.I., Christensen, J.E., Huser, T.R., Chan, J.W., and Tarantal, A.F. (2008). Growth, differentiation, and biochemical signatures of rhesus monkey mesenchymal stem cells. Stem Cells and Development *17*, 185-198.

50. Kohler, A., Böcker, U., Warringer, J., Blomberg, A., Omholt, S., Stark, E., and Martens, H. (2009). Reducing inter-replicate variation in Fourier transform

infrared spectroscopy by extended multiplicative signal correction. Applied Spectroscopy *63*, 296-305.

51. Konorov, S.O., Glover, C.H., Piret, J.M., Bryan, J., Schulze, H.G., Blades, M.W., and Turner, R.F. (2007). In situ analysis of living embryonic stem cells by coherent anti-Stokes Raman microscopy. Analytical Chemistry *79*, 7221-7225.

52. Krafft, C., Salzer, R., Seitz, S., Ern, C., and Schieker, M. (2007). Differentiation of individual human mesenchymal stem cells probed by FTIR microscopic imaging. Analyst *132*, 647-653.

53. Langer, R. (2007). Tissue engineering: perspectives, challenges, and future directions. Tissue Engineering *13*, 1-2.

54. Lewis, E.N., Treado, P.J., Reeder, R.C., Story, G.M., Dowrey, A.E., Marcott, C., and Levin, I.W. (1995). Fourier transform spectroscopic imaging using an infrared focal-plane array detector. Analytical Chemistry *67*, 3377-3381.

55. Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68-73.

56. Liu, J., Sumer, H., Leung, J., Upton, K., Dottori, M., Pébay, A., & Verma, P. J. (2011). Late passage human fibroblasts induced to pluripotency are capable of directed neuronal differentiation. Cell Transplantation *20*, 193-203.

57. Maitra, A., Arking, D.E., Shivapurkar, N., Ikeda, M., Stastny, V., Kassauei, K., Sui, G., Cutler, D.J., Liu, Y., Brimble, S.N., et al. (2005). Genomic alterations in cultured human embryonic stem cells. Nature Genetics *37*, 1099-1103.

58. Mantsch, H.H., Yang, P.W., and Casal, H.L. (1986). Infrared spectrometry of living systems: current trends and perspectives. Journal of Molecular Structure *141*, 237-242.

59. Marchetto, M.C.N., Yeo, G.W., Kainohana, O., Marsala, M., Gage, F.H., and Muotri, A.R. (2009). Transcriptional signature and memory retention of human-induced pluripotent stem cells. PloS ONE *4*, e7076.

60. Marcsisin, E.J. (2011). Infrared spectroscopy to monitor drug response of individual live cells.

61. Martens, H., and Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new pre-processing methods for near infrared spectroscopy. Journal of Pharmaceutical and Biomedical Analysis *9*, 625-635.

62. Martens, H., Nielsen, J.P., and Engelsen, S.B. (2003). Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. Analytical Chemistry *75*, 394-404.

63. Martens, H.; Nielsen, J. P.; Engelsen, S. B. (2003). Light scattering and light absorbance separated by Extended Multiplicative Signal Correction. Application to Near-Infrared transmission analysis of powder mixtures. Analytical Chemistry *75*, 394-404.

64. Miller, L.M. (2009). Biomedical applications of infrared microspectroscopy using synchrotron radiation. Infrared and Raman Spectroscopic Imaging.

65. Miljković, M. ; Bird, B.; Lenau, K.; Mazur, A.I.; Diem, M. (2013). Spectral cytopathology: new aspects of data collection, manipulation and confounding effects. Analyst, 138, 3975-3982

66. Mitalipova, M. M.; Rao, R. R.; Hoyer, D. M.; Johnson, J. A.; Meisner, L. F.; Jones, K. L.; Dalton, S.; Stice, S. L. (2005). Preserving the genetic integrity of human embryonic stem cells. Nature. Biotechnology. *23*, 19-20.

67. Mohlenhoff, B., Romeo, M., Diem, M., and Wood, B.R. (2005). Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy. Biophysical Journal *88*, 3635-3640.

68. Nakamura, T., Kelly, J.G., Trevisan, J., Cooper, L.J., Bentley, A.J., Carmichael, P.L., Scott, A.D., Cotte, M., Susini, J., and Martin-Hirsch, P.L, Kinoshita, S, Fullwood, N.J and Martin, F.L (2010). Microspectroscopy of spectral biomarkers associated with human corneal stem cells. Molecular Vision *16*, 359.

69. Narsinh, K.H., Plews, J., and Wu, J.C. (2011). Comparison of human induced pluripotent and embryonic stem cells: fraternal or identical twins? Molecular Therapy *19*, 635-638.

70. Nasse, M., Ratti, S., Giordano, M., and Hirschmugl, C. (2009). Demountable liquid/flow cell for in vivo infrared microspectroscopy of biological specimens. Applied Spectroscopy *63*, 1181-1186.

71. Nasse, M.J., Walsh, M.J., Mattson, E.C., Reininger, R., Kajdacsy-Balla, A., Macias, V., Bhargava, R., and Hirschmugl, C.J. (2011). High-resolution Fourier-transform infrared chemical imaging with multiple synchrotron beams. Nature Methods *8*, 413-416.

72. Ng, E.S., Davis, R., Stanley, E.G., and Elefanty, A.G. (2008). A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nature Protocols *3*, 768-776.

73. Ng, E.S., Davis, R.P., Azzola, L., Stanley, E.G., and Elefanty, A.G. (2005). Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. Blood *106*, 1601-1603.

74. Ng, E.S., Davis, R.P., Hatzistavrou, T., Stanley, E.G., and Elefanty, A.G. (2007). Directed differentiation of human embryonic stem cells as spin embryoid bodies and a description of the hematopoietic blast colony forming assay. In Current Protocols in Stem Cell Biology (John Wiley & Sons, Inc.).

75. Notingher, I., Bisson, I., Bishop, A.E., Randle, W.L., Polak, J.M., and Hench, L.L. (2004). In situ spectral monitoring of mRNA translation in embryonic stem cells during differentiation in vitro. Analytical Chemistry *76*, 3185-3193.

76. Ó Faoláin, E., Hunter, M.B., Byrne, J.M., Kelehan, P., McNamara, M., Byrne, H.J., and Lyng, F.M. (2005). A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. Vibrational Spectroscopy *38*, 121-127.

77. Phanstiel, D.H., Brumbaugh, J., Wenger, C.D., Tian, S., Probasco, M.D., Bailey, D.J., Swaney, D.L., Tervo, M.A., Bolin, J.M., Ruotti, V., et al. (2011). Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nature Methods *8*, 821-827.

78. Pijanka, J.K., Kumar, D., Dale, T., Yousef, I., Parkes, G., Untereiner, V., Yang, Y., Dumas, P., Collins, D., Manfait, M., et al. (2010). Vibrational spectroscopy differentiates between multipotent and pluripotent stem cells. Analyst *135*, 3126-3132.

79. Puppels, G.J., Olminkhof, J.H.F., Segers-Nolten, G.M.J., Otto, C., de Mul, F.F.M., and Greve, J. (1991). Laser irradiation and Raman spectroscopy of single

living cells and chromosomes: Sample degradation occurs with 514.5 nm but not with 660 nm laser light. Experimental Cell Research *195*, 361-367.

80. Rahmelow, K., and Hubner, W. (1997). Infrared spectroscopy in aqueous solution: Difficulties and accuracy of water subtraction. Applied Spectroscopy *51*, 160-170.

81. Richards, M., Fong, C.-Y., Chan, W.-K., Wong, P.-C., and Bongso, A. (2002). Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells. Nature Biotechnology *20*, 933-936.

82. Richards, M., Fong, C.-Y., Chan, W.-K., Wong, P.-C., and Bongso, A. (2002). Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells. Nature Biotechnology *20*, 933-936.

83. Rosler, E. S.; Fisk, G. J.; Ares, X.; Irving, J.; Miura, T.; Rao, M. S.; Carpenter, M. K. (2004). Long-term culture of human embryonic stem cells in feeder-free conditions. Developmental. Dynamics. *229* 259-274.

84. Sandt, C., Féraud, O., Oudrhiri, N., Bonnet, M.L., Meunier, M.C., Valogne, Y., Bertrand, A., Raphaël, M., Griscelli, F., Turhan, A.G., et al. (2012). Identification of spectral modifications occurring during reprogramming of somatic cells. PloS ONE *7*, e30743.

85. Sato, N.; Sanjuan, I. M.; Heke, M.; Uchida, M.; Naef, F.; Brivanlou, A. H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. Developmental Biology, *260*, 404-413.

86. Savitzky, A., and Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry *36*, 1627-1639.

87. Schulze, H.G., Konorov, S.O., Caron, N.J., Piret, J.M., Blades, M.W., and Turner, R.F. (2010). Assessing differentiation status of human embryonic stem cells

noninvasively using Raman microspectroscopy. Analytical Chemistry *82*, 5020-5027.

88. Sjögren-Jansson, E., Zetterström, M., Moya, K., Lindqvist, J., Strehl, R., and Eriksson, P.S. (2005). Large-scale propagation of four undifferentiated human embryonic stem cell lines in a feeder-free culture system. Developmental Dynamics *233*, 1304-1314.

89. Sperger, J. M.; Chen, X.; Draper, J. S.; Antosiewicz, J. E.; Chon, C. H.; Jones, S. B.; Brooks, J. D.; Andrews, P. W.; Brown, P. O.; Thomson, J. A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proceedings of the National Academy of Sciences. U.S.A, *100*, 13350-13355.

90. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

91. Tan, Y., Konorov, S.O., Schulze, H.G., Piret, J.M., Blades, M.W., and Turner, R.F.B. (2012). Comparative study using Raman microspectroscopy reveals spectral signatures of human induced pluripotent cells more closely resemble those from human embryonic stem cells than those from differentiated cells. Analyst *137*, 4509-4515.

92. Tanthanuch, W., Thumanu, K., Lorthongpanich, C., Parnpai, R., and Heraud, P. (2010). Neural differentiation of mouse embryonic stem cells studied by FTIR spectroscopy. Journal of Molecular Structure *967*, 189-195.

93. Taupin, P. (2006). Derivation of embryonic stem cells for cellular therapy: challenges and new strategies. Medical Science Monitor *12*.

94. Tecirlioglu, R. T.; Nguyen, L.; Koh, K.; Trounson, A. O.; Michalska, A. E. (2010). Derivation and maintenance of human embryonic stem cell line on human adult skin fibroblast feeder cells in serum replacement medium. In Vitro Cellular & Developmental Biology-Animal *46*, 231-235.

95. Thomson, J. A.; Itskovitz-Eldor, J.; Shapiro, S. S.; Waknitz, M. A.; Swiergiel, J. J.; Marshall, V. S.; Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. Science, *282*, 1145-1147.

96. Thumanu, K., Tanthanuch, W., Ye, D., Sangmalee, A., Lorthongpanich, C., Parnpai, R., and Heraud, P. (2011). Spectroscopic signature of mouse embryonic stem cell--derived hepatocytes using synchrotron Fourier transform infrared microspectroscopy. Journal of Biomedical Optics *16*, 057005-057005.

97. Tobin, M.J., Puskar, L., Barber, R.L., Harvey, E.C., Heraud, P., Wood, B.R., Bambery, K.R., Dillon, C.T., and Munro, K.L. (2010). FTIR spectroscopy of single live cells in aqueous media by synchrotron IR microscopy using microfabricated sample holders. Vibrational Spectroscopy *53*, 34-38.

98. Trevisan, J., Angelov, P.P., Carmichael, P.L., Scott, A.D., and Martin, F.L. (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. Analyst *137,* 3202-3215.

99. Vaccari, L., Birarda, G., Grenci, G., Pacor, S., and Businaro, L. (2012). Synchrotron radiation infrared microspectroscopy of single living cells in microfluidic devices: advantages, disadvantages and future perspectives. Journal of Physics: Conference Series *359*, 12007-12016.

100. Venyaminov, S.Y., and Prendergast, F.G. (1997). Water ($H_2O$ and $D_2O$) Molar absorptivity in the 1000–4000 $cm^{-1}$ range and quantitative infrared spectroscopy of aqueous solutions. Analytical Biochemistry *248*, 234-245.

101. Wagers, A.J., and Weissman, I.L. (2004). Plasticity of Adult Stem Cells. Cell *116*, 639-648.

102. Walsh, M.J., Fellous, T.G., Hammiche, A., Lin, W.-R., Fullwood, N.J., Grude, O., Bahrami, F., Nicholson, J.M., Cotte, M., Susini, J., et al. (2008). Fourier Transform infrared microspectroscopy identifies symmetric $PO_2$- modifications as a marker of the putative stem cell region of human intestinal crypts. Stem Cells *26*, 108-118.

103. Walsh, M.J., Hammiche, A., Fellous, T.G., Nicholson, J.M., Cotte, M., Susini, J., Fullwood, N.J., Martin-Hirsch, P.L., Alison, M.R., and Martin, F.L. (2009). Tracking the cell hierarchy in the human intestine using biochemical signatures derived by mid-infrared microspectroscopy. Stem Cell Research *3*, 15-27.

104. Weida, M.J., and Yee, B. (2011). Quantum cascade laser-based replacement for FTIR microscopy. 79021C-79021C.

105. Whelan, D.R., Bambery, K.R., Heraud, P., Tobin, M.J., Diem, M., McNaughton, D., and Wood, B.R. (2011). Monitoring the reversible B to A-like transition of DNA in eukaryotic cells using Fourier transform infrared spectroscopy. Nucleic Acids Research *39*, 5439-5448.

106. Wold, S., & Sjostrom, M. (1977). SIMCA: a method for analyzing chemical data in terms of similarity and analogy. Chemometrics: theory and application, *52*, 243-282.

107. Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. Chemometrics and Intelligent Laboratory Systems *2*, 37-52.

108. Yanes, O., Clark, J., Wong, D., Patti, G., Sánchez-ruiz, A., Benton, H., Trauger, S., Desponts, C., Ding, S., and Siuzdak, G. (2010). Metabolic oxidation regulates embryonic stem cell differentiation. Nature Chemical Biology *6*, 411-417.

109. Ye, D., Tanthanuch, W., Thumanu, K., Sangmalee, A., Parnpai, R., and Heraud, P. (2012). Discrimination of functional hepatocytes derived from mesenchymal stem cells using FTIR microspectroscopy. Analyst *137*, 4774-4784.

110. Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797-801.

111. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

112. Yu, Q.C., Hirst, C.E., Costa, M., Ng, E.S., Schiesser, J.V., Gertow, K., Stanley, E.G., and Elefanty, A.G. (2012). Apelin promotes hematopoiesis from human embryonic stem cells. Blood *119*, 6243-6254.

113. Zelig, U., Dror, Z., Iskovich, S., Zwielly, A., Ben-Harush, M., Nathan, I., Mordechai, S., and Kapelushnik, J. (2010). Biochemical analysis and quantification of hematopoietic stem cells by infrared spectroscopy. Journal of Biomedical Optics *15*, 037008-037008.

APPENDICES

FULL ARTICLE

# Fourier transform infrared microspectroscopy reveals unique phenotypes for human embryonic and induced pluripotent stem cell lines and their progeny

*Julie Cao*[1,2], *Elizabeth S. Ng*[1,4], *Don McNaughton*[2], *Edouard G. Stanley*[1,4], *Andrew G. Elefanty*[1,4], *Mark J. Tobin*[3], *and Philip Heraud**[*,1,2]

[1] Monash Immunology and Stem Cell Laboratories, Monash University, Building 75, STRIP 1, West Ring Road, Clayton, Victoria 3800, Australia
[2] Centre for Biospectroscopy and the School of Chemistry, Monash University, Clayton, Victoria 3800, Australia
[3] Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168
[4] Murdoch Childrens Research Institute, The Royal Children's Hospital, Parkville, Victoria, Australia, 3052

➡ **Supporting information** for this article is available free of charge under http://dx.doi.org/10.1002/jbio.201200217

Fourier transform infrared (FTIR) microspectroscopy was employed to elucidate the macromolecular phenotype of human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) and their differentiated progeny. Undifferentiated hESCs and hiPSC lines were found to be not clearly distinguishable from each other. However, although both hESC and hiPSC variants appeared to undergo similar changes during differentiation in terms of cell surface antigens, the derived cell types from all cell lines could be discriminated using FTIR spectroscopy. We foresee a possible future role for FTIR microspectroscopy as a powerful and objective investigative and quality control tool in regenerative medicine.
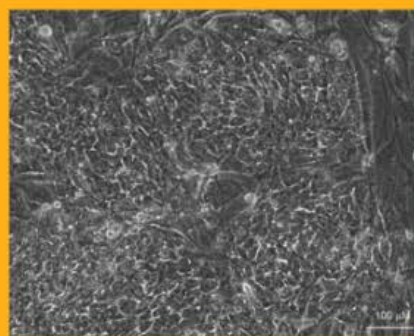
Bright field image of human embryonic stem cells in culture.

