BIOINFORMATICS APPROACHES TO STUDYING MESENCHYMAL STEM CELL BEHAVIOUR ON ARTIFICIAL EXTRACELLULAR MATRICES

Ross Brian Gurden

(BMedSc, MRes)

Submitted in fulfilment of requirements for the degree of Doctor of Philosophy (PhD)

Centre for Cell Engineering
Institute of Molecular, Cell and Systems Biology
School of Medical, Veterinary and Life Sciences
University of Glasgow

December 2017

Abstract

Stem cells have potential use in tissue engineering and regenerative medicine, and as they underlie the development and maintenance of tissues throughout life, how they function is also of interest. The extracellular matrix presents a variety of physical and chemical signals to stem cells to regulate their behaviour *in vivo*. Recapitulation of these signals *in vitro* could enable the control of explanted stem cells to facilitate their study. Biomaterials that display extracellular-matrix inspired cues are one way to do this. By combining surface chemistry and fibronectin, an extracellular matrix protein with cell binding and growth factor binding domains, the conformation of fibronectin was controlled to create artificial extracellular matrices. Adsorbed on a film of poly(ethyl acrylate), fibronectin adopted a network-like conformation which ostensibly increased the exposure of its functional domains, whereas on poly(methyl acrylate) it had an unconnected organisation with more concealed domains. The growth factors bone morphogenetic protein 2 and vascular endothelial growth factor, known to bind to fibronectin, were adsorbed to the network conformation. Prior studies have reported that these artificial extracellular matrices differentially affected cell behaviour. In this work, the growth and differentiation of human bone marrow stromal cell surface marker-1 positive mesenchymal stem cells was characterised on these substrates. It was shown that all combinations of fibronectin conformation and growth factors supported cell adhesion and growth. A high-content image processing and analysis pipeline was developed to take advantage of automated fluorescence microscopy to show that cytoskeletal, nuclei, and differentiation-associated protein features distinguished cells cultured on the artificial extracellular matrices. Those on the isolated conformation and the network conformation with vascular endothelial growth factor were particularly distinct. Further, metabolomics revealed several metabolic pathways that differed in activity between the fibronectin conformations. To analyse the metabolomics data a Quick Results web application was built, which extended the existing Polyomics integrated Metabolomics Pipeline. The application improves the visualisation and interpretation of untargeted liquid chromatography—mass spectrometry metabolomics data. This work gives insights into how these artificial extracellular matrices can control stem cell behaviour, and developed and

demonstrated several tools to improve the understanding of these biomaterials and the use of metabolomics data.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank the BBSRC for funding and providing resources for my PhD (BBSRC WestBio DTP - Grant Number: BB/J013854/1).

I am extremely grateful for my supervisors Matthew Dalby, Manuel Salmerón-Sánchez, Rónán Daly, and Karl Burgess. Matthew provided so much support for me during the project, and I will always appreciate his effort in creating opportunities for me. I'm aware Manuel worked tirelessly to ensure I had all the resources and foundations for the project, and he has always supported new things I wanted to try. From Rónán and Karl I learnt so much in a short space of time and had a lot of fun doing so. Without this exact team, I probably would not have had such an enjoyable time during the PhD.

The groups working at the CCE, Polyomics, and MiMe are made of brilliant people. I don't think I could have found better shoulders to work alongside. They are both great company and hard-working, and this enthusiasm helped me along at several points. I am also grateful to the IR team at the Uni of Strathclyde for our work at Diamond, and Diamond Light Source and the team at B22 for our beamtime.

To my friends, I can't thank you enough.

My family have provided limitless support, happiness, encouragement, and been eternally patient. For this, I will forever be indebted.

Author's Declaration

I hereby declare the research conducted in this thesis is my own work, unless otherwise stated, and at the time of submission, is not being considered elsewhere for any other academic qualification.