* Corresponding author: e-mail: phil.heraud@monash.edu, Phone: +61 3 9905 0765

2      J. Cao et al.: Spectroscopic difference between embryonic and re-programmed stem cells

## 1. Introduction

Human embryonic (hESC) and induced pluripotent (hiPSC) stem cells can form all the foetally-derived cells and tissues of the body and hava been advanced as a potential source of new cells for the emerging field of regenerative medicine. However, before their envisaged introduction into the clinic, the question of whether undifferentiated hESCs and hiPSCs and their differentiated progeny are equivalent must be answered [1]. In regards to their pluripotency, ability to form the three embryonic germ layers, and overall global gene expression programs [2–4], the two stem cell classes appear to be very similar [3, 5]. Nevertheless, others have observed subtle differences between their RNA levels, DNA methylation, protein expression and protein phosphorylation [6]. Some of these differences are attributed to the current limitations of the various reprogramming processes used in the generation of iPSCs [7, 8] and may contribute to the variable efficiencies with which hiPSCs can undergo differentiation [9].

Our current knowledge concerning iPSCs and their embryonic stem cell (ESC) counterparts has been acquired via the use of conventional molecular biology techniques such as in vitro and in vivo assays, bisulfite sequencing, flow cytometry, and gene array analysis. Although these methods are highly informative, they cannot provide a global overview of the biochemical similarities and differences between these two cell types. An attractive modality that provides such information is Fourier transform infrared (FTIR) microspectroscopy.

Fourier transform infrared microspectroscopy is a powerful, non-destructive, optical technique that is reliant on the intrinsic property of molecular systems to vibrate in resonance with different frequencies of infrared light. It is able to elucidate the macromolecular chemistry of a wide spectrum of biological samples and unlike the other conventional methods, can readily attain quantitative and qualitative data without the use of additional reagents or contrast agents. For example, FTIR spectroscopy is sensitive to the relative concentrations of macromolecules such as proteins, lipids, carbohydrates and nucleic acids within biological cells. Vibrational modes in macromolecular functional groups within biological samples give rise to a series of identifiable bands in the FTIR spectrum, providing information about their relative concentrations. Spectroscopic information is therefore complementary to other methodologies used to characterise stem cells and their differentiated progeny in the sense that the differences in macromolecular composition revealed by the FTIR spectrum can provide a unique phenotypic chemical 'signature' for a particular cell or tissue structure [10].

At present, there is only a limited body of literature reporting the application of FTIR microspectro-scopy in the study of embryonic and adult stem cells [11–13]. Furthermore, the only extant study using FTIR to classify human iPSCs [14] has been restricted to undifferentiated cells, where as iPSC derived lineage committed progeny progenitors have not been examined.

We have previously demonstrated the success of this method in discriminating between undifferentiated human embryonic stem cell lines and their early differentiated mesendodermal and ectodermal progeny [15]. As a continuation of this earlier work, we have used laboratory-based FTIR microspectroscopy to elucidate the FTIR spectral profiles of several readily available hESC and hiPSC lines and their early differentiated progeny. To confirm that the differentiation had been successful, flow cytometry was used to verify the presence or absence of well-established cell surface markers indicative of differentiation state. Our study showed that FTIR microspectroscopy was able to successfully discriminate different human embryonic and human induced pluripotent stem cells from each other and from their mesendodermal and ectodermal lineage committed progeny with high levels of sensitivity and specificity.

## 2. Materials and methods

A schematic of the general methodology used in this study is shown in Figure 1.

### 2.1 FTIR microspectroscopy of mouse embryonic fibroblasts

A monolayer of inactivated mouse embryonic fibroblasts (MEFs) is commonly used as a substrate in the culture of hES and hiPS cells [16]. In our attempts to minimize any spectral contamination that may have arisen from their presence in samples measured by spectroscopy, all MEFs that were used in the experiments were seeded onto the gelatinized tissue culture flasks at half their normal density. Further, the hESCs and hiPSCs were always harvested from cultures that were ~80% confluent to ensure a greater stem cell to feeder ratio.

Focal Plane Array (FPA) FTIR images comprised of a $64 \times 64$ array of spectra (1024 spectra in total, after the binning of each grouping of four adjacent pixels) were acquired from fields of cells. First, fields of MEF cells in the absence of hESCs and hiPSC were imaged to address the question of how spectroscopically similar the different cell types were to each other, and whether the presence of MEFs would confound the interpretation of their spectra.

J. Biophotonics (2013)  3

# Experimental procedure
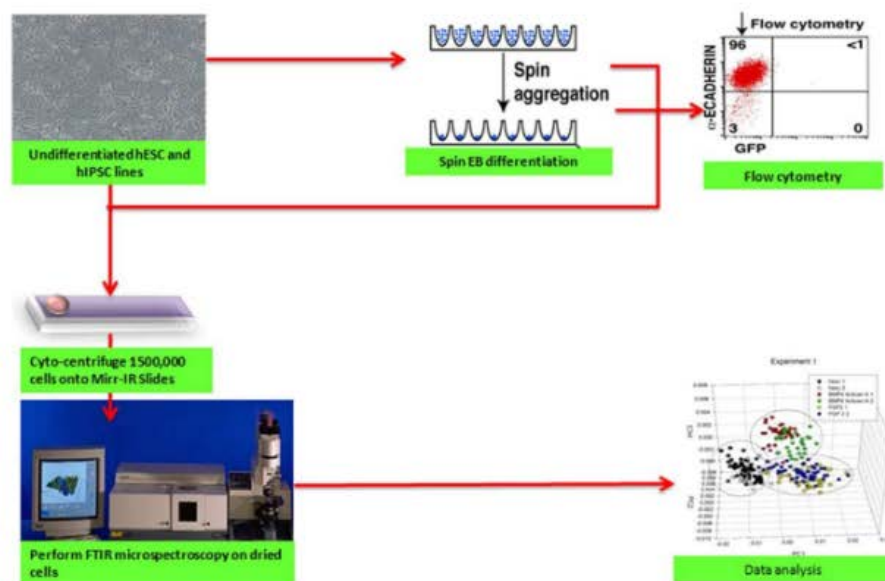


**Figure 1** Flow diagram of the experimental procedure employed during the study. Human ESC lines (HES3, H9, *MIXL1*[GFP/w], NKX 2.5[GFP/w] and ENVY) and hiPSC lines (IMR90C2, ES4CL1 and AUS) maintained in bulk culture were harvested and cytospun onto MirrIR slides for FPA-FTIR microspectroscopy. Some of these lines were also differentiated towards mesendodermal and ectodermal lineages before being cytospun onto the same slides for data acquisition. To verify that the differentiation process had been successful, FACs analysis was performed on the samples to check for the presence or absence of well-established stem cell and differentiation associated cell surface markers.

MEFs were maintained for 2 days on gelatin coated T25 flasks in KOSR-based hESC medium before being dissociated via enzymatic passaging [17]. The harvested cells were then washed 3 times in phosphate buffered saline (PBS) before approximately 150,000 cells were cyto-centrifuged (Cytospin III, Thermo Fischer Scientific, Waltham, MA) onto infrared reflective slides (MirrIR slides, Tienta Technologies, OH, USA).

## 2.2 FTIR microspectroscopy of undifferentiated human embryonic and human induced pluripotent stem cells

Five human embryonic stem cell lines, HES3 [18], H9 obtained from WiCell, and, NKX 2.5[GFP/w]

*MIXL1*[GFP/w], and ENVY [19–21] in addition to 3 human induced stem cell lines, ES4CL1 derived from foreskin, and lung fibroblast derived IMR90C2 provided by the Australian Stem Cell Centre [5], and skin fibroblast derived AUS [22] from the Monash Institute of Medical Research, were cultured on top of a low density feeder monolayer on gelatinized T25 tissue culture flasks in KOSR-based hESC medium using methods previously described [17]. The medium was changed daily and when the cells were at least 80% confluent, they were enzymatically passaged using TrypLE [17] and washed 3 times in PBS before 150,000 cells were cyto-centrifuged onto MirrIR slides for FTIR microspectroscopy using two replicate cultures.

4        J. Cao et al.: Spectroscopic difference between embryonic and re-programmed stem cells

### 2.3 FTIR microspectroscopy of human embryonic and human induced pluripotent stem cells differentiated towards a mesendodermal and ectodermal lineage

Two human embryonic stem cell lines, $MIXL1^{GFP/w}$ (a genetically modified hESC line expressing green fluorescent protein (GFP) from the MIXL1 locus [20]) and MEL1, in addition to one of the human induced pluripotent stem cell lines IMR90C2, were used for differentiation experiments. Differentiation towards mesendodermal lineages occurred in serum-free medium supplemented with bone morphogenetic protein (BMP4) and Activin A (BMP4/Act A medium) or toward ectodermal lineages in medium supplemented with fibroblast growth factor (FGF2), using the previously described spin embryoid body (EB) method [23, 24]. After 4 days of differentiation, the cells were immuno-stained for flow cytometry to confirm that the differentiation process had been successful. Once differentiation had been confirmed by the expression of the appropriate cell biomarkers, 10 EBs were dissociated with TrypLE select, and washed 3 times with phosphate buffered saline (PBS) before being cytospun onto MirrIR slides.

### 2.4 Fluorescence Activated Cell Sorting (FACS)

To verify that the differentiation process had been successful, flow cytometry was carried out using antibodies directed against human E-CADHERIN (Zymed, San Francisco, CA), TRA-1-60 (Millipore Corporation, MA), SSEA-4 (Millipore Corporation), and CD9 (BD Biosciences, CA) as previously described [20, 24].

### 2.5 FPA FTIR imaging

FPA FTIR infrared images were acquired on a Agilent Technologies FTIR spectrometer (Model FTS 7000; Agilent Technologies Inc., Palo Alto, CA, USA) in transflectance mode, coupled to an infrared microscope (model 600 UMA; Agilent Technologies) using a 15× objective and equipped with a 64 × 64 pixel MCT liquid nitrogen cooled FPA detector located at the Centre for Biospectroscopy at Monash University in Melbourne, Australia. FTIR spectra were acquired with 256 co-added background scans, at 8 $cm^{-1}$ spectral resolution, with signals from groups of 4 adjacent pixels binned and co-added for 128 scans. The binning of 4 adjacent pixels on the FPA resulted in each spectrum corresponding to an area of 11 μm × 11 μm on the sample plane, which was the approximate dimensions of single cells in the dried monolayer. A Happ Genzel apodisation function was used to truncate the interferogram, and a zero-filling factor of 2 was employed.

### 2.6 Data preprocessing and multivariate data analysis

The infrared images were taken into Cytospec 1.03 IR imaging software (Cytospec Inc., NY, and USA) in order to extract the spectral data deemed suitable for preprocessing and subsequent analysis. Various analysis techniques are employed in the analysis of FTIR biospectroscopy spectra, and the rationales for all these methods are explained extensively in Trevisan et al.'s review [25]. However, the analyses methods which were used by our group are described as follows. First, to ensure measurements had an appropriate signal to noise ratio and were in the linear range of the detector's response, the data first underwent a quality test whereby spectra with absorbance values outside the range of 0.2–0.8 (arising from sample regions devoid of cells or where cells were clumped and overlaid) were rejected. The spectra that passed the quality test underwent further preprocessing in The Unscrambler v 10.1 (Camo, Oslo, Norway) software package. Here, spectra were converted to their second derivatives using a Savitsky-Golay algorithm using 9 smoothing points in order to minimize baseline effects and to resolve spectral components that would otherwise be overlapping in the underivatised spectra. Normalisation was performed using Extended Multiplicative Signal Correction [26] using the following spectral ranges: 3050–2800 $cm^{-1}$, 1770–1730 $cm^{-1}$, 1470–900 $cm^{-1}$, as these ranges contain biological bands and therefore provided the best possibility of discrimination between different sample types. The amide I and amide II region were omitted from the analysis as this region is greatly affected by resonant Mie scattering, which may have confounded our interpretation of the spectra [27]. The data set was then randomly sorted into 2 new data sets comprising 2 thirds and 1 thirds of the spectra. These new sets were then used as the calibration and datasets respectively for Principal Component Analysis (PCA) and Soft Independent Modeling of Class Analogy (SIMCA) [28, 29], and Partial Least Squares Discriminant Analysis (PLS-DA) [30]. To enable comparison between spectral averages and their standard deviation, mean and standard deviation spectra for each hESC and hiPSC line were calculated using the 2nd derivative spectra that passed the quality test.

### 2.7 Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), Soft Independent Modeling of Class Analogy (SIMCA)

PCA was carried out on the calibration set using 5 PCs, and the score plots were used to visualize any clustering of the data, while the loading plots were used to determine which spectral regions contributed most to the variance in the data set, accounting for the observed clustering. SIMCA was then applied to the PCA models from the calibration set using an alpha level of 0.05 to determine the interclass distances. A SIMCA model distance of greater than 3 is generally indicative that the two clusters are classifiable and hence significantly different [31].

The data set was then randomly sorted into 2 new data sets comprising 2 thirds and 1 thirds of the spectra and used as the calibration and validation datasets respectively for classification of spectra using Partial Least Squares Discriminant Analysis (PLS-DA) [30]. The calibration data matrix employed for PLS-DA consisted of the spectral dataset (multivariate $X$) and two $Y$ variables with integer values of 0 or 1 coding for the each of the two modelled spectral classes. Classification of the dataset was then carried out by predicting a $Y$ value for each spectrum in the independent validation using PLSR

models that had been generated from the calibration sets. Correct classification of each class was arbitrarily assigned to samples with predicted $Y > 0.5$ for respective spectra. PLS loading plots were used to determine which spectral regions contributed most to the ability to classify spectra using PLS-DA.

## 3. Results and discussion

### 3.1 Some independent human embryonic stem cell lines exhibit unique FTIR spectra

Previous work suggests that different origins and culture conditions can lead to different hESC lines having distinct gene expression signatures [32]. Accordingly, our initial experiments involved determining the spectroscopic similarities between five independently derived hESC lines grown under standardized conditions.

Inspection of the average second derivative spectra and their standard deviations revealed line-specific FTIR absorbance differences in the spectral regions attributed to lipids, protein, amino acids, carbohydrates and nucleic acids (Figure 2C). PCA revealed limited clustering of spectra from the same cell lines in the scores plot along PC1 – the principal
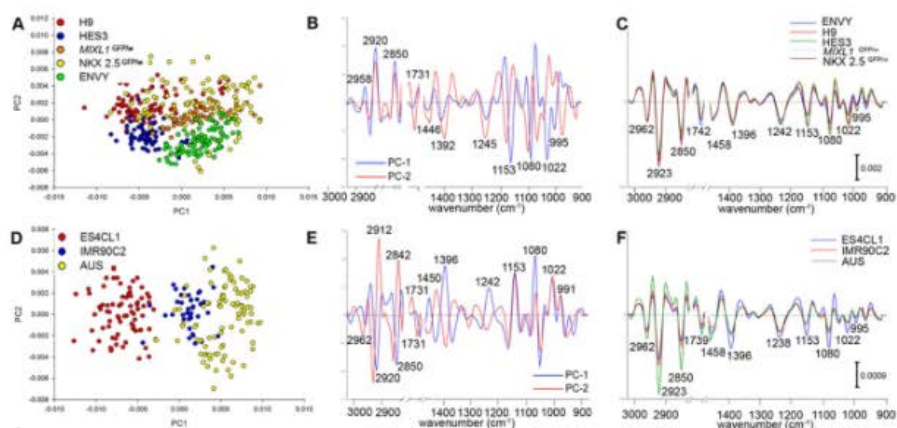


**Figure 2** Some, but not all hESCs and hiPSC lines can be readily discriminated from each other due to their unique FTIR spectroscopic profiles. PCA scores plot (**A**) loadings plot (**B**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**C**) for spectra from 5 different hESC lines acquired using FPA-FTIR microspectroscopy. PCA scores plot (**D**) loadings plot (**E**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**F**) for spectra from 3 different hiPSC lines acquired using FPA-FTIR microspectroscopy.

6                    J. Cao et al.: Spectroscopic difference between embryonic and re-programmed stem cells

component indicating the greatest source of variation in the dataset (Figure 2A). For example, some separation was observed between clusters of spectra from the H9, ENVY and MIXL1GFP/w groups, whereas considerable overlapping was evident between the other sample classes. Since the spectra had been converted to their $2^{nd}$ derivatives, the negatively scored spectra corresponded to positive PC loadings, and positively scored spectra corresponded to negative PC loadings. That is, positively scored spectra would have highest absorbance for negatively-loaded bands and vice versa (Figure 2B). The PC1 loadings bands that explained the variation in the dataset, were the bands at 2920 cm$^{-1}$ and 2850 cm$^{-1}$ and 1731 cm$^{-1}$, assigned to lipid groups, 2920 cm$^{-1}$ and 1450 cm$^{-1}$ assigned to proteins and lipids, at 1404 cm$^{-1}$ assigned to free amino acids, 1242 cm$^{-1}$, 1080 cm$^{-1}$ and 995 cm$^{-1}$, from nucleic acids, and 1153 cm$^{-1}$, 1080 cm$^{-1}$ and 1022 cm$^{-1}$, assigned to C−O vibrations from glycogen and other carbohydrates (See Table 1 for details of the band assignments).