Ross Brian Gurden

20th December 2017

Abbreviations/Definitions

HSC	Haematopoietic stem cell
MSC	Mesenchymal stem cell
TERM	Tissue engineering and regenerative medicine
STRO-1	Stromal cell surface marker 1
VCAM1	Vascular cell adhesion molecule 1
(p)Runx2	(phosphorylated) Runt-related transcription factor 2
CD146	Cluster of differentiation 146
CD271	Cluster of differentiation 271
NESTIN	Neuroectodermal stem cell marker
ALP	Alkaline phosphatase
OCN	Osteocalcin
OPN	Osteopontin
ECM	Extracellular matrix
FN	Fibronectin
BMP2	Bone morphogenetic protein 2
VEGF	Vascular endothelial growth factor
RGD	arginine-glycine-aspartic acid
PHSRN	proline-histidine-serine-arginine-asparagine
(k)Pa	(kilo) Pascal
PMA	Poly(methyl acrylate)
PEA	Poly(ethyl acrylate)
HUVEC	Human vascular endothelial cell
PBS	Phosphate buffered saline
FBS	Foetal bovine serum
DMEM	Dulbecco's Modified Eagle's medium
EDTA	Ethylenediaminetetraacetic acid
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid
BM	Basal medium/media
OM	Osteogenic medium/media
AM	Adipogenic medium/media
FTIR	Fourier-transform infrared (spectroscopy)
AFM	Atomic force microscopy

IgG	Immunoglobulin G
DAPI	4',6-diamidino-2-phenylindole (blue fluorescence)
ICW	In-Cell Western
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
SR-FTIR	Synchrotron radiation-FTIR
ANOVA	Analysis of variance
PMA + FN	PMA with FN
PEA + FN	PEA with FN
PEA + FN + BMP2	PEA with FN and BMP2
PEA + FN + VEGF	PEA with FN and VEGF
ICA	Independent components analysis
TRITC	Tetramethylrhodamine (red fluorescence)
FITC	Fluorescein isothiocyanate (green fluorescence)
QR	Quick Results (application for visualising metabolomics data)
PiMP	Polyomics integrated Metabolomics Pipeline
m/z	mass-to-charge ratio
TIC	Total ion current (chromatogram)
LC	Liquid chromatography
MS	Mass spectrometry/spectrometer
RT	Retention time
LC-MS	Liquid chromatography—mass spectrometry
FiRDI	Filter related data interactively (application for displaying related data)
RPM	Rotations per minute
RSD	Relative standard deviation
ppm	Parts per million
PLAGE	Pathway level analysis of gene expression
JSON	JavaScript object notation
SQL	Structured query language
ID	Identifier
InChiKey	International chemical identifier key
KEGG	Kyoto Encyclopaedia of Genes and Genomes
PCA	Principal components analysis
GLCM	Grey-level co-localisation matrix

1 General Introduction

1.1 Stem Cells, Tissue Engineering, and Regenerative Medicine

Stem cells are self-renewing cells that can differentiate to produce specialised cells throughout life as part of tissue development and maintenance (Odorico et al. 2001). Types of stem cell are classified by the stage of development in which they exist and the cells they produce. For example, during embryogenesis there are totipotent stem cells, which can produce any cell type, and pluripotent stem cells, which can produce any cell type of the developing embryo but not extra-embryonic tissues (Bissels et al. 2013). Gradual reductions in lineage choice define postnatal or adult tissue-specific multipotent stem cells, such as haematopoietic stem cells (HSCs), which produce blood cells (Orkin 2000), and mesenchymal stem cells (MSCs), which produce skeletal tissues (Pittenger et al. 1999). Further restrictions lead to unipotent stem cells that produce only one cell type, such as spermatogonial stem cells (Bissels et al. 2013). Fundamental stem cell biology is therefore a research focus. Further, because of these properties, stem cells are studied for their potential to create substitutes for or to repair tissues that have been lost or impaired by illness, ageing, or injury, known as the field of tissue engineering and regenerative medicine (TERM) (Bianco & Robey 2001; Fisher & Mauck 2013).