Further analysis of the dataset employing SIMCA, revealed significant differences between some of the cell lines (Supplementary Figure 1A). The model distance plot showed that in relation to the H9 line, HES3, NKX2.5 GFP/w and ENVY all had SIMCA model distances of greater than 3, indicating that these groups were more likely to be spectroscopically distinct from each other. In contrast, spectra from the MIXL1GFP/w line had a SIMCA model distance of less than 3, in relation to spectra from the H9 group, which suggested that they were very similar. PLS-DA modeling was also performed for the calibration and dataset shown in this experiment (Supplementary Figure 1D). The validation set

was used to predict Y using the PLS-DA model and sensitivity and specificities for each of the lines was as follows: 83% and 100% for the MIXL1GFP/w line, 85% and 100% for NKX 2.5GFP/w, 93% and 100% for the ENVY line, 55% and 96% for the H9 group and 48% and 100% for the HES3 line. The poorer classification outcome of some of these lines was not surprising given the observed overlapping of spectra from different classes in the scores plot, and the small SIMCA model distances as described above.

Considering that the NKX2.5GFP/w, ENVY and MIXL1GFP/w hESC lines are sub-clones of HES3, it was surprising that some of them were spectroscopically distinguishable from each other. This result might suggest that, in addition to underlying genetics, the history of each line may impact on its overall biochemical composition. This study also highlights the power of FTIR microspectroscopy to provide invaluable information about a sample that may have otherwise gone undetected by conventional laboratory methods used in stem cell biology.

### 3.2 Different human induced pluripotent stem cell lines possessed unique FTIR 'spectral signatures'

Using the same rationale for performing the previous comparison, a similar multivariate data analysis was carried out using FTIR spectra collected from 3 different human iPSC lines. Upon inspection of the PCA plots, the spectra were found to cluster into distinct groups according to cell type, and these

**Table 1** Band maxima of human embryonic and human induced pluripotent stem cells (replicate no. 1).

| Band maxima $2^{nd}$ derivative spectra (cm$^{-1}$) hESC, hIPSC | PC1 loadings (cm$^{-1}$) hESC, hIPSC | PC2 loadings (cm$^{-1}$) hESC, hIPSC | Band assignments |
|---|---|---|---|
| ~2962 | ~2962 | ~2962 | C−H stretching peak of lipids and proteins |
| ~2923 | ~2920 | ~2920 | C−H asymmetrical stretching of lipid groups and protein |
| ~2850 | ~2850 | ~2850 | C−H symmetrical stretching of lipid groups |
| ~1743 | ~1731 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1454 | ~1446 | Methyl and methylene groups from lipids and protein |
| ~1396 | ~1396 | ~1392 | COO$^-$ stretching vibrations of amino acid side chains |
| ~1238 | ~1238 | ~1245 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| ~1080 | ~1080 | ~1091 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules, and glycogen |
| ~1153 | ~1153 | ~1164 | C−O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1018 | C−O vibrations from glycogen and other carbohydrates |
| ~995 | ~995 | ~991 | C−O stretch from RNA ribosechain and other carbohydrates |

groups appeared to be more disparate than was observed for the hESC lines (Figure 2D). Unlike the prior result, where there was considerable overlap between the lines, the PCA scores plot from this data showed distinctly negatively and positively scored clusters displaying a clear separation along PC1. The negatively scored spectra for the ES4CL1 line clustered away from the other two co-localising hiPSC lines, the IMR90C2 and AUS lines. In agreement with their second derivative spectra (Figure 2F), loadings plots indicated that the latter two cell lines had higher IR absorbances at 2920 cm$^{-1}$, 2850 cm$^{-1}$, and 1731 cm$^{-1}$, the spectral region that corresponded to lipids, compared to the ES4CL1 line. However, they had lower IR absorbances for bands ascribed to proteins and lipids 1454 cm$^{-1}$, free amino acids 1396 cm$^{-1}$, phosphorylated molecules (1080 cm$^{-1}$ and 1238 cm$^{-1}$), glycogen and other carbohydrates (1080 cm$^{-1}$, 1153 cm$^{-1}$ and 1018 cm$^{-1}$) (Figure 2E and Table 1).

Subsequent SIMCA modeling further verified that the groups were divergent in terms of macromolecular composition, as both the IMR90C2 and AUS cell lines had SIMCA interclass distances of greater than 3 in relation to the ES4CL1 group (Supplementary Figure 1B). The spectral data sets of these lines were found to classify very well with PLS-DA modeling, which was successful at discriminating between the three lines, achieving sensitivities and specificities of 100% for all three groups (Supplementary Figure 1E). We hypothesised that the spectroscopic disparities between the different cell lines were probably due to subtle differences between them resulting from the methods used to derive them. The protocols that are typically used to generate human induced pluripotent stem cells – viral transduction, DNA- based induction, mRNA transfection, and recombinant proteins- can be applied to a variety of somatic cell types. All of the hiPSC lines used in this study were 'reprogrammed' using viral vectors, but each had different somatic cell origins. The ES4CL1 line was derived from foreskin, the IMR90C2 line, from lung fibroblast, and the AUS line from skin fibroblast. It has been reported that many human iPSC lines may retain an 'epigenetic memory' of the somatic cells from which they were created [33], and it is therefore plausible that the differences we observe between these cells lines could be related to this effect.

In the only other study that has investigated FTIR spectral 'signatures' of iPS cells, Sandt et al. found somatic amniotic fluid cells (AFC) derived iPSC lines to be spectrally indistinguishable from human ESC lines [14]. Conversely, small chemical differences were found to exist between hiPSC lines generated from H9 derived mesenchymal stem cells and hESC lines, indicated by their ability to be classified with PLS-DA.

### 3.3 Line dependent differences in the spectroscopic profiles of human embryonic stem cells and human induced pluripotent stem cells

Given the macromolecular variability that was seen amongst lines of the same stem cell type, we subsequently compared spectra from both stem cell types (hESC and hiPSC) together, to see if those differences were still detectable when compared to each other. From the average second derivative spectra, differences could certainly be seen across the spectral regions ascribed to lipids, protein, amino acids, carbohydrates, and nucleic acids but subsequent multivariate analyses did not reveal any consistent clustering based on whether the cell lines were hESC or hiPSC (data not shown).

Previous non-spectroscopic studies have identified a number of differences between the hiPSCs and hESCs, including their epigenetic state and transcriptional signature [9, 33, 34]. How these molecular differences impact on the overall phenotype of the cells remains contentious. A recent investigation by Bock et al., for instance, found the effect of epigenetic memory on the DNA methylation and gene expression of 12 fibroblast derived iPSC lines to be statistically insignificant [9]. The group also found that most, but not all of the iPS cell lines could be readily distinguished from their ESC counterparts based on their DNA methylation and/or gene expression profiles. However, an epigenetic or transcriptional variation profile that was common to all the investigated iPS cell lines could not be detected, in line with the results presented here.

### 3.4 Undifferentiated human embryonic stem cells and human induced pluripotent stem cells can be discriminated from their differentiated progeny by their unique FTIR spectroscopic 'signatures'

We then wished to determine whether hESC and hiPSC undergoing differentiation towards mesendodermal or ectodermal lineages underwent similar macromolecular chemical changes. Firstly, in order to verify that the differentiation process had been successful, in addition to enabling a correlation between cell differentiation and spectral signature, aliquots of the ectodermally or mesendodermally differentiated cultures that were interrogated by FTIR were analysed by flow cytometry. This analysis verified the assigned differentiation status of the samples and enabled a comparison of the differentiation obtained

**Figure 3** Flow cytometry profiles of ectodermally and mesendodermally differentiated human pluripotent cell lines subsequently analysed by vibrational spectroscopy. (**A**) *MIXL1*[GFP/w] cells differentiated for four days as spin EBs in APEL medium supplemented with FGF2 (Ectoderm differentiation) or with a combination of BMP4 and ACTIVIN A (Mesendoderm differentiation), were disaggregated and single cell suspensions analysed by flow cytometry for the expression of GFP from the MIXL1 locus (MIXL1 GFP) or for the expression of the indicated cell surface markers. (**B**, **C**) The human ESC lines *MIXL1*[GFP/w], MEL1 or the human iPSC line IMR90C2 were differentiated as above towards (**B**) ectoderm of (**C**) mesendoderm and analysed by flow cytometry for the expression of the indicated cell surface markers. The quadrants were set using appropriate isotype control antibodies (data not shown).

9

using the three pluripotent cell lines. As shown in Figure 3A, *MIXL1*[GFP/w] cells differentiated towards ectoderm using FGF2 [15], did not express GFP from the MIXL1 locus, consistent with the lack of mesendoderm differentiation under these conditions. Similarly, these cells did not express PDGFRα or CXCR4, other markers associated with early mesoderm and endoderm, but retained expression of the epithelial marker EpCAM. Conversely, when these cells were differentiated in response to a combination of BMP4 and ACTIVIN A for four days, ~97% of the cells expressed MIXL1 in a graded fashion, consistent with mesendodermal commitment [20]. Most cells also expressed PDGFRα and many were CXCR4 positive,

but very few cells expressed EpCAM in the absence of MIXL1 expression. We have previously shown that during early mesendodermal differentiation, MIXL1 is co-expressed with epithelial markers such as E-CADHERIN or EpCAM [15, 20, 35]. Analyses examining the co-expression of EpCAM or PDGFRα with CXCR4 in the hESC lines *MIXL1*[GFP/w] and MEL1 and the hiPSC line IMR90C2, demonstrated similar patterns of marker expression under ectodermal conditions, with retention of EpCAM and failure to induce significant proportions of the cells to express PDGFRα or CXCR4 (Figure 3B). In response to mesendoderm differentiation, very few cells expressed EpCAM alone (2−8%), although the propor-
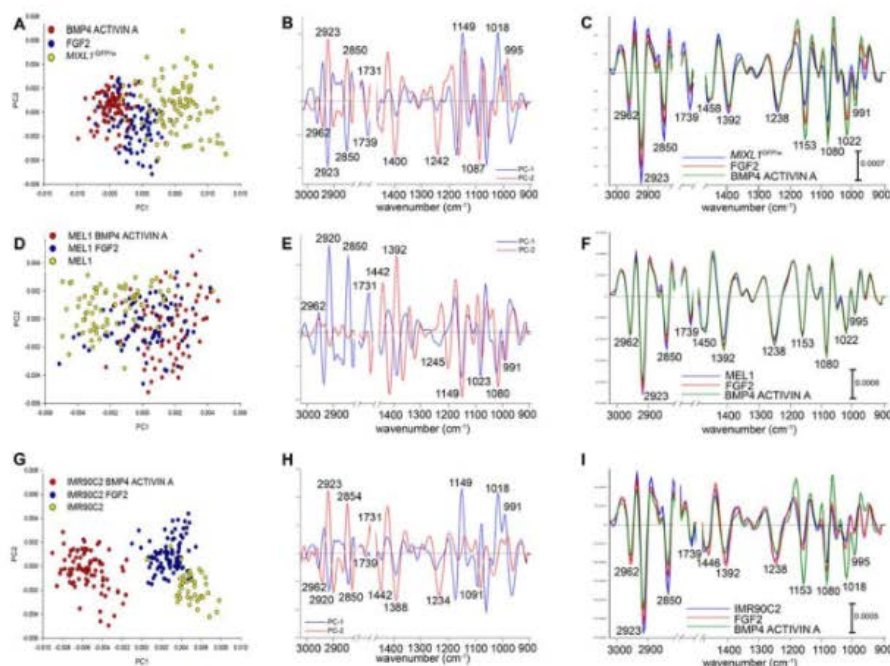


**Figure 4** hESC and hiPSC lines undergo the same macromolecular chemical changes upon differentiation towards mesendodermal and ectodermal lineages. PCA scores plot (**A**) loadings plot (**B**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**C**) for spectra from the hESC line *MIXL1*[GFP/w] differentiated towards mesendoderm using BMP4/ACTIVIN A or towards ectoderm with FGF2 acquired using FPA-FTIR microspectroscopy. PCA scores plot (**D**) loadings plot (**E**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**F**) for spectra from the hESC line MEL1 differentiated towards mesendoderm using BMP4/ACTIVIN A or towards ectoderm with FGF2 acquired using FPA-FTIR microspectroscopy. PCA scores plot (**G**) loadings plot (**H**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**I**) for spectra from the hiPSC line IMR90C2 differentiated towards mesendoderm using BMP4/ACTIVIN A or towards ectoderm with FGF2 acquired using FPA-FTIR microspectroscopy.

tion of PDGFRα and CXCR4 expressing cells varied a small amount between the three lines examined (Figure 4C). These data suggested that the concentrations of BMP4 and ACTIVIN A used for these experiments imparted a more mesodermal bias to the *MIXL1*[GFP/w] cells (leading to more cells expressing PDGFRα and fewer expressing CXCR4) and a more endodermal bias to the IMR90C2 line, in which a higher percentage of cells expressed CXCR4.

Spectral differences that were observed between undifferentiated hESC and hiPSC lines and their progeny resembled those previously described by our group [15]. In the present study, the chemical differences were characterised by higher signal intensities for lipid components at ~2920 cm⁻¹, ~2850 cm⁻¹ and ~1739 cm⁻¹ in the average second derivative spectra of all the undifferentiated cells across all three cell lines (Figure 4C, F and I). In contrast, higher FTIR absorbances were consistently seen in the spectral regions corresponding to glycogen and other carbohydrates at ~1153 cm⁻¹ and ~1022 cm⁻¹ in the spectra of the BMP4/ACTIVIN A differentiated progeny. Observations from the second derivative spectra were in agreement with the PCA results, which showed that in the PCA scores plots along PC1, spectra derived from undifferentiated cells clus-

tered separately from spectra acquired from their differentiated progeny (Figure 4A, D, and G). In data derived from both stem cell types, the PC1 loadings plots (Figure 4B, H, and Table 2) were heavily loaded by bands attributable to lipids from C−H stretching vibrations from methylene groups at around 2920 cm⁻¹, 2850 cm⁻¹, and the ester carbonyl stretching band at 1739 cm⁻¹. Lipid absorbance was found to be higher in the undifferentiated cell lines, than in mesendoderm committed cells. Spectral clusters derived from BMP4/ACTIVIN A differentiated cells were found to be heavily loaded by bands ascribed to glycogen and other carbohydrates (1149 cm⁻¹, ~1080 cm⁻¹ and 1018 cm⁻¹).

Overall, the SIMCA model distances were greater than 3 between the undifferentiated cells and the differentiated progeny, suggesting they should be highly classifiable (Supplementary Figure 2A, B and C). Indeed, for the *MIXL1*[GFP/w] differentiation experiment, PLS-DA classification could be applied with the following sensitivities and specificities: 100% and 100% for the BMP4/ACTIVIN A treated samples, 97% and 97% for the FGF2 treated samples, and 80% and 100% for the undifferentiated line, (Supplementary Figure 2D). Classification of independent test spectra from the MEL1 differentiation experiment

**Table 2** Band maxima of human embryonic and human induced pluripotent stem cells differentiated with BMP4 ACTIVIN A and FGF2.

| Band maxima 2ⁿᵈ derivative spectra (cm⁻¹) | PC1 loadings (cm⁻¹) | PC2 loadings (cm⁻¹) | Band assignments |
|---|---|---|---|
| *MIXL1*[GFP/w], MEL1, IMR90C2 +BMP4 ACTIVIN A/ FGF2 | *MIXL1*[GFP/w], MEL1, IMR90C2 +BMP4 ACTIVIN A/ FGF2 | *MIXL1*[GFP/w], MEL1, IMR90C2 +BMP4 ACTIVIN A/ FGF2 | |
| ~2962 | ~2962 | ~2962 | C−H stretching peak of lipids and proteins |
| ~2923 | ~2920 | ~2923 | C−H asymmetrical stretching of lipid groups |
| ~2850 | ~2850 | ~2850 | C−H symmetrical stretching of lipid groups |
| ~1739 | ~1735 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1450 | ~1450 | Methyl and methylene groups from lipids and protein |
| ~1392 | ~1392 | ~1392 | COO⁻ stretching vibrations of amino acid side chains |
| ~1238 | ~1245 | ~1249 | P=O asymmetrical stretching of PO₂ phosphodiesters from phosphorylated molecules |
| ~1080 | ~1080 | ~1091 | P=O symmetrical stretching of PO₂ phosphodiesters from phosphorylated molecules and glycogen |
| ~1153 | ~1153 | ~1149 | C−O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1018 | C−O vibrations from glycogen and other carbohydrates |
| ~995 | ~991 | ~991 | C−O stretch from RNA ribose chain and other carbohydrates |

had sensitivities and specificities of 76% and 90% for the BMP4/ACTIVIN A treated group, 85% and 88% for the FGF2 treated group, and 100% and 96% for the undifferentiated MEL1 line (Supplementary Figure 2E). The validation set from the differentiation of the hiPSC line, IMR90C2 was found to classify at 100%, 93% and 80% for the BMP4/ACTIVIN A, FGF2, and undifferentiated groups respectively, and their specificities that were obtained were 100%, 97% and 100% (Supplementary Figure 2F).

The finding that lipid stores diminished during the loss of pluripotency by hiPSCs corroborates hESC differentiation data previously published by our laboratory and other groups [11, 15]. The reasons why lipid depletion occurs during the differentiation process is unclear but one group has suggested that the eicosanoid pathway may play an important role [36]. They discovered that the inhibition of the eicosanoid signaling pathway of murine embryonic stem cells fostered pluripotency when the cells underwent neural differentiation. However, other groups have also observed either higher lipid levels in more differentiated stem cells [37] or non significant lipid differences [12] between their undifferentiated and their progeny which suggests that this may be a cell-type specific phenomenon.

### 3.5 Differentiated progeny of human ESCs and iPSCs are spectroscopically distinguishable

Considering that FTIR microspectroscopy indicated that similar macromolecular changes were occurring during the differentiation of both the hESC and hiPSC lines, this prompted us to investigate whether these changes yielded progeny that were biochemically identical.
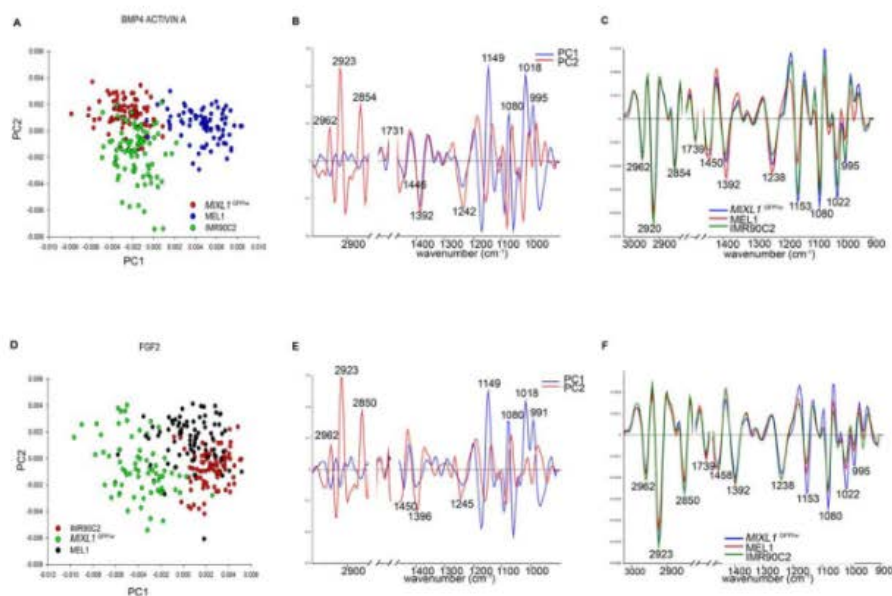


**Figure 5** hESC and hiPSC differentiated towards mesendodermal and ectodermal lineages do not yield biochemically equivalent progeny. PCA scores plot (**A**) loadings plot (**B**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**C**) for spectra from the hESC lines *MIXL1*^GFP/w, MEL1 and the hiPSC line IMR90C2 treated with BMP4/ACTIVIN A acquired using FPA-FTIR microspectroscopy. PCA scores plot (**D**) loadings plot (**E**) and average second derivative spectra plot with the highest standard deviation across all of the spectra displayed (**F**) for spectra from the hESC lines *MIXL1*^GFP/w and MEL1 and the hiPSC line IMR90C2 treated with FGF2 acquired using FPA-FTIR microspectroscopy.

Journal of
**BIOPHOTONICS**

PCA was performed using the spectra of BMP4/ ACTIVIN A treated cells derived from the *MIXL1*[GFP/w] hESC line, MEL1 hESC line, and IMR90C2 iPSC line (Figure 5). Along PC1 of the PCA scores plot, a separation could be seen between clusters of spectra derived from the *MIXL1*[GFP/w] and IMR90C2 lines away from the MEL1 spectral clusters. In accordance with the average second derivative spectra. High FTIR absorbances were consistently seen in the spectral regions corresponding to lipids (at ~2916 cm$^{-1}$, 2842 cm$^{-1}$ and 1731 cm$^{-1}$), carbohydrates (at ~1153 cm$^{-1}$ and 1022 cm$^{-1}$) in the spectra of the BMP4/ACTIVIN A differentiated progeny derived from the *MIXL1*[GFP/w] and IMR90C2 lines (Table 3). Whereas spectral clusters derived from MEL1 progeny were found to have high FTIR intensities from bands ascribed to proteins and lipids (1446 cm$^{-1}$), and free amino acids (1392 cm$^{-1}$).

The SIMCA model distances were consistently greater than 3 when the MEL1 line was taken as the model center, whereas the other two lines appeared to be more similar to each other (Supplementary Figure 3). Further, PLS-DA classification could be applied with the following sensitivities and specificities, respectively: 53% and 100% for the *MIXL1*[GFP/w] line, 100% and 85% for the MEL1 line, and 100% and 82% for the IMR90C2 line.

Multivariate analysis of spectra derived from the FGF2 differentiated progeny of the same three lines showed the *MIXL1*[GFP/w] spectral cluster separating away from spectra of the two other cell lines along PC1 of the PCA scores plot (Figure 5D). The loadings bands explaining the variance contributing to the clustering pattern were the bands that corresponded to lipid groups (~2916 cm$^{-1}$, 2842 cm$^{-1}$ and 1731 cm$^{-1}$), in addition to those ascribed to glycogen and other carbohydrates (~1149 cm$^{-1}$, 1080 cm$^{-1}$ and 1018 cm$^{-1}$). FTIR intensities from these aforementioned bands were higher in the *MIXL1*[GFP/w] spectral cluster compared to spectral clusters of the other two lines, which exhibited higher IR intensities for the spectral regions which corresponded to protein (~1446 cm$^{-1}$), free amino acids (~1392 cm$^{-1}$), and phosphorylated molecules (~1242 cm$^{-1}$). SIMCA classification results showed the *MIXL1*[GFP/w] line to be highly classifiable compared to the other two lines, consistently having a model distance of greater than 3 when taken as the model centre. The spectral data sets of these lines were found to be classifiable with PLS-DA modeling, which was successful at discriminating between the three lines, achieving sensitivities and specificities of 100% and 77% for the MEL1 line, 86% and 96% for the IMR90C2 line, and 54% and 100% for the *MIXL1*[GFP/w] line.

**Table 3** Band maxima of comparisons between BMP4 ACTIVIN A and FGF2 treated progeny of hESC lines *MIXL1*[GFP/w] and MEL1 and hiPSC line IMR90C2.

| Band maxima 2$^{nd}$ derivative spectra (cm$^{-1}$) | PC1 loadings (cm$^{-1}$) | PC2 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|
| *MIXL1*[GFP/w], MEL1, IMR90C2 +BMP4 ACTIVIN A, +FGF2 | *MIXL1*[GFP/w], MEL1, IMR90C2 +BMP4 ACTIVIN A, +FGF2 | *MIXL1*[GFP/w], MEL1, IMR90C2 +BMP4 ACTIVIN A, +FGF2 | |
| ~2962 | ~2962 | ~2962 | C–H stretching peak of lipids and proteins |
| ~2923 | ~2927 | ~2920 | C–H asymmetrical stretching of lipid groups |
| ~2854 | ~2854 | ~2850 | C–H symmetrical stretching of lipid groups |
| ~1739 | ~1739 | ~1739 | C=O stretching of lipid esters |
| ~1458 | ~1446 | ~1446 | Methyl and methylene groups from lipids and protein |
| ~1392 | ~1392 | ~1392 | COO$^-$ stretching vibrations of amino acid side chains |
| ~1238 | ~1242 | ~1238 | P=O asymmetrical stretching of phosphodiesters from phosphorylated molecules |
| ~1080 | ~1091 | ~1080 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules and glycogen |
| ~1153 | ~1149 | ~1157 | C–O vibrations from glycogen and other carbohydrates |
| ~1022 | ~1018 | ~1022 | C–O vibrations from glycogen and other carbohydrates |
| ~995 | ~991 | ~995 | C–O stretch from RNA ribose chain and other carbohydrates |

This result is noteworthy because it shows that despite both stem cell variants undergoing the same macromolecular changes during the 4 days of differentiation, the resulting cells from all lines tested were spectroscopically distinguishable from each other. It is feasible that the aforementioned disparities between the different cell lines may reflect different responses to cytokine cues resulting in lineage committed progeny that appear phenotypically equivalent in terms of the cell surface marker expression (Figure 3), but are still distinct in terms of their macromolecular composition.

This difference in differentiation potential between different hES and hiPS cell lines has also been observed by others using non-spectroscopic analyses. In one recent study, Bock et al. compared the capacity of 12 human iPSC lines generated via lentiviral, retroviral, or non-integrating episomal methods to form neural lineages with 5 hESC lines [9]. The study found that the iPSC lines were able to differentiate into neuroepithelial (NE) cells and functional neurons or glia with the same kinetics as that observed for the hESC lines. Nevertheless, the differentiation outcome was found to proceed with reduced efficiency and higher variability when compared to human embryonic stem cells. Similarly, Narsinh et al. described, poorer *in vitro* differentiation outcomes with hiPSC lines derived from lentiviral transgenesis and non-viral minicircle based reprogramming differentiating into beating cardiomyocytes and ECs, as a consequence of higher variability and lower expansion capabilities [20].

Certainly, to achieve greater success in hiPSC differentiation experiments, there needs to be an appropriate screening method, capable of selecting the iPSC lines with the greatest propensity for differentiating towards specific lineages. One such quality control technique was described in Bock et al.'s study whereby a 'lineage scorecard' assay, consisting of the profiles of 500 relevant genes was created to predict the hiPSC lines that would most effectively differentiate into motor neurons [9].

Nonetheless, the last finding in our study suggests that it may be insufficient to use parameters such as genetics and differentiation ability, as the sole criteria for determining the hiPSC lines that most closely resemble hESCs. Our data raise a critical point that despite hiPSC lines exhibiting similar differentiation kinetics to their hESC counterparts they may still be divergent, in terms of their macromolecular chemistry. At present the exact implications of this finding are unclear, but one should remain mindful of these differences and their potential to affect experimental outcomes when designing future studies involving hESC and hiPSC lines.

Perhaps an alternative solution to Bock et al.'s strategy could be an integrative testing method, encompassing features of the lineage scorecard assays

with other robust phenotypic-based modalities, including FTIR microspectroscopy. Not only is this chemical tool able to provide invaluable, complementary information alongside conventional molecular biology methods, but it is also able to do so rapidly, and non-invasively. In addition, its capacity for generating large quantities of data (up to 4096 spectra with the use of a $64 \times 64$ focal plane detector in a single experiment) would be advantageous in a clinical environment where rapid yet detailed assessment of a cells phenotype was required.

A recent exciting development in the field has been the utilisation of quantum cascade laser sources of mid infrared light [38]. These lasers provide sources of IR light more brilliant than synchrotron sources, allowing FTIR measurements with high signal to noise ratios to be achieved in the laboratory, even with single living cells. For future experiments we hope to use either laser based, or synchrotron IR systems to probe the macromolecular chemical signatures of these clinically anticipated cell lines *in vivo*. This would be an invaluable exercise as it would enable the investigation of IR spectral differences at the single cell level, and to more closely study the nucleic acid region of the IR spectrum, which can be ambiguous in dehydrated samples [39].

Advances in instrumentation and analysis techniques will allow for the feasible implementation of FTIR microspectroscopy as an objective, label-free, non-destructive technique for the screening of clinically destined stem cells and their derivatives. This modality would not only be advantageous in the interrogation of pluripotent and/or multipotent stem cells to be used for regenerative therapies, but also holds great promise for the detection of cancer stem cells [40, 41].

## 4. Conclusions

The work presented here shows that some, but not all, human embryonic and human induced pluripotent stem cells have significant spectroscopic and hence, macromolecular differences from each other. Whilst both stem cell variants undergo similar macromolecular changes during differentiation, with the main change involving lipid depletion, they generate spectroscopically distinct, lineage committed cells. We have demonstrated the robust abilities of FTIR microspectroscopy for providing further insight into the contentious hESC versus hiPSC equivalency debate.

Journal of
**BIOPHOTONICS**

## References

[1] K. H. Narsinh, J. Plews, and J. C. Wu, Mol Ther **19**(4), 635–638 (2011).

[2] J. Yu, K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, I. Slukvin, and J. A. Thomson, Science **324**(5928), 797–801 (2009).

[3] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka, Cell **131**(5), 861–872 (2007).

[4] M. G. Guenther, G. M. Frampton, F. Soldner, D. Hockemeyer, M. Mitalipova, R. Jaenisch, and R. A. Young, Cell Stem Cell **7**(2), 249–257 (2010).

[5] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin, and J. A. Thomson, Science **318**(5858), 1917–1920 (2007).

[6] D. H. Phanstiel, J. Brumbaugh, C. D. Wenger, S. Tian, M. D. Probasco, D. J. Bailey, D. L. Swaney, M. A. Tervo, J. M. Bolin, Victor Ruotti, Ron Stewart, James A. Thomson, and Joshua J. Coon, Nat Meth **8**(10), 821–827 (2011).

[7] M. C. N. Marchetto, G. W. Yeo, O. Kainohana, M. Marsala, F. H. Gage, and A. R. Muotri, PLoS ONE **4**(9), e7076 (2009).

[8] M. H. Chin, M. J. Mason, W. Xie, S. Volinia, M. Singer, C. Peterson, G. Ambartsumyan, O. Aimiuwu, L. Richter, J. Zhang, I. Khvorostov, V. Ott, M. Grunstein, N. Lavon, N. Benvenisty, C. M. Croce, A. T. Clark, T. Baxter, A. D. Pyle, M. A. Teitell, M. Pelegrini, K. Plath, and W. E. Lowry, Cell Stem Cell **5**(1), 111–123 (2009).

[9] C. Bock, E. Kiskinis, G. Verstappen, H. Gu, G. Boulting, Z. D. Smith, M. Ziller, G. F. Croft, M. W. Amoroso, D. H. Oakley, A. G. K. Eggan, and A. Meissner, Cell **144**(3), 439–452 (2011).

[10] P. Heraud and M. J. Tobin, Stem Cell Research **3**(1), 12–14 (2009).

[11] J. K. Pijanka, D. Kumar, T. Dale, I. Yousef, G. Parkes, V. Untereiner, Y. Yang, P. Dumas, D. Collins, M. Manfait, G. D. Sockalingum, N. R Forsyth, and J. Sule-Suso, Analyst **135**(12), 3126–3132 (2010).

[12] D. Ami, T. Neri, A. Natalello, P. Mereghetti, S. M. Doglia, M. Zanoni, M. Zuccotti, S. Garagna, and C. A. Redi, Biochimica et Biophysica Acta (BBA) – Molecular Cell Research **1783**(1), 98–106 (2008).

[13] M. J. Walsh, T. G. Fellous, A. Hammiche, W. R. Lin, N. J. Fullwood, O. Grude, F. Bahrami, J. M. Nicholson, M. Cotte, J. Susini, H. M. Pollock, M. Brittan,

P. L. Martin-Hirsch, M. R. Alison, and F. L. Martin, Stem Cells **26**(1), 108–118 (2008).

[14] C. Sandt, O. Féraud, N. Oudrhiri, M. L. Bonnet, M. C. Meunier, Y. Valogne, A. Bertrand, M. Raphaël, F. Griscelli, A. G. Turhan, P. Dumas, and A. Bennaceur-Griscelli, PLoS ONE **7**(4), e30743 (2012).

[15] P. Heraud, E. S. Ng, S. Caine, Q. C. Yu, C. Hirst, R. Mayberry, A. Bruce, B. R. Wood, D. McNaughton, E. G. Stanley, and A. G. Elefanty, Stem Cell Research **4**(2), 140–147 (2010).

[16] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones, Science **282**(5391), 1145–1147 (1998).

[17] M. Costa, K. Sourris, T. Hatzistavrou, A. G. Elefanty, and E. G. Stanley, Curr. Protoc. Stem Cell Biol **5**, 1C.1.11C.1.7. (2008)

[18] M. Richards, C. Y. Fong, W. K. Chan, P. C. Wong, and A. Bongso, Nat Biotech **20**(9), 933–936 (2002).

[19] M. Costa, M. Dottori, E. Ng, S. M. Hawes, K. Sourris, P. Jamshidi, M. F. Pera, A. G. Elefanty, and E. G. Stanley, Nat Meth **2**(4), 259–260 (2005).

[20] R. P. Davis, E. S. Ng, M. Costa, A. K. Mossman, K. Sourris, A. G. Elefanty, and E. G. Stanley, Blood **111**(4), 1876–1884 (2008).

[21] D. A. Elliott, S. R. Braam, K. Koutsis, E. S. Ng, R. Jenny, E. L. Lagerqvist, C. Biben, T. Hatzistavrou, C. E. Hirst, Q. C. Yu, R. J. P. Skelton, D. Ward-van Oostwaard, S. M Lim, O. Khammy, X. Li, S. M. Hawes, R. P. Davis, A. L. Goulburn, R. Passier, O. W. J. Prall, J. M. Haynes, C. W. Pouton, D. M. Kaye, C. L. Mummery, A. G. Elefanty, and E. G. Stanley, Nat Meth **8**(12), 1037–1040 (2011).

[22] J. Liu, H. Sumer, J. Leung, K. Upton, M. Dottori, A. Pebay, and P. J. Verma, Cell Transplant **20** (2), 193–203 (2011)

[23] E. S. Ng, R. P. Davis, L. Azzola, E. G. Stanley, and A. G. Elefanty, Blood **106**(5), 1601–1603 (2005).