1.2 Mesenchymal Stem Cells

MSCs are of particular interest for TERM and basic biology because they form bone, fat, cartilage, and reticular tissues (Pittenger et al. 1999) (Figure 1.1), and are part of the haematopoiesis-supporting stroma of bone marrow (Sacchetti et al. 2007; Pinho et al. 2013). They are believed to predominantly reside at perivascular sites on bone marrow sinusoids, but have reportedly been observed at other sites around the body (Bianco 2014). Their isolation for culture is prospective and relies on selecting bone marrow cells that express certain markers. Several markers with overlapping expression profiles have been proposed as truly enriching for MSCs, such as cluster of differentiation 146 (CD146) (Sacchetti et al. 2007), neuroectodermal stem cell marker (NESTIN) (Méndez-Ferrer et al. 2010; Isern et al. 2013), cluster of differentiation CD271

(Tormin et al. 2011), vascular cell adhesion molecule 1 (VCAM1) (Gronthos et al. 2003), and stromal cell surface marker 1 (STRO-1) (Gronthos et al. 1999; Gronthos et al. 1994; Simmons & Torok-Storb 1991; Kulterer et al. 2007). The relative merits of each marker are not clear but they can be used to enrich for a cell population displaying stem cell phenotypes (Lv et al. 2014). After isolation from bone marrow they readily adhere to tissue culture plastic and proliferate in standard culture conditions (serum-containing media) (Williams et al. 2013). There is as yet no consensus on whether the proliferative and differentiation capacity of MSCs vary with age or donor (Zhou et al. 2008; Stenderup et al. 2001; Surdo & Bauer 2012; Siddappa et al. 2007), but it is generally believed they cannot be cultured indefinitely in standard conditions and still remain multipotent (Javazon et al. 2004).

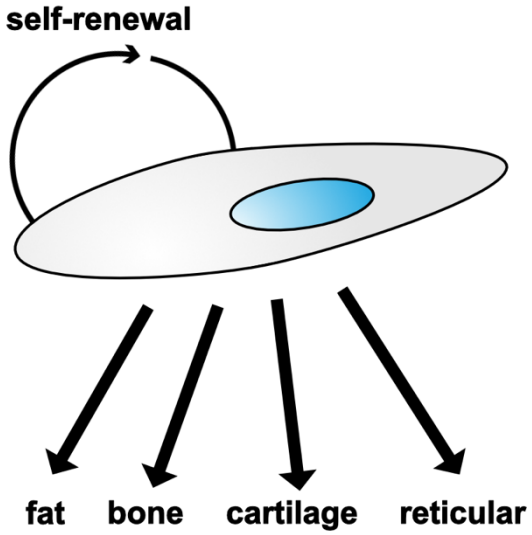

Figure 1.1 MSCs can self-renew and differentiate into bone, fat, cartilage and other reticular cell types.

1.2.1 MSC Osteogenesis

Bone marrow MSCs have been used to produce bone *in vitro* and *in vivo* (Sacchetti et al. 2007; Lian & Stein 1992). The differentiation of MSCs into terminal osteoblasts that produce mineralized bone matrix can be followed by monitoring the ordered expression of various osteogenesis-related markers (Stein & Lian 1993). Runt-related transcription factor 2 (Runx2), also commonly known as core-binding factor alpha 1, is a transcriptional activator of osteogenic differentiation (Ducy 2000; Stein et al. 2004). The presence of the phosphorylated form of Runx2 in the nucleus is a hallmark of early osteogenesis

(Stein et al. 2004; Vimalraj et al. 2015; Komori 2010). During the initial stages of differentiation MSCs proliferate and form nodule-like features, at which point proliferation reduces (Stein & Lian 1993). Alkaline phosphatase (ALP) is an enzyme expressed at the cell surface and is used as a mid-to-late stage osteogenesis marker (Sharma et al. 2014; Pittenger et al. 1999; Štefková et al. 2015; Stein & Lian 1993). It is required to initiate bone matrix mineralisation at these nodule-like features. (Wennberg et al. 2000; Bellows et al. 1991). The expression and secretion of osteopontin (OPN) and osteocalcin (OCN) into the extracellular environment is used as an indicator of terminal differentiation (Pittenger et al. 1999; Kulterer et al. 2007; Lian & Stein 1992; Stein & Lian 1993). The whole process is believed to occur over 3-4 weeks (Stein & Lian 1993).