[24] E. S. Ng, R. P. Davis, T. Hatzistavrou, E. G. Stanley, and A. G. Elefanty, in Current Protocols in Stem Cell Biology (John Wiley & Sons, Inc., 2007).

[25] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott, and F. L. Martin, Analyst (2012).

[26] H. Martens, J. P. Nielsen, and S. B. Engelsen, Analytical Chemistry **75**(3), 394–404 (2003).

[27] B. Mohlenhoff, M. Romeo, M. Diem, and B. R. Wood, Biophysical Journal **88**(5), 3635–3640 (2005).

[28] S. Wold, K. Esbensen, and P. Geladi, Chemometrics and Intelligent Laboratory Systems **2**(1–3), 37–52 (1987).

[29] S. Wold and M. Sjostrom, in Chemometrics: Theory and Application American Chemical Society, **52**, 243–282 (1977).

[30] P. Geladi, Journal of Chemometrics **2**(4), 231–246 (1988).

[31] K. H. Esbensen, D. Guyot, F. Westad, and L. P. Houmøller, Multivariate data analysis-in practice: An introduction to multivariate data analysis and experimental design (Multivariate Data Analysis, 2002).

[32] M. J. Abeyta, A. T. Clark, R. T. Rodriguez, M. S. Bodnar, R. A. Reijo Pera, and M. T. Firpo, Human Molecular Genetics **13**(6), 601–608 (2004).

[33] Z. Ghosh, K. D. Wilson, Y. Wu, S. Hu, T. Quertermous, and J. C. Wu, PLoS ONE **5**(2), e8975 (2010).

[34] R. Lister, M. Pelizzola, Y. S. Kida, R. D. Hawkins, J. R. Nery, G. Hon, J. Antosiewicz-Bourget, R. O'Malley, R. Castanon, S. Klugman, M. Downes, R. Yu, R. Stewart, B. Ren, J. A. Thomson, R. M. Evans, and J. R. Ecker, Nature **471**(7336), 68–73 (2011).

[35] Q. C. Yu, C. E. Hirst, M. Costa, E. S. Ng, J. V. Schiesser, K. Gertow, E. G. Stanley, and A. G. Elefanty, Blood **119**(26), 6243–6254 (2012).

[36] O. Yanes, J. Clark, D. Wong, G. Patti, A. Sánchezruiz, H. Benton, S. Trauger, C. Desponts, S. Ding, and G. Siuzdak, Nature Chemical Biology **6**(6), 411 (2010).

[37] W. Tanthanuch, K. Thumanu, C. Lorthongpanich, R. Parnpai, and P. Heraud, Journal of Molecular Structure **967**(1–3), 189–195 (2010).

[38] F. Haibach, A. Erlich, and E. Deutsch, SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, 803208–803208 (2011).

[39] D. Whelan, K. Bambery, P. Heraud, M. Tobin, M. Diem, D. McNaughton, and B. Wood, Nucleic Acids Research, **39**(13), 5439–5448 (2011).

[40] J. G. Kelly, T. Nakamura, S. Kinoshita, N. J. Fullwood, and F. L. Martin, Analyst **135**(12), 3120–3125 (2010).

[41] C. Hughes, M. Liew, A. Sachdeva, P. Bassan, P. Dumas, C. A. Hart, M. D. Brown, N. W. Clarke, and P. Gardner, Analyst **135**(12), 3133–3141 (2010).

# Analyst

**PAPER**

# Fourier transform infrared microspectroscopy reveals that tissue culture conditions affect the macromolecular phenotype of human embryonic stem cells†

Julie Cao,[ab] Elizabeth S. Ng,[ad] Don McNaughton,[b] Edouard G. Stanley,[ad] Andrew G. Elefanty,[ad] Mark J. Tobin[c] and Philip Heraud*[ab]

We employed Fourier transform infrared (FTIR) microspectroscopy to investigate the effects of different tissue culture environments on the FTIR spectra of undifferentiated human embryonic stem cells (hESCs) and their differentiated progeny. First we tested whether there were any possible spectral artifacts resulting from the use of transflectance measurements by comparing them with transmission measurements and found no evidence of these concluding that the lack of any differences resulted from the homogeneity of the dried cytospun cellular monolayers. We found that hESCs that were enzymatically passaged onto mouse embryonic fibroblasts (MEFs) in KOSR based hESC medium, hESCs enzymatically passaged onto Matrigel in mTESR medium and hESCs mechanically passaged onto MEFs in KOSR-based hESC medium, possessed unique FTIR spectroscopic signatures that reflect differences in their macromolecular chemistry. Further, these spectroscopic differences persisted even upon differentiation towards mesendodermal lineages. Our results suggest that FTIR microspectroscopy is a powerful, objective, measurement modality that complements existing methods for studying the phenotype of hESCs and their progeny, particularly changes induced by the cellular environment.

## Introduction

Owing to their unique ability to differentiate into the three embryonic germ layers,[1] human embryonic stem cells (hESCs) have been advanced as a potential source of cells for the replacement of disease-compromised cells and tissue. Since the initial derivation of hESCs from blastocysts,[1] cell culturing conditions have undergone multiple refinements. Today, hESC lines are either co-cultured with embryonic fibroblast feeder cells (feeder-dependent systems),[1-4] or in the absence of feeders (feeder-free systems).[5-7] An example of the latter system involves culturing onto Matrigel (Becton Dickinson, Bedford, MA), a basement membrane preparation extracted from a murine Englebreth-Holm-Swarm sarcoma. Stem cells grown using either platform are maintained in a range of growth media formulations and routinely passaged *via* either mechanical

dissociation,[1,8,9] or disaggregation methods that include both non-enzymatic (cell dissociation buffer, CDB) and enzymatic treatments (collagenase/trypsin, CT).

The diversity of culture methodologies has contributed to a range of morphological[8] and genotypic[10] disparities which may affect the success of stem cell maintenance[11] and differentiation outcomes. There is currently no data documenting how different methodologies affect the macromolecular phenotype of hESCs, as traditional stem cell biology techniques cannot provide this information. However, with the recent emergence of spectroscopic modalities such as Fourier transform infrared (FTIR) microspectroscopy in stem cell research[12] such information can now be easily attained and modeled using a variety of chemometric-based approaches.

FTIR microspectroscopy has provided new insights into a variety of different stem cells, including hESC, adult (tissue derived) and induced pluripotent stem cells (iPSC)[13-17] and their lineage committed cells. Our laboratory has employed it to discriminate undifferentiated hESCs from their mesendodermal progeny.[13] However, no studies to our knowledge have applied it to investigate the influence of the culturing environment on the macromolecular chemistry of hESCs. Since hESCs can be both genotypically and phenotypically affected by their growth conditions, we hypothesised that hESCs and their progeny, maintained under different conditions, will possess unique FTIR "spectroscopic signatures".

*a*Monash Immunology and Stem Cell Laboratories, Monash University, Building 75, STRIP 1, West Ring Road, Clayton, Victoria 3800, Australia. E-mail: ▮▮▮▮▮▮▮

*b*Centre for Biospectroscopy, The School of Chemistry, Monash University, Clayton, Victoria 3800, Australia

*c*Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia

*d*Murdoch Childrens Research Institute, The Royal Children's Hospital, Parkville, Victoria, 3052, Australia

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3an00321c

## Materials and methods

### Atomic force microscopy methodology

It has been reported recently that transflectance measurements may be affected by the electric field standing wave (EFSW),[18,19] which is claimed to give rise to distorted spectra and absorbance changes and thereby influencing the classification outcomes. The artifact has been reported to be most pronounced in transflectance measurements compared with transmission measurements. Since the phenomena is thought to be most problematic when there are differences in the thickness of compared samples, we measured the height of day 4 mesendodermal differentiated $MIXL1^{GFP/w}$ cells that were initially cultured on Matrigel in mTESR medium, and $MIXL1^{GFP/w}$ cells that were initially cultured on MEFs in KOSR-based hESC medium, on MirrIR slides using Atomic Force Microscopy (AFM). This comparison was chosen because this constituted the most significant comparison in the study in terms of the findings, and the differences in thickness between cells derived from the two conditions would allow us to determine whether the spectroscopic differences between the cells in each condition were of a physical (attributable to the EFSW phenomenon) or biochemical origin.

AFM measurements of cell height were made in tapping mode using a JPK Nanowizard 3 AFM. This instrument is equipped with capacitive sensors to ensure accurate reporting of height, $z$, and $x$-$y$ lateral distances. Cantilevers used were Bruker NCHV model tapping mode levers, with nominal resonant frequencies of 340 kHz and spring constants of 20–80 N m$^{-1}$ respectively. Imaging was performed with a set-point force of <1 nN. Sectional heights of cells were measured by taking line profiles (effectively a single $x$-$z$ fast-axis scan line from an image) that had clear areas of substrate on either side of a cell, and calculating the cell's highest point with respect to the substrate. Repeated imaging at different scan angles was performed to ensure that no sample damage or deformation occurred during measurement.

### Preparation of stem cells for FTIR microspectroscopy

**Testing the effects of different tissue culture methodologies on the FTIR spectroscopic signatures of human embryonic stem cells.** Four karyotypically normal hESC lines, HES3 (ref. 2) and H9,[1] MISCES-01 (ref. 20) and $MIXL1^{GFP/w21}$ were used for the various experiments in this study.

Three replicate cultures of the hESC lines HES3 (ref. 2) and H9,[1] and MISCES-01 (ref. 20) were either enzymatically passaged onto mouse embryonic fibroblasts (MEFs) using TrypLE Select (Invitrogen) in Knock Out Serum Replacer (KOSR)-based hESC medium,[22] mechanically passaged onto MEFs in KOSR medium or enzymatically passaged onto Matrigel (Becton Dickinson, Bedford, MA), in mTESR medium (Stem Cell Technologies, Vancouver, BC, Canada), using methods previously described.[23]

**Testing the persistence of culture condition induced spectroscopic effects upon differentiation towards mesendodermal lineages.** Three replicate cultures of the hESC line $MIXL1^{GFP/w}$, a genetically modified hESC line expressing green fluorescent protein (GFP) from the $MIXL1$ locus,[21] were differentiated toward mesendoderm or ectoderm as monolayers, in 6 well tissue culture plates. Differentiation towards mesendoderm was performed in serum-free APEL medium[24] supplemented with 20 ng ml$^{-1}$ of BMP4, 50 ng ml$^{-1}$ of Activin A, 10 ng ml$^{-1}$ of FGF2 and10 ng ml$^{-1}$ of VEGF. After 3 and 4 days of differentiation, cells were analysed for induction of GFP from the MIXL1 locus (denoted MIXL1-GFP) by flow cytometry[21,25] to confirm that differentiation had occurred, before being prepared as previously described for FTIR microspectroscopy.[13]

**Testing the effects of changing culture conditions on the FTIR spectroscopic signatures of hES cells.** For the adaptation experiments, $MIXL1^{GFP/w}$ cells were enzymatically passaged three times onto Matrigel in mTESR medium before being enzymatically passaged onto MEFs in KOSR medium. hESCs were maintained for 5 passages under the same conditions, and during each passage, 120 000 cells were harvested and prepared as previously described for FTIR microspectroscopy above.

**Testing the effects of the growth medium on the FTIR spectroscopic signatures of hESC cells.** The effect of different growth media on cell spectra was tested by co-culturing hESCs with MEFs in either KOSR or mTESR medium. $MIXL1^{GFP/w}$ cells were cultured in 6 well plates, on top of low density MEFs in either KOSR or mTESR medium. hESCs were maintained for 5 passages in the same conditions, and during each passage, 120 000 of the cells were collected and prepared as previously described for FTIR microspectroscopy above.

### Transmission *versus* transflectance comparison

To test whether the recently reported spectral artifact caused by the EFSW effect may have affected our data acquired from cells deposited on the transflectance substrate, we compared the spectral datasets of two replicates of $MIXL1^{GFP/w}$ cells, that had been co-cultured with MEFs in KOSR based hESC medium acquired in both experimental modes. For the former studies, 120 000 cells were collected and cytospun onto MirrIR slides as previously described. For the transmission experiments, the same number of cells were cytospun onto 0.5 mm thick CaF$_2$ windows.

### Focal plane array (FPA) FTIR imaging

FPA FTIR infrared images were acquired at the Centre for Biospectroscopy at Monash University in Melbourne, Australia, in transflectance mode using an Agilent Technologies FTIR spectrometer (Model FTS 7000; Agilent Technologies Inc., Palo Alto, CA, USA) coupled to an infrared microscope (model 600 UMA; Agilent Technologies) using a 15× cassegrain objective coupled with a 64 × 64 pixel MCT liquid nitrogen cooled FPA detector. The areas of the deposits chosen for spectral acquisiton were selected *via* visual examination by light microscopy where cells were found to be in a true monolayer, and were generally around the edges of the deposits. FTIR spectra were acquired with 128 co-added scans, at 8 cm$^{-1}$ spectral resolution, with the binning of signals from groups of 4 adjacent pixels on the

Paper

Analyst

FPA resulting in each spectrum corresponding to an area of 11 μm × 11 μm on the sample plane, the approximate dimensions of single cells in the dried monolayer A Happ-Genzel apodisation function was used to truncate the interferogram, and a zero-filling factor of 2 was employed.

### Data preprocessing for multivariate data analysis

The infrared images were taken into Cytospec 1.03 FTIR imaging software (Cytospec Inc., NY, and USA) in order to extract the spectral data that were deemed suitable for further analysis. The spectra first underwent a quality test whereby spectra that had maximum absorbance values outside the range of 0.6–0.8, arising from sample regions devoid of cells or where cells were clumped and overlaid, were rejected. This was to ensure that measurements with a good signal to noise ratio and spectral signal within the linear range of the detector's response were used in further steps. The spectra which passed the quality test underwent further preprocessing in The Unscrambler v 10.1 (Camo, Oslo, Norway) software package. Here, they were converted to their second derivatives with a Savitsky-Golay algorithm using 9 smoothing points in order to minimize baseline effects and to resolve spectral components that would otherwise be overlapping in the underivatised spectra. This procedure results in the inversion of the peaks in the underivatised spectra and so appear as minima bands. Extended Multiplicative Signal Correction (EMSC),[26] which also normalizes the data, was carried out using the following spectral ranges: 3050–2800, 1770–1730, and 1470–900 cm$^{-1}$. These ranges were selected because they contain key biological absorption bands and therefore provide the best discrimination of different sample types. The amide I and amide II regions were omitted from the analysis as these regions are thought to be influenced by the effects of resonant Mie scattering, which may have confounded our interpretation of the spectra.

### Partial least squares discriminant analysis (PLS-DA)

The dataset was randomly sorted into 2 new datasets comprising 2 thirds and 1 third of the spectra and used as the calibration and independent test datasets respectively for classification of spectra using Partial Least Squares Discriminant Analysis (PLS-DA).[27] The calibration data matrix employed for PLS-DA consisted of the spectral dataset (multivariate $X$) and two $Y$ variables with integer values of 0 or 1 coding for the each of the two modelled spectral classes. Data outliers in the calibration set were identified and removed by the examination of residual influence plots. The number of outliers constituted less than 1% of the original number of spectra for all datasets. Classification of the dataset was then carried out by predicting a $Y$ value for each spectrum in the independent validation using PLSR models that had been generated from the calibration sets. Correct classification of each class was arbitrarily assigned to samples with predicted $Y > 0.5$ for respective spectra. PLS scores and loadings plots were used to determine which spectral regions contributed most to the ability to classify spectra using PLS-DA.

## Results and discussion

### Testing for possible biases introduced by the transflectance measurement

To determine whether there were differences between spectra acquired in transflectance and transmission, which were attributable to EFSW effects, spectra were acquired from $MIXL1^{GFP/w}$ cells in both measurement modes and analysed using PLS-DA. No clear clustering was seen in the PLS-DA scores plots of FTIR spectral clusters derived from both replicates of $MIXL1^{GFP/w}$ cells that were grown on MEFs in KOSR-based medium, suggesting that spectra from the two types of measurement were highly similar (ESI Fig. 1A†). Along PLS Factor 1, the loadings bands that explained the greatest variance in the dataset were the bands at ~1149 cm$^{-1}$, ~1076 cm$^{-1}$, ~1014 cm$^{-1}$, and ~991 cm$^{-1}$ (ESI Fig. 1B†). This was the type of variation typically seen between spectra from cells drawn from the same populations.

PLS-DA classification of the independent test spectra was poor, particularly with the two transflectance experiment derived datasets which had sensitivities of 11% and ~34% at for replicates 1 and 2 respectively. Discrimination of spectra derived from the transmission experiments was better, but still very low, with sensitivities of 54% for replicate 1 and 43% for replicate 2.