1.3 Control of Stem Cell Behaviour

Stem cell behaviour can be guided by signals from the extracellular environment — the surrounding chemical and physical cues — that influence internal cellular state (Scadden 2006; Huang 2010) leading to changes in, for example, shape, proliferation, and differentiation (Burridge & Chrzanowska-Wodnicka 1996) (Figure 1.2). Different spatiotemporal combinations of cell state and extracellular signals lead to the variety of behaviours necessary to develop, sustain, and repair the body. This, together with the means to generate variations of a cell's culture environment (provided by advances in engineering and molecular biology (Murphy et al. 2014)), suggests that for every *in vivo* stem cell function there is a corresponding combination of biomaterials, biomolecules, and cells for achieving that same behaviour *in vitro*. Indeed, the use of biomaterials to control stem cell fate is a focus of TERM and stem cell biology research (Discher et al. 2009; Lutolf et al. 2009).

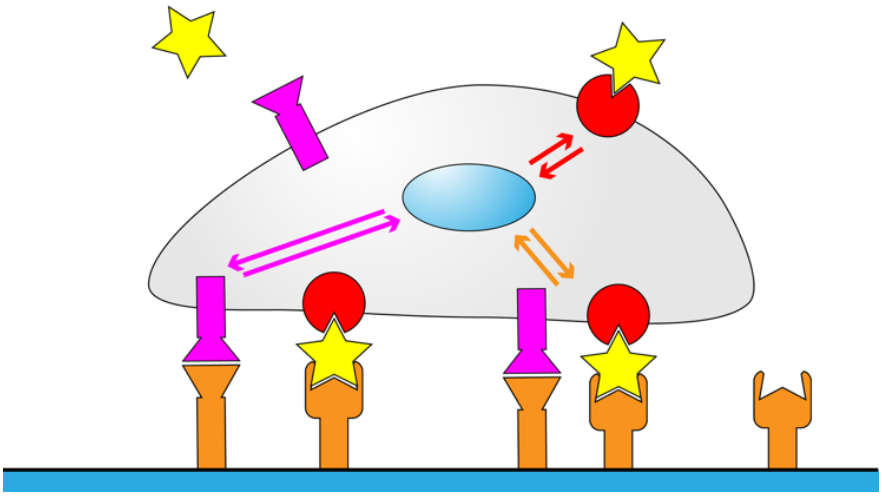

Figure 1.2 Extracellular signals can regulate cell behaviour. Soluble signals (free star) or their bound counterparts (star bound to orange star receptor that is substrate bound) and innate extracellular ligands (orange flat ligand) can be sensed by cells. Adhesion or binding to these ligands (red circular receptor, purple flat receptor) modify downstream signalling cascades (arrows) to affect cell behaviour. Specific combinations of ligand-receptor pairings have different effects on cell behaviours. Substrate, blue and black.

1.3.1 The Extracellular Matrix, Cell Adhesion, and Bidirectional Signalling

The extracellular matrix (ECM) provides anchorage and signals to cells, which, in turn, can rearrange, breakdown, or synthesise new ECM (Humphrey et al. 2014; Geiger et al. 2009). It is mainly composed of water, proteins — predominantly fibrillar proteins like collagen and proteoglycans like fibronectin (FN) — and polysaccharides (Mouw et al. 2014). Cells sense, respond to, and manipulate their local ECM *via* adhesion molecules on their periphery (Geiger et al. 2009). Foremost of these are integrins, which are transmembrane proteins that allow bidirectional signalling between cells and the ECM. They bind extracellular ligands, and their intracellular tails are nucleation points for cytoskeletal and signalling protein attachment and organisation (Parsons et al. 2010; Hynes 2002). The best understood integrin-based signal transduction hubs are called focal adhesions, which form when cytoskeletal tension pulls the ECM-connected integrins into a cluster (Burridge et al. 1988; Geiger et al. 2009; Geiger et al. 2001; Burridge & Chrzanowska-Wodnicka 1996). They are micron-sized, elongated, flat complexes found at the cell periphery (Geiger et al. 2001). The degree of focal adhesion formation and signal protein recruitment can be affected by the distribution of ligands in the ECM, namely the density and

average distance between them (Coyer et al. 2012; Cavalcanti-Adam et al. 2007; Cavalcanti-Adam et al. 2008). Cells can also control the activation of integrins (Hynes 2002).

Integrins recognise specific ECM domains as one of 24 known heterodimers composed of an alpha (α) and beta (ß) subunit (Barczyk et al. 2010). Ligands include the arginine-glycine-aspartic acid (RGD) tripeptide found on ECM proteoglycans like laminin and FN, and the leucine-aspartic acid-valine tripeptide also found in ECM proteoglycans and growth factor receptors like VCAM1 (Humphries et al. 2006). Ligand binding induces conformational changes in integrins, initiating the formation of adhesion complexes at the intracellular integrin domains (Hynes 2002). Many proteins are purported to be involved in these intracellular signalling complexes (Geiger et al. 2001). Some of these interact directly with the intracellular domains of integrins; others are platforms for the recruitment of signalling molecules; enzymes are also present, such as phosphatases and kinases; and yet others form the linkage between the complex and cytoskeleton (Wozniak et al. 2004). They can be roughly grouped into functional layers: from outside the cell to within are the extracellular and intracellular integrin domains, an integrin-associated signalling complex layer, a force transduction layer, an actin regulatory layer, and then the cytoskeleton itself (Kanchanawong et al. 2010). The complexity of focal adhesions was embodied in an analysis that assembled a consensus integrin adhesome complex from several focal adhesion proteomics datasets of mammalian cells cultured on FN, comprised of 41 proteins involved in 92 interactions (Horton et al. 2015).

Vinculin is a central protein in focal adhesion formation and signalling, recruited to focal adhesion complexes after initial integrin binding events (Geiger et al. 2001). It is a prominent adaptor protein, binding to both the actin cytoskeleton and signalling enzymes (Carisey & Ballestrem 2011). Its roles in mechanotransduction can be divided into the transmission of intracellular actomyosin-generated tension to the linked ECM (Parsons et al. 2010), and mechanically responding to ECM resistance to this tension to mediate intracellular signalling pathways (Atherton et al. 2016).

Actin is the main structural cytoskeletal protein with roles in cell motility, shape, and signal transduction. It polymerises from a globular to fibrous

structure (F-actin) (Dominguez & Holmes 2011). Actin fibres transmit tension from and to the ECM via focal adhesions to sustain isometric tension between cell and ECM (Geiger et al. 2001). As well as for the assembly of focal adhesions, which regulates intracellular signalling cascades (Burridge & Chrzanowska-Wodnicka 1996), tension is part of several cellular functions such as traction-based motility using myosin-generated forces (Li et al. 2005), assembling ECM (Singh et al. 2010; Mouw et al. 2014), and regulating gene expression (Wang et al. 2009; Swift et al. 2013). Microtubules, for example ß-tubulin, are another cytoskeletal protein associated with focal adhesions, contributing to their turnover and associated ECM remodelling (Stehbens & Wittmann 2012; Ng et al. 2014). They also have a role in general cell metabolism, for example in controlling glucose transport and enzymes (Cassimeris et al. 2012). The integrity of microtubules and the actin cytoskeleton have been reported as necessary for MSC differentiation (Rodríguez et al. 2004; Yourek et al. 2007; Sonowal et al. 2013)

1.3.1.1 Fibronectin

FN is a major ECM protein that is involved in ECM assembly, cell binding, and soluble signal sequestration, which is in part due to a variety of binding domains (Zollinger & M. L. Smith 2017). It is also found in a soluble form in plasma (To & Midwood 2011). FN has three types of repeating domains, named I-III, of which the 9th and 10th type III repeats contain the proline-histidine-serine-arginine-asparagine (PHSRN) synergy and RGD peptides critical for FN matrix assembly. They are also cell-binding domains, so are referred to as the main integrin binding site (Singh et al. 2010). The 12-14th type III repeats bind several growth factors including bone morphogenetic protein 2 (BMP2) and vascular endothelial growth factor (VEGF) (Martino & Hubbell 2010). FN matrix deposition, known as FN fibrillogenesis, is the process whereby FN is secreted, dimerised, and assembled into insoluble fibres by cells (Singh et al. 2010). Cell-mediated tension of FN is believed to unfold FN domains, particularly the type III repeats, to expose cryptic binding sites necessary for matrix formation and binding activity (Smith et al. 2007; Baneyx et al. 2002).