The mean thickness of the day 4 mesendodermal differentiated $MIXL1^{GFP/w}$ cells originally cultured on MEFs in KOSR-based medium were 0.93 μm ± 0.05, and 0.91 μm ± 0.07 for the cells originally cultured on Matrigel in mTESR based medium. The $p$ value from the two sample independent $t$-test was 0.80. This result showed that the cells in the monolayer deposits derived from both conditions had no statistically significant difference in thickness. In any case, differences were well below the 0.5 μm, which was the sample height difference were at which EFSW effects have been reported to be detectable.[19]

### Both cellular morphology and spectroscopic signature of hESCs are influenced by culture conditions

Mechanically passaged hESCs formed discrete colonies of tightly packed cells with scant cytoplasm and prominent nucleoli (Fig. 1A and D). The morphology of enzymatically passaged cells was generally similar, although cells passaged on Matrigel in mTeSR medium appeared smaller than those passaged on MEFs as the cultures became denser (compare Fig. 1E with F). Also, for hESCs passaged by either enzymatic protocol, cultures became confluent with time, blurring any separation between colonies.

We compared the spectral profiles of hESCs grown under the three different culture conditions. Examination of results from this experiment clearly showed that the different culture conditions affected the spectro-phenotype of all the hESC lines tested, with PLS discriminant analysis (PLS-DA) allowing classification of independent test spectra with very high levels of accuracy (>92% Fig. 2A–C). This was in accordance with both the average second derivative spectra (Fig. 3A–C) and distinct
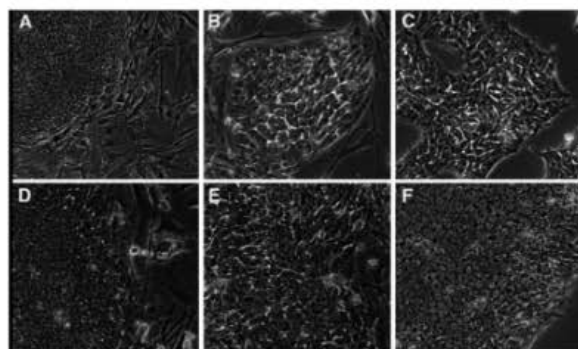
Analyst

Analyst

Paper



**Fig. 1** Brightfield images of human embryonic stem cells propagated using (A and D) manual passaging, (B and E) enzymatic passaging on mouse embryonic fibroblast feeder cell layers and (C and F) enzymatic passaging on Matrigel coated plates using mTeSR1 medium. (A) Low power view of part of a manual passaged colony. (B and C) Subconfluent and (C and F) confluent cultures of enzymatically passaged cultures imaged at the same magnification. Original magnifications (A) ×50, (B–F) ×100.
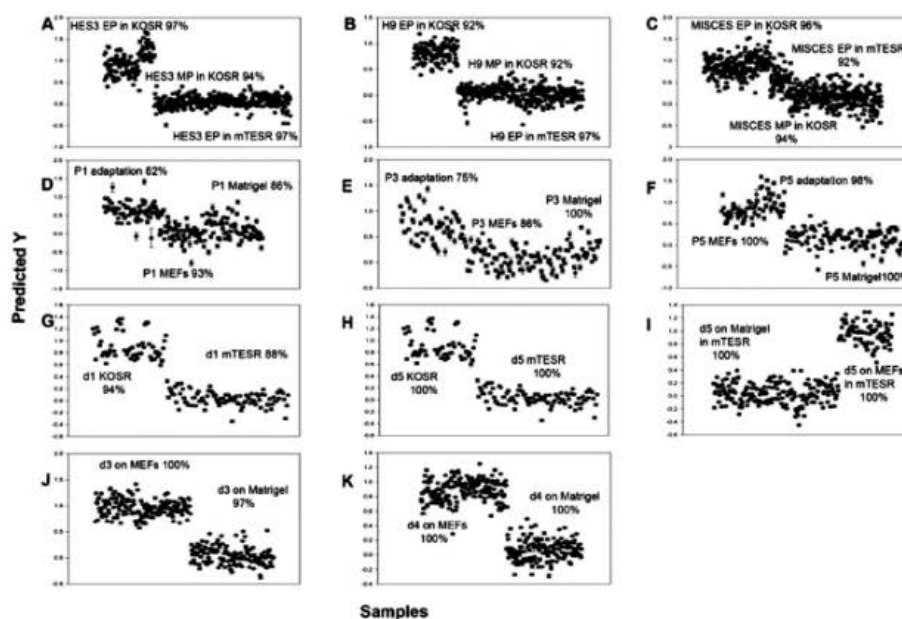


**Fig. 2** Predicted Y values from PLS-DA modelling for the calibration and validation set resulting from the PLS-DA discriminant analyses of all of the previously described experiments in Fig. 3–8.

separation of clusters in the PLS scores plots (Fig. 4A, C and E), with HES3 and H9 showing very similar responses compared with the MISCES-01 line.

In general, cells enzymatically passaged onto Matrigel in mTESR medium had spectra consistent with higher concentrations of amino acids and phosphorylated molecules, with an
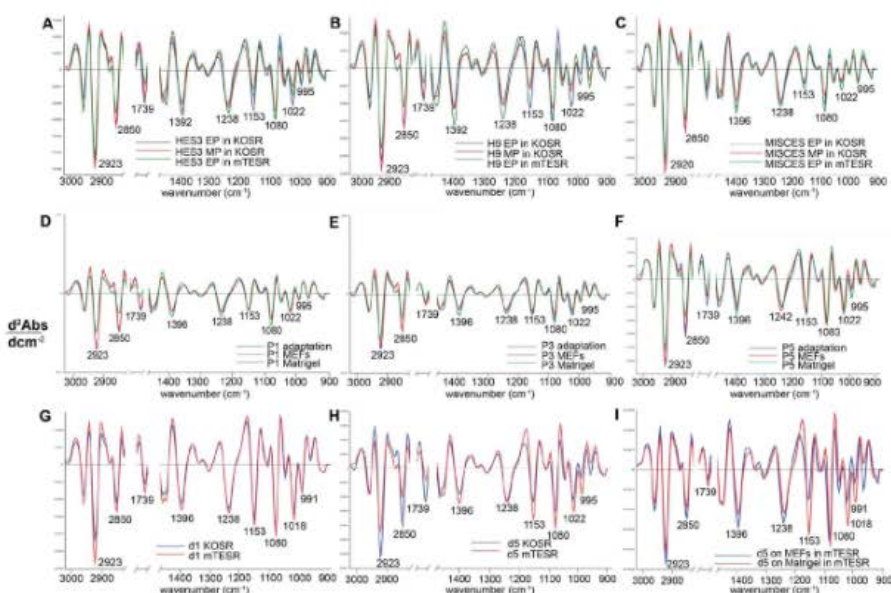
Analyst

**Fig. 3** Average second derivative spectra for the (A) HES3, (B) H9 and (C) MISCES-01 lines enzymatically passaged onto MEFs in KOSR medium, mechanically passaged onto MEFs in KOSR medium and enzymatically passaged onto Matrigel in mTESR medium. Average second derivative spectra for *MIXL1*$^{GFP/w}$ lines previously cultured on Matrigel in mTESR media, and transferred onto MEFs in KOSR media (adaptation group), continuously cultured on MEFs in KOSR medium (MEFs control) and continuously cultured on Matrigel in mTESR medium, after 1 passage (D), 3 passages (E) and 5 passages (F) back into their assigned conditions. Average second derivative spectra for *MIXL1*$^{GFP/w}$ cells in co culture with MEFs and supplemented with either KOSR medium or mTESR medium after 1 passage (G) and 5 passages (H). Average second derivative spectra of *MIXL1*$^{GFP/w}$ co-cultured with MEFs and maintained for 5 passages in mTESR medium *versus* the same line, after 5 passages on Matrigel, in mTESR medium.

intense carboxylate band at ~1392 cm$^{-1}$, and bands from PO$_2^-$ stretching vibrations at ~1238 cm$^{-1}$ and ~1080 cm$^{-1}$, respectively, seen in the average second derivative spectra (Fig. 3A–C). This was also evident by the clustering patterns along Factor 1 (Fig. 4C and E) or Factor 2 (Fig. 4A) in PLS scores plots associated with prominent loadings for these bands (Fig. 4B, D and F and Table 1). By contrast, HES3 and H9 hESCs enzymatically passaged onto MEFs in KOSR medium contained higher levels of carbohydrates, as shown by intense negative peaks at ~1153 cm$^{-1}$ and ~1022 cm$^{-1}$ (Fig. 3A–C). Spectral clusters from cells that were enzymatically passaged onto MEFs in KOSR medium clearly separated along Factor 1 (Fig. 4A and C), explained by strong positive loadings for these bands ascribed to carbohydrates (Fig. 4B and D). For all three cell lines, mechanically dissociated cells had higher lipid concentrations, as indicated by more intense $\nu_s$CH$_2$, $\nu_{as}$CH$_2$ stretching modes in the average second derivative spectra at ~2850 cm$^{-1}$ and 2920 cm$^{-1}$ respectively (Fig. 3), and separation of clusters of spectra from mechanically dissociated cells along Factor 1 with the MISCES-01 hESC line (Fig. 4E) or along Factor 2 with the HES3 and H9 hESC lines (Fig. 4A and C).

The effect of the culture conditions appeared to be least pronounced with the MISCES-01 line, as demonstrated by the less prominent separation of spectral clusters from the three treatment cohorts in scores plots. This contrasted with the excellent separation of spectra observed between clusters of spectra from similarly treated samples in the HES3 and H9 groups. The data from this initial study convincingly demonstrated that different tissue culture platforms affected the IR spectroscopic phenotypes of human embryonic stem cells, and that the extent of these effects appeared to be dependent on the individual hESC line used.

Our findings are in agreement with other studies which have reported alterations in hESCs caused by the culturing environment.[6,8–10,23] However, those studies focused on changes in morphology, gene expression, or cell surface antigen expression, rather than macromolecular information. Here we present the first study showing environmentally induced changes in hESCs occurring at the biochemical level. For instance, different media and substrates have both been found to influence proliferation rates and cellular attachment abilities.[6] Subtle morphological differences between cells cultured *via*
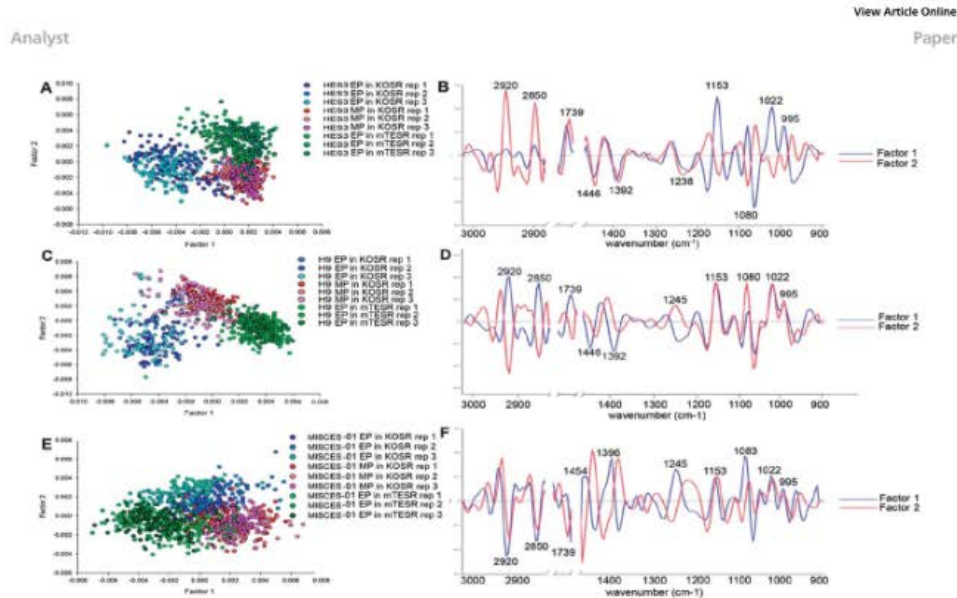
Analyst

Paper



**Fig. 4** (A) PLS-DA scores plots from the analysis of the spectral data from the human embryonic stem cell line, HES3, enzymatically passaged onto MEFs in KOSR medium, mechanically passaged, and enzymatically passaged onto Matrigel in mTESR medium, and (B) PLS-DA loadings plots used to explain the spectral differences explaining the clustering observed in the previous scores plot. (C) PLS-DA scores plot from the analysis of the spectral data from the human embryonic stem cell line, H9, enzymatically passaged onto MEFs in KOSR medium, mechanically passaged onto MEFs in KOSR medium, and enzymatically passaged onto Matrigel in mTESR medium, and (D) PLS-DA loadings plots. (E) PLS-DA scores plot from the analysis of the spectral data from the human embryonic stem cell line, MISCES-01, enzymatically passaged onto MEFs in KOSR medium, mechanically passaged onto MEFs in KOSR medium, and enzymatically passaged onto Matrigel in mTESR medium and (F) PLS-DA loadings plots.

different methods have also been consistently observed in our laboratory (Fig. 1), with Matrigel-based culture systems producing smaller cells than feeder-based systems. Further, the choice of passaging technique has also been found to significantly influence cellular genotype and phenotype. Bulk culture passaging methods,[1,8,9] which include non-enzymatic (cell dissociation buffer, CDB) and enzymatic passaging (collagenase/trypsin, CT) promote the development of karyotypic abnormalities in some late passage embryonic stem cell lines.[10] In contrast, chromosomal alterations have been found to either not occur or occur less frequently,[10,23] after extended periods of mechanical passaging.

### Dependence of phenotype on growth media and substrate

To determine the effect on phenotype of the growth medium alone, we cultured hESCs on the same substrate (MEFs) but in different growth media (KOSR based or mTESR). PLS-DA classification showed the *MIXL1*[GFP/w] cells that were grown in KOSR-based medium could be correctly discriminated from those grown in mTESR medium with 94% accuracy. Furthermore, the group that was grown in mTESR medium could also be discriminated from the KOSR treated group with 88% accuracy after one passage (Fig. 2G). In addition, after 5 passages, these spectroscopic differences became more distinct as verified by

the improved classification outcomes with greater separation of the spectral clusters on the PLS scores plots. Average spectra (Fig. 3G and H) and PLS scores and loadings plots (Fig. 6A and C and Table 3) indicated that cells grown in mTESR medium for 1 and 5 passages had higher levels of amino acids, phosphorylated molecules and carbohydrates compared to the cells cultured in KOSR-based medium.

Nevertheless, when we then compared the spectra of the hESCs grown in mTESR medium on MEFs with previously acquired spectra of the same cells grown in mTESR medium on a Matrigel substrate, we found significant differences between the two treatment classes, which suggested that the growth substrate also contributed to the observed phenotypic differences between hESCs grown in the different culture systems (Fig. 6E and F).

This finding that the substrate type affected the cell phenotype was verified in further experiments whereby spectra from *MIXL1*[GFP/w] hESCs grown in mTESR medium and cultured either on MEFs or Matrigel were compared. Classification of independent test spectra by PLS-DA could distinguish spectra from each treatment group with 100% accuracy (Fig. 2). This was confirmed by PLS scores plots that revealed an excellent separation of clusters of spectra representing cells cultured on feeder layers and on Matrigel along Factor 1 (Fig. 6E). Loadings plots showed that carbohydrate (bands at ~1149 cm$^{-1}$,
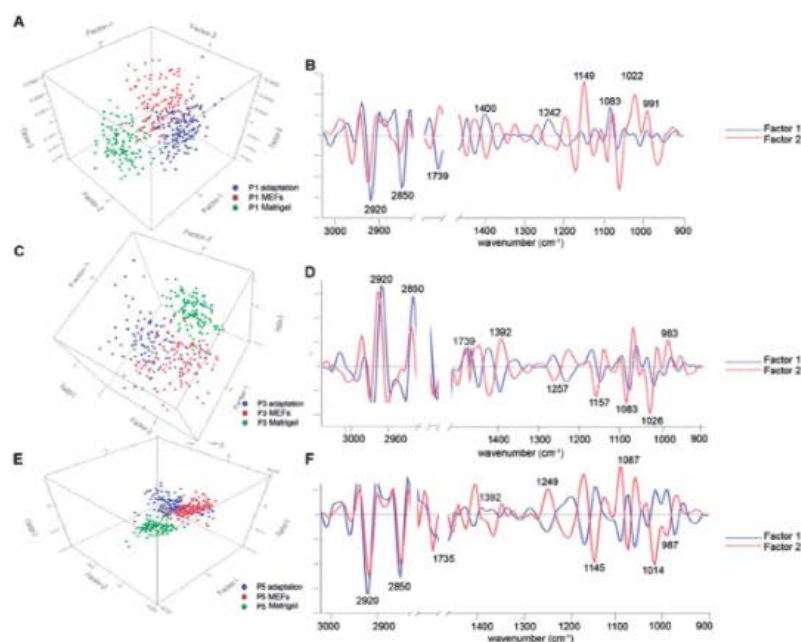
**Fig. 5** PLS-DA scores plots (A), (C) and (E) and loadings plots (B), (D) and (F) from the analysis of spectra acquired from *MIXL1*^GFP/w cells previously enzymatically passaged onto Matrigel in mTESR medium and readapted to co culture with MEFs in KOSR medium, compared with spectra of *MIXL1*^GFP/w cells continuously maintained on MEFs in KOSR medium, and *MIXL1*^GFP/w cells continuously Matrigel. PLS-DA analysis was performed using FTIR spectra acquired of the cells following 1 passage (A) and (B), 3 passages (C) and (D), and 5 passages (E) and (F) back onto the same growth conditions.