1.3.2 Biomaterial Control of Stem Cell Fate

Various substrate properties can be changed with the aim of controlling cultured cell behaviour. Adjusting the cell-substrate interface is a common approach to controlling stem cell fate (Lutolf et al. 2009; Murphy et al. 2014).

Changing the elasticity of a substrate by varying the crosslinking in polyacrylamide gels was shown to influence MSC differentiation: a stiffer substrate (34 kPa) increased the expression of osteogenic markers compared to softer substrates (Engler et al. 2006). In addition, when the elasticity of subcellular-sized regions was varied, instead of the entire substrate, MSC osteogenesis was also greater on stiffer patterns and there was a higher expression of multipotency markers on the softer counterparts (Yang et al. 2016). Altering substrate topography by printing different symmetries of nanoscale pit arrays into poly(methyl methacrylate) also influenced MSC fate; on a disordered topography MSCs differentiated to the osteoblastic lineage (Dalby et al. 2007), whereas a symmetric topography maintained MSCs (McMurray et al. 2011). Cell-adhesive islands surrounded by a non-adhesive substrate can be used to control cell shape and spreading. Using this approach, it was shown that MSCs cultured on large or star-shaped islands underwent osteogenesis whereas those on small or flower-shaped islands underwent adipogenesis (McBeath et al. 2004; Kilian et al. 2010).

The modification of substrates to affect MSC differentiation is not limited to physical changes; the display of different chemical groups can affect MSC behaviour as well. Substrates comprised of methyl-terminated chemicals were reported to maintain stem cell phenotype, whereas those made of amino- or silane-terminated molecules induced osteogenesis, and hydroxyl- and carboxyl-groups promoted chondrogenesis (Curran et al. 2005; Curran et al. 2006). Further, creating an array of submicron dots of the methyl-terminated chemicals were reported to maintain MSC phenotype in comparison to those made from amino-terminated molecules (Curran et al. 2010). The effects of these tail groups may not result from direct interaction with MSCs but instead through a layer of adsorbed protein. Proteins from the culture medium can adsorb rapidly to a substrate before cell attachment, and substrate properties can affect

protein adhesion and conformation (Steiner et al. 2007; Baujard-Lamotte, et al. 2008).

Another approach to controlling cell behaviour is by presenting ECM-inspired ligands, such as protein domains or growth factors, in a cell culture system (Lutolf & Hubbell 2005; Hettiaratchi et al. 2016). These systems can use chemical and physical approaches to control ligand tethering and function, or allow passive adsorption to a culture substrate (Kinney & McDevitt 2013). Using proteins inspired by the bone marrow ECM, MSC osteogenic differentiation was reported to be greater on passively adsorbed vitronectin and collagen compared to FN or control plain tissue culture plastic (Salasznyk et al. 2004). Instead of whole proteins, only the domains of interest can be presented. MSCs cultured on substrates displaying the cell-binding RGD and PHSRN peptides of FN at a fixed distance were shown to have undergone an equal or greater degree of osteogenesis compared to a random arrangement of these domains or passively adsorbed FN itself (Fraioli et al. 2016).

As well as using fibrillar or proteoglycan ECM proteins, growth factors that are known to influence MSC phenotype can be used; BMP2, believed to induce osteogenesis in MSCs (Carreira et al. 2014), directly coupled to silk fibroin polymers attached to a solid substrate induced greater osteogenesis in MSCs in comparison to that passively adsorbed onto the silk substrate (Karageorgiou et al. 2004). The use of ligands has also been coupled more directly with feature geometry. Compared to displaying BMP2-mimetic and RGD peptides randomly or as ordered alternating micron-sized rectangles, those patterned into squares or triangles appeared to induce cultured MSCs to undergo osteogenesis to a greater extent (Bilem et al. 2017a). Displaying the same BMP2-mimetic peptides alone in the square or triangle shapes was also reported to increase MSC osteogenesis over the rectangle or random presentations, and this phenomenon was not observed with shapes displaying solely RGD peptides (Bilem et al. 2017b).