~1080 cm$^{-1}$, ~1018 cm$^{-1}$ and ~991 cm$^{-1}$) levels were higher in cells grown on mouse embryonic fibroblast feeder cells (MEFs) compared to those grown on Matrigel (Fig. 6F and Table 5). In contrast to cells grown on MEFs in KOSR medium, those grown on MEFs in mTESR medium had lower lipid levels compared to hESCs grown on Matrigel (Fig. 3 compared with Fig. 6E and F). This data correlated with differences seen in the average second derivative spectra (Fig. 3A–C compared with 3H).

Since the growth medium, substrate and dissociation methods had such a profound influence on cell phenotype, we then wanted to determine if this influence persisted following several passages onto different tissue culture platforms. We therefore compared the spectra of *MIXL1*^GFP/w hESCs cultured on Matrigel and then enzymatically passaged onto MEFs in KOSR medium, to spectra of hESCs continuously cultured on MEFs in KOSR medium, and to those continuously cultured on Matrigel in mTESR medium. Results from the PLS-DA analysis on the independent validation set showed that cells originally enzymatically passaged onto Matrigel in mTESR medium and subsequently co-cultured with MEFs in KOSR medium, maintained a phenotypic distinctness from cells that had been continuously cultured on MEFs, even after five days in the new

environment. PLS-DA (Fig. 2D–F): showed that these cells were classified as distinct from the continuously passaged cells with 97% correct discrimination after day 1 of growth in the new conditions; 75% correct after 3 days; and 98% correct after 5 days. However, the phenotype of the cells that were continuously passaged onto Matrigel in mTESR medium could be distinguished from both hESCs continuously passaged onto feeder layers in KOSR medium and from hESCs which had been transferred from a Matrigel-based to MEF dependent conditions with 100% accuracy at the two later time points (Fig. 2). This suggested that although some adaptation to the new environment had occurred, the original Matrigel-based growth environment had some continuing effect on phenotype even after several passages.

Average second derivative spectra (Fig. 3D–F) and PLS scores plots (Fig. 5B, D and F) showed that the major spectroscopic differences between cells enzymatically passaged continuously onto Matrigel in mTESR medium and the other treatment groups was lower absorbance for lipid components at ~2920 cm$^{-1}$, ~2850 cm$^{-1}$ and ~1739 cm$^{-1}$ at all measured time points. PLS-DA indicated that cells passaged onto Matrigel in mTESR medium and subsequently passaged onto MEFs in
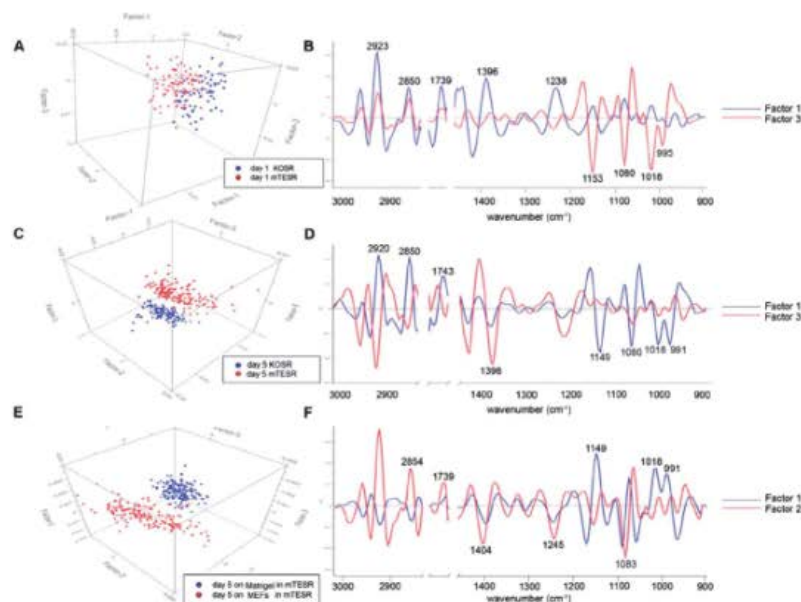
Analyst

Paper



**Fig. 6** PLS-DA scores plots (A) and (C) and loadings plots (B), (D) from the analysis of spectra acquired from *MIXL1*^GFP/w in co culture with MEFs and supplemented with either KOSR medium or mTESR medium. PLS-DA analysis was performed using FTIR spectra acquired of cells following 1 passage (A) and (B), and 5 passages (C) and (D), in their assigned growth conditions. PLS-DA scores plots (E) and corresponding loadings plot (F) from the comparison made between the day 5 *MIXL1*^GFP/w co-cultured with MEFs, in mTESR as shown in (C) and (D) and *MIXL1*^GFP/w cells in mTESR medium and maintained on a Matrigel substrate after 5 days.

KOSR medium were consistently different from hESCs continuously grown on MEFs in terms of carbohydrate absorbance, with loadings plots indicating lower levels in the cells grown continuously on feeder layers at all time points (C–O stretching bands at 1150 cm$^{-1}$ and 1022 cm$^{-1}$; Fig. 5). Further, spectra from the hESCs grown on MEFs initially had less lipid than those recently adapted to these conditions (days 1 and 3) as indicated by PLS loadings for bands at 2920 cm$^{-1}$ and 2850 cm$^{-1}$, and then higher lipid levels at the last time point (day 5). These differences are summarized in Table 2.

### The effect of culture conditions on the phenotype of differentiated cells

Upon observing culture condition influenced "spectroscopic signatures" in undifferentiated hESCs, we wished to determine the heritability of these spectroscopic differences in lineage committed progeny. To verify that the cells had differentiated to mesendodermal lineages, flow cytometry was performed on the *MIXL1*^GFP/w cells after 3 and 4 days of differentiation in BMP4/Act A containing medium (Fig. 7A–F). Undifferentiated hESCs, which were used as the control for these experiments, are known to express a variety of stem cell surface antigens such as EpCAM and do not express GFP from the *MIXL1* locus. By 4 days

of differentiation, some of the *MIXL1*^GFP/w hESCs had down-regulated cell surface marker EpCAM and the majority expressed MIXL1-GFP and platelet derived growth factor PDGFRα, an early mesendodermal marker, whereas undifferentiated cells were EpCAM positive but negative for MIXL1 and PDGFRα expression.

Surprisingly, even after 4 days of differentiation, and despite the flow cytometry data showing highly comparable differentiation kinetics, the phenotypic effects of the initial growth conditions were not entirely eliminated. PLS-DA classification could discriminate between spectra from day 3 progeny derived from hESCs maintained under the two growth conditions with greater than 97% accuracy (Fig. 2J). After 4 days of differentiation, both groups could be discriminated with 100% accuracy (Fig. 2K). PLS-DA scores plots at day 3 and day 4 (Fig. 8A and C) showed excellent separation along Factor 1 between clusters of spectra from the differentiated cells. Loadings values (Fig. 8B and D and Table 4) revealed that spectra of progeny derived from *MIXL1*^GFP/w hESCs maintained in mTESR medium on a Matrigel substrate, from both time points, exhibited more intense IR absorbances at ~2923 cm$^{-1}$, ~2850 cm$^{-1}$ and ~1739 cm$^{-1}$ from lipids, and at ~1396 cm$^{-1}$ from the COO$^-$ stretching vibrations of amino acid side chains, and at ~1238 cm$^{-1}$ from phosphorylated molecules. Conversely,

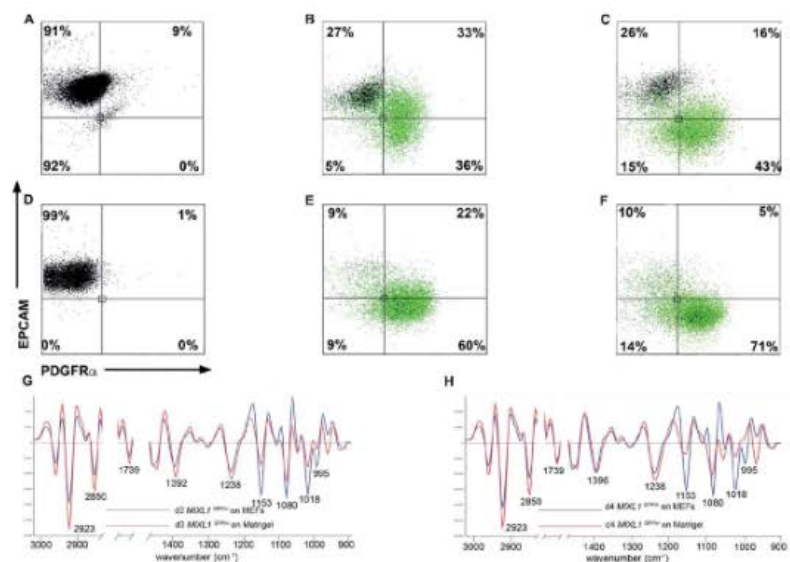Paper                                                                 Analyst



**Fig. 7** Flow cytometric analysis results of undifferentiated *MIXL1*[GFP/w] cells maintained on MEFs in KOSR medium (A) and differentiated in BMP4 ACTIVIN A towards mesendodermal lineages and stained for expression of EPCAM and PDGFRα at day 3 (B) and at day 4 (C). Percentages of cells within the indicated quadrants or regions are shown. Panels (D–F) show the flow cytometric analysis results of undifferentiated *MIXL1*[GFP/w] cells maintained on Matrigel in mTESR medium (D) and differentiated in BMP4 ACTIVIN A and stained for expression of EPCAM and PDGFRα at day 3 (E) and at day 4 (F). Percentages of cells within the indicated quadrants or regions are shown. Panels (G–H) show the average second derivative spectra for *MIXL1*[GFP/w] cells after 3 days (G) and 4 days (H) of differentiation with BMP4 ACTIVIN A.



**Fig. 8** PLS-DA scores plots (A), and loadings plots (B) of spectra acquired from *MIXL1*[GFP/w] after 3 days of differentiation towards mesendodermal lineages using BMP4 ACTI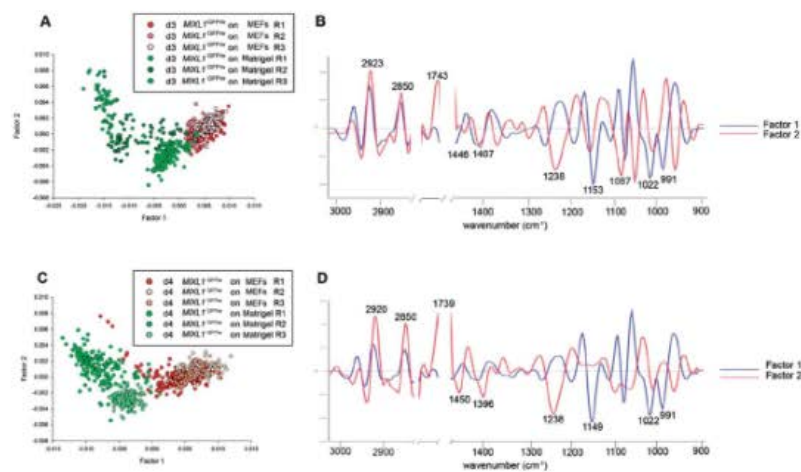VIN A. PLS-DA scores plots (C), and loadings plots (D) of spectra acquired from *MIXL1*[GFP/w] after 4 days of differentiation towards mesendodermal lineages.

Analyst

**Table 1** Band maxima of the human embryonic stem cell lines HES3, H9 and MISCES maintained on MEFs in KOSR medium and enzymatically passaged, maintained on MEFs in KOSR medium and mechanically passaged, and maintained on Matrigel and enzymatically passaged

| Band maxima 2$^{nd}$ derivative spectra (cm$^{-1}$) | | | PC1 loadings (cm$^{-1}$) | PC2 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|---|---|
| HES3, H9 and MISCES lines | | | | | |
| Enzymatically passaged onto MEFs in KOSR | Mechanically passaged onto MEFs in KOSR | Enzymatically passaged onto Matrigel in mTESR | HES3, H9 and MISCES lines | HES3, H9 and MISCES lines | HES3, H9 and MISCES lines |
| 2923, 2927, 2923 | 2923, 2923, 2923 | 2923, 2923, 2923 | 2920, 2920, 2920 | 2920, 2920, 2916 | C–H asymmetrical stretching of lipid groups |
| 2850, 2850, 2850 | 2850, 2850, 2850 | 2850, 2850, 2850 | 2850, 2850, 2850 | 2850, 2850, 2846 | C–H symmetrical stretching of lipid groups |
| 1743, 1739, 1739 | 1739, 1739, 1739 | 1739, 1739, 1739 | 1747, 1739, 1731 | 1739, 1735, 1739 | C=O stretching of lipid esters |
| 1396, 1396, 1396 | 1396, 1396, 1392 | 1392, 1396, 1392 | 1392, 1392, 1396 | 1388, 1404, 1484 | COO$^-$ stretching vibrations of amino acid side chains |
| 1242, 1224, 1238 | 1238, 1238, 1238 | 1238, 1238, 1238 | 1234, 1238, 1234 | 1238, 1245, 1253 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1080, 1080, 1083 | 1083, 1080, 1080 | 1080, 1083, 1080 | 1091, 1091, 1091 | 1080, 1080, 1091 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1153, 1153, 1153 | 1157, 1153, 1153 | 1153, 1153, 1153 | 1153, 1153, 1153 | 1149, 1153, 1149 | C–O vibrations from glycogen and other carbohydrates |
| 1022, 1022, 1022 | 1022, 1022, 1022 | 1022, 1022, 1022 | 1022, 1018, 1022 | 1018, 1022, 1018 | C–O vibrations from glycogen and other carbohydrates |
| 995, 995, 995 | 995, 995, 995 | 995, 995, 995 | 995, 995, 995 | 991, 995, 991 | C–O stretch from RNA ribose chains and carbohydrates |

**Table 2** Band maxima of the human embryonic stem cell line *MIXL1*$^{GFP/w}$ previously cultured on Matrigel in mTESR media and transferred over to culture with feeders, after 1, 3 and 5 passages. Band maxima are also shown of the two control groups, which consisted of *MIXL1*$^{GFP/w}$ cells that had been continuously co-cultured with feeders in KOSR medium (feeder control), and *MIXL1*$^{GFP/w}$ cells that had been continuously cultured on Matrigel in mTESR medium (Matrigel control)

| Band maxima 2$^{nd}$ derivative spectra (cm$^{-1}$) | | | | | |
|---|---|---|---|---|---|
| P1, P3, P5 adaptation | P1, P3, P5 on MEFs | P1, P3, P5 on Matrigel | Factor 1 loadings (cm$^{-1}$) | Factor 2 loadings (cm$^{-1}$) | Band assignments |
| 2923, 2923, 2923 | 2923, 2923, 2923 | 2923, 2923, 2923 | 2920, 2920, 2920 | 2927, 2923, 2920 | C–H asymmetrical stretching of lipid groups |
| 2850, 2850, 2850 | 2850, 2850, 2850 | 2850, 2850, 2850 | 2850, 2850, 2850 | 2854, 2850, 2850 | C–H symmetrical stretching of lipid groups |
| 1739, 1739, 1739 | 1739, 1739, 1739 | 1739, 1739, 1739 | 1739, 1731, 1735 | 1747, 1739, 1739 | C=O stretching of lipid esters |
| 1396, 1396, 1396 | 1396, 1396, 1396 | 1396, 1396, 1396 | 1400, 1396, 1392 | 1388, 1396, 1404 | COO$^-$ stretching vibrations of amino acid side chains |
| 1238, 1238, 1238 | 1238, 1238, 1242 | 1238, 1238, 1238 | 1242, 1238, 1249 | 1249, 1238, 1245 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1080, 1080, 1080 | 1080, 1080, 1083 | 1083, 1080, 1080 | 1083, 1080, 1087 | 1091, 1080, 1087 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1153, 1153, 1153 | 1153, 1153, 1153 | 1153, 1153, 1153 | 1157, 1149, 1145 | 1145, 1153, 1145 | C–O vibrations from glycogen and other carbohydrates |
| 1022, 1022, 1022 | 1022, 1022, 1022 | 1022, 1022, 1018 | 1026, 1018, 1014 | 1014, 1022, 1018 | C–O vibrations from glycogen and other carbohydrates |
| 995, 995, 995 | 995, 995, 995 | 995, 995, 995 | 991, 991, 991 | 987, 995, 987 | C–O stretch from RNA ribose chains and carbohydrates |

spectra from hESC progeny grown on MEFs had higher absorbances from glycogen and other carbohydrates (bands at ~1153 cm$^{-1}$, ~1080 cm$^{-1}$ and ~1018 cm$^{-1}$). This result was noteworthy, as we had previously found that undifferentiated MEF-derived *MIXL1*$^{GFP/w}$ cells had higher lipid absorbances than the Matrigel-derived hESCs. Average second derivative

**Table 3** Band maxima of the human embryonic stem cell line *MIXL1*[GFP/w] co-cultured with feeders and maintained in either KOSR medium or mTESR medium, after 1 and 5 passaging events

| Band maxima 2nd derivative spectra (cm⁻¹) | | | | |
|---|---|---|---|---|
| Day 1 and day 5 in KOSR based medium | Day 1 and day 5 in mTESR medium | Factor 1 loadings (cm⁻¹) | Factor 3 loadings (cm⁻¹) | Band assignments |
| 2923, 2923 | 2923, 2923 | 2923, 2920 | 2923, 2923 | C–H asymmetrical stretching of lipid groups |
| 2850, 2850 | 2850, 2850 | 2854, 2850 | 2850, 2854 | C–H symmetrical stretching of lipid groups |
| 1739, 1739 | 1739, 1739 | 1739, 1743 | 1739, 1739 | C=O stretching of lipid esters |
| 1396, 1396 | 1396, 1396 | 1392, 1396 | 1384, 1396 | COO⁻ stretching vibrations of amino acid side chains |
| 1238, 1238 | 1238, 1238 | 1234, 1238 | 1242, 1234 | P=O asymmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1080, 1080 | 1080, 1080 | 1080, 1080 | 1080, 1080 | P=O symmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1153, 1153 | 1153, 1153 | 1149, 1149 | 1153, 1157 | C–O vibrations from glycogen and other carbohydrates |
| 1018, 1022 | 1018, 1018 | 1018, 1018 | 1018, 1018 | C–O vibrations from glycogen and other carbohydrates |
| 991, 995 | 991, 995 | 983, 991 | 995, 995 | C–O stretch from RNA ribose chain |

**Table 4** Band maxima of the human embryonic stem cell line *MIXL1*[GFP/w] for spectra from the hESC line *MIXL1*[GFP/w], differentiated towards mesendodermal lineages in BMP4 ACTIVIN A supplemented differentiation medium

| Band maxima 2nd derivative spectra (cm⁻¹) | | PC1 loadings (cm⁻¹) | PC2 loadings (cm⁻¹) | Band assignments |
|---|---|---|---|---|
| Day 3 and day 4 *MIXL1*[GFP/w] on MEFs in KOSR based medium | Day 3 and day 4 *MIXL1*[GFP/w] on Matrigel in mTESR medium | Day 3 and day 4 *MIXL1*[GFP/w] | Day 3 and day 4 *MIXL1*[GFP/w] | Day 3 and day 4 *MIXL1*[GFP/w] |
| 2920, 2923 | 2923, 2923 | 2923, 2923 | 2920, 2920 | C–H asymmetrical stretching of lipid groups |
| 2850, 2850 | 2850, 2850 | 2854, 2850 | 2850, 2850 | C–H symmetrical stretching of lipid groups |
| 1739, 1739 | 1739, 1739 | 1743, 1739 | 1735, 1731 | C=O stretching of lipid esters |
| 1396, 1396 | 1392, 1392 | 1388, 1407 | 1407, 1396 | COO⁻ stretching vibrations of amino acid side chains |
| 1238, 1238 | 1238, 1234 | 1234, 1234 | 1238, 1238 | P=O asymmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1080, 1080 | 1080, 1080 | 1080, 1091 | 1087, 1072 | P=O symmetrical stretching of $PO_2$ phosphodiesters from phosphorylated molecules |
| 1153, 1153 | 1153, 1157 | 1153, 1149 | 1157, 1157 | C–O vibrations from glycogen and other carbohydrates |
| 1022, 1018 | 1022, 1018 | 1022, 1018 | 1018, 1014 | C–O vibrations from glycogen and other carbohydrates |
| 995, 995 | 995, 983 | 991, 991 | 987, 991 | C–O stretch from RNA ribose chain |

spectra of the day 3 and day 4 progeny (Fig. 7G and H) corroborated the spectroscopic differences indicated by the PLS-DA.

Interestingly, differences between the progeny from both culture conditions were different from those that had initially allowed their discrimination from the undifferentiated hESCs from which they were derived. Prior to differentiation, spectra from MEF-derived hESCs had higher lipid absorbances compared to spectra from Matrigel-derived hESCs. However the converse was true following differentiation. The finding that lipid stores diminished during the loss of pluripotency corroborates data previously published by our laboratory and other groups.[13,16] A starting hypothesis may be that the

eicosanoid signaling pathway,[28] thought to be involved in lipid depletion during differentiation, may be impaired in the Matrigel cultured cells, causing less efficient lipid down-regulation than the MEF-derived hESCs.

A possible explanation for the hESCs' "memory" of their original conditions may be due to different culturing environments affecting the endogenous production of growth factors which impacted differentiation. FGF and nodal,[29,30] activators of the SMAD1/5/8 signaling pathway such as BMPs,[31] as well as nodal activin signaling antagonists[32] are all possible candidates. An alternative explanation may involve culture-induced epigenetic mechanisms, which have been well documented in the

Analyst

Paper

**Table 5**  Band maxima of the human embryonic stem cell line *MIXL1*<sup>GFP/w</sup> co-cultured with feeders and maintained for 5 passages in mTESR medium *versus* the same line, after 5 passages on Matrigel, in mTESR medium

Band maxima 2nd derivative spectra (cm$^{-1}$)

| Day 5 on Matrigel in mTESR medium | Day 5 on feeders in mTESR medium | Factor 1 loadings (cm$^{-1}$) | Factor 3 loadings (cm$^{-1}$) | Band assignments |
|---|---|---|---|---|
| 2923 | 2923 | 2923 | 2923 | C–H asymmetrical stretching of lipid groups |
| 2850 | 2850 | 2850 | 2854 | C–H symmetrical stretching of lipid groups |
| 1739 | 1739 | 1728 | 1739 | C=O stretching of lipid esters |
| 1396 | 1396 | 1400 | 1404 | COO$^-$ stretching vibrations of amino acid side chains |
| 1242 | 1238 | 1245 | 1245 | P=O asymmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1083 | 1080 | 1091 | 1083 | P=O symmetrical stretching of PO$_2$ phosphodiesters from phosphorylated molecules |
| 1153 | 1153 | 1149 | 1157 | C–O vibrations from glycogen and other carbohydrates |
| 1022 | 1018 | 1018 | 1022 | C–O vibrations from glycogen and other carbohydrates |
| 995 | 991 | 991 | 995 | C–O stretch from RNA ribose chain |

literature.[33,34] For example, one group described various culture-dependent methylation changes that were heritable even after differentiation of their hESC lines.[33] Another study reported transcriptional and epigenetic modifications, thought to reduce transcriptional heterogeneity, in normoxic cultured (2% O$_2$) hESCs compared to 21% O$_2$ cultured hESCs.[34] Transcriptional heterogeneity within hESCs has been known to give rise to suboptimal differentiation efficiencies although no evidence of this was found in our study.

## Conclusion

This study emphasises the view that stem cell biologists need to be aware that hESCs and their differentiated progeny, maintained by different means, are not spectroscopically equivalent and this knowledge underscores the need to develop standardised protocols for the production of cells destined for use in the clinic.

## Acknowledgements

## References

1 J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall and J. M. Jones, Embryonic Stem Cell Lines Derived from Human Blastocysts, *Science*, 1998, **282**(5391), 1145–1147.

2 M. Richards, C.-Y. Fong, W.-K. Chan, P.-C. Wong and A. Bongso, Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells, *Nat. Biotechnol.*, 2002, **20**(9), 933–936.

3 M. Amit, V. Margulets, H. Segev, K. Shariki, I. Laevsky, R. Coleman and J. Itskovitz-Eldor, Human Feeder Layers for Human Embryonic Stem Cells, *Biol. Reprod.*, 2003, **68**(6), 2150–2156.

4 O. Hovatta, M. Mikkola, K. Gertow, A.-M. Strömberg, J. Inzunza, J. Hreinsson, B. Rozell, E. Blennow, M. Andäng and L. Ahrlund-Richter, A culture system using human foreskin fibroblasts as feeder cells allows production of human embryonic stem cells, *Hum. Reprod.*, 2003, **18**(7), 1404–1409.

5 E. S. Rosler, G. J. Fisk, X. Ares, J. Irving, T. Miura, M. S. Rao and M. K. Carpenter, Long-term culture of human embryonic stem cells in feeder-free conditions, *Dev. Dyn.*, 2004, **229**(2), 259–274.

6 E. Sjögren-Jansson, M. Zetterström, K. Moya, J. Lindqvist, R. Strehl and P. S. Eriksson, Large-scale propagation of four undifferentiated human embryonic stem cell lines in a feeder-free culture system, *Dev. Dyn.*, 2005, **233**(4), 1304–1314.

7 S. R. Braam, L. Zeinstra, S. Litjens, D. Ward-van Oostwaard, S. van den Brink, L. van Laake, F. Lebrin, P. Kats, R. Hochstenbach, R. Passier, A. Sonnenberg and C. L. Mummery, Recombinant Vitronectin Is a Functionally Defined Substrate That Supports Human Embryonic Stem Cell Self-Renewal *via* αVβ5 Integrin, *Stem Cells*, 2008, **26**(9), 2257–2265.

8 M. M. Mitalipova, R. R. Rao, D. M. Hoyer, J. A. Johnson, L. F. Meisner, K. L. Jones, S. Dalton and S. L. Stice, Preserving the genetic integrity of human embryonic stem cells, *Nat. Biotechnol.*, 2005, **23**(1), 19–20.

9 M. K. Carpenter, E. Rosler and M. S. Rao, Characterization and Differentiation of Human Embryonic Stem Cells, *Cloning Stem Cells*, 2003, **5**(1), 79–88.

10 A. Maitra, D. E. Arking, N. Shivapurkar, M. Ikeda, V. Stastny, K. Kassauei, G. Sui, D. J. Cutler, Y. Liu, S. N. Brimble, K. Noaksson, J. Hyllner, T. C. Schulz, X. Zeng, W. J. Freed, J. Crook, S. Abraham, A. Colman, P. Sartipy, S.-I. Matsui,

Paper

Analyst

M. Carpenter, A. F. Gazdar, M. Rao and A. Chakravarti, Genomic alterations in cultured human embryonic stem cells, *Nat. Genet.*, 2005, **37**(10), 1099–1103.

11 V. Akopian, P. Andrews, S. Beil, N. Benvenisty, J. Brehm, M. Christie, A. Ford, V. Fox, P. Gokhale, L. Healy, F. Holm, O. Hovatta, B. Knowles, T. Ludwig, R. McKay, T. Miyazaki, N. Nakatsuji, S. Oh, M. Pera, J. Rossant, G. Stacey and H. Suemori, Comparison of defined culture systems for feeder cell free propagation of human embryonic stem cells, *In Vitro Cell. Dev. Biol.*, 2010, **46**(3/4), 247.

12 P. Heraud and M. J. Tobin, The emergence of biospectroscopy in stem cell research, *Stem Cell Res.*, 2009, **3**(1), 12–14.

13 P. Heraud, E. S. Ng, S. Caine, Q. C. Yu, C. Hirst, R. Mayberry, A. Bruce, B. R. Wood, D. McNaughton, E. G. Stanley and A. G. Elefanty, Fourier transform infrared microspectroscopy identifies early lineage commitment in differentiating human embryonic stem cells, *Stem Cell Res.*, 2010, **4**(2), 140–147.

14 D. Ami, T. Neri, A. Natalello, P. Mereghetti, S. M. Doglia, M. Zanoni, M. Zuccotti, S. Garagna and C. A. Redi, Embryonic stem cell differentiation studied by FT-IR spectroscopy, *Biochim. Biophys. Acta, Mol. Cell Res.*, 2008, **1783**(1), 98–106.

15 M. J. Walsh, T. G. Fellous, A. Hammiche, W.-R. Lin, N. J. Fullwood, O. Grude, F. Bahrami, J. M. Nicholson, M. Cotte, J. Susini, H. M. Pollock, M. Brittan, P. L. Martin-Hirsch, M. R. Alison and F. L. Martin, Fourier Transform Infrared Microspectroscopy Identifies Symmetric $PO_2^-$ Modifications as a Marker of the Putative Stem Cell Region of Human Intestinal Crypts, *Stem Cells*, 2008, **26**(1), 108–118.

16 J. K. Pijanka, D. Kumar, T. Dale, I. Yousef, G. Parkes, V. Untereiner, Y. Yang, P. Dumas, D. Collins, M. Manfait, G. D. Sockalingum, N. R. Forsyth and J. Sule-Suso, Vibrational spectroscopy differentiates between multipotent and pluripotent stem cells, *Analyst*, 2010, **135**(12), 3126–3132.

17 M. J. German, H. M. Pollock, B. Zhao, M. J. Tobin, A. Hammiche, A. Bentley, L. J. Cooper, F. L. Martin and N. J. Fullwood, Characterization of Putative Stem Cell Populations in the Cornea Using Synchrotron Infrared Microspectroscopy, *Invest. Ophthalmol. Visual Sci.*, 2006, **47**(6), 2417–2421.

18 J. Filik, M. D. Frogley, J. K. Pijanka, K. Wehbe and G. Cinque, Electric field standing wave artefacts in FTIR microspectroscopy of biological materials, *Analyst*, 2012, **137**(4), 853–861.

19 P. Bassan, J. Lee, A. Sachdeva, J. Pissardini, K. M. Dorling, J. S. Fletcher, A. Henderson and P. Gardner, The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications, *Analyst*, 2013, **138**(1), 144–157.

20 R. T. Tecirlioglu, L. Nguyen, K. Koh, A. O. Trounson and A. E. Michalska, Derivation and maintenance of human embryonic stem cell line on human adult skin fibroblast feeder cells in serum replacement medium, *In Vitro Cell. Dev. Biol.*, 2010, **46**(3/4), 231–235.

21 R. P. Davis, E. S. Ng, M. Costa, A. K. Mossman, K. Sourris, A. G. Elefanty and E. G. Stanley, Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors, *Blood*, 2008, **111**(4), 1876–1884.

22 M. Costa, K. Sourris, T. Hatzistavrou, A. G. Elefanty and E. G. Stanley, Expansion of Human Embryonic Stem Cells In Vitro, in *Current Protocols in Stem Cell Biology*, John Wiley & Sons, Inc., 2007.

23 M. Costa, K. Sourris, T. Hatzistavrou, A. G. Elefanty and E. G. Stanley, Expansion of human embryonic stem cells *in vitro*, in *Current Protocols in Stem Cell Biology*, 2008, pp. 1C.1.1–1C.1.7.

24 E. S. Ng, R. Davis, E. G. Stanley and A. G. Elefanty, A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies, *Nat. Protoc.*, 2008, **3**(5), 768–776.

25 E. S. Ng, R. P. Davis, T. Hatzistavrou, E. G. Stanley and A. G. Elefanty, Directed Differentiation of Human Embryonic Stem Cells as Spin Embryoid Bodies and a Description of the Hematopoietic Blast Colony Forming Assay, in *Current Protocols in Stem Cell Biology*, John Wiley & Sons, Inc., 2007.

26 H. Martens, J. P. Nielsen and S. B. Engelsen, Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures, *Anal. Chem.*, 2003, **75**(3), 394–404.

27 P. Geladi, Notes on the history and nature of partial least squares (PLS) modelling, *J. Chemom.*, 1988, **2**(4), 231–246.

28 O. Yanes, J. Clark, D. Wong, G. Patti, A. Sánchez-ruiz, H. Benton, S. Trauger, C. Desponts, S. Ding and G. Siuzdak, Metabolic oxidation regulates embryonic stem cell differentiation, *Nat. Chem. Biol.*, 2010, **6**(6), 411.

29 I. Ginis, Y. Luo, T. Miura, S. Thies, R. Brandenberger, S. Gerecht-Nir, M. Amit, A. Hoke, M. K. Carpenter, J. Itskovitz-Eldor and M. S. Rao, Differences between human and mouse embryonic stem cells, *Dev. Biol.*, 2004, **269**(2), 360–380.

30 J. M. Sperger, X. Chen, J. S. Draper, J. E. Antosiewicz, C. H. Chon, S. B. Jones, J. D. Brooks, P. W. Andrews, P. O. Brown and J. A. Thomson, Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**(23), 13350–13355.

31 N. Sato, I. M. Sanjuan, M. Heke, M. Uchida, F. Naef and A. H. Brivanlou, Molecular signature of human embryonic stem cells and its comparison with the mouse, *Dev. Biol.*, 2003, **260**(2), 404–413.

32 R. Brandenberger, H. Wei, S. Zhang, S. Lei, J. Murage, G. J. Fisk, Y. Li, C. Xu, R. Fang, K. Guegler, M. S. Rao, R. Mandalam, J. Lebkowski and L. W. Stanton, Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation, *Nat. Biotechnol.*, 2004, **22**(6), 707–716.

Analyst

Paper

33 C. Allegrucci, Y.-Z. Wu, A. Thurston, C. N. Denning, H. Priddle, C. L. Mummery, D. Ward-van Oostwaard, P. W. Andrews, M. Stojkovic, N. Smith, T. Parkin, M. E. Jones, G. Warren, L. Yu, R. M. Brena, C. Plass and L. E. Young, Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome, *Hum. Mol. Genet.*, 2007, **16**(10), 1253–1268.

34 A. Downing, N. R. Forsyth, K. Hampson, A. Kay, J. McWhir and R. Talbot, Transcriptome alterations due to physiological normoxic (2% $O_2$) culture of human embryonic stem cells, *Regener. Med.*, 2008, **3**(6), 817–833.