1.3.2.1 Fibronectin Based Biomaterials

The work in this thesis focuses on the control of FN and growth factor presentation to control MSC behaviour. Various approaches have been used to alter the conformation of adsorbed FN to explore changes in domain

arrangement and availability and their influence on cell behaviour (Zollinger & Smith 2017). For example, FN appeared to have a more extended conformation on hydrophobic surfaces in comparison to hydrophilic surfaces (Klotzsch et al. 2014). Further, by adsorbing FN on substrates displaying different chemical functional groups (hydroxyl, carboxyl, methyl/alkyl, amino), it was suggested that there was differential availability of cell-binding domains as determined by the degree of antibody binding to FN domains (Keselowsky et al. 2003), and that MSCs underwent osteogenesis on FN-coated amino-terminated substrates but not on others (Phillips et al. 2010).

The presence of certain ECM molecules, such as heparin or its functional groups, which interact with FN *in vivo* (Zollinger & Smith 2017), were shown to change FN conformation *in vitro* to alter its growth factor binding activity (Mitsi et al. 2006) and subsequent focal adhesion formation in adipose tissue-derived mesenchymal stem cells (Araújo et al. 2016). The spreading and migration of endothelial cells on FN fibres was altered by changing the strain applied to the fibres, presumably because the strain affected the unfolding of cryptic FN functional domains (Hubbard et al. 2016). FN conformational changes induced by adsorption to bacterial, tissue, or collagen-coated culture plastics, as indicated by exposure of cell-binding domains using antibodies that targeted specific FN domains, altered C2C12 muscle myoblast proliferation and differentiation (García et al. 1999).

Several studies have suggested that there is cooperation between integrins and growth factor receptors in response to bound growth factors and protein ligands in the ECM — particularly to FN — referred to as synergistic signalling (Salmerón-Sánchez & Dalby 2016; Kim et al. 2011; Dingal & Discher 2014). Studies of the presentation of growth factors on ECM proteins indicate that this collaboration enhances growth factor signalling beyond its effect as a soluble molecule. The presence of both the FN RGD cell binding domain and growth factor binding domains was reported to be necessary for mouse FN-null fibroblasts adhesion and survival on tissue culture plastic (Lin et al. 2011). By exposing cultured human bone marrow MSCs to a recombinant protein containing both the FN integrin binding site and growth factor binding region together with BMP2, MSC proliferation, migration, and expression of osteogenic markers *in vitro* were

enhanced in comparison to the recombinant protein alone; the same model was shown to improve repair of an *in vivo* bone defect (Martino et al. 2011). It was reported that VEGF bound to FN caused VEGF-receptor and $\alpha_5\beta_1$ integrin heterodimer — which recognises the RGD peptide in the FN cell binding domain (Hynes 2002) — clustering and potentiated endothelial cell migration (Wijelath et al. 2002). A recombinant protein comprised of FN integrin and growth factor binding sites in conjunction with VEGF was shown to increase endothelial cell migration and proliferation relative to VEGF or the recombinant protein alone (Wijelath et al. 2006). Improved healing of *in vivo* bone and skin wound models was shown using bound growth factors compared to that free in solution (Martino et al. 2014; Martino et al. 2015).

1.3.2.2 Poly(methyl acrylate) and poly(ethyl acrylate)-driven FN and Growth Factor-Based Biomaterials

The systems used in this thesis to control FN and growth factor presentation are based on the use of two polymers, poly(methyl acrylate) and poly(ethyl acrylate) (PMA and PEA, respectively). These polymers differ in their constituents by the length of the alkyl chain that makes up the acrylate monomer: where methyl acrylate contains only a single carbon (a methyl group $-CH_3$), ethyl acrylate has two (an ethyl group $-CH_2-CH_3$). When prepared as an approximately 1 μm thick film on a glass substructure (Vanterpool et al. 2014), they reportedly have similar wettability and surface roughness (a broad measure of topography or smoothness). The density of adsorbed FN from a solution of fixed concentration is also comparable (Salmerón-Sánchez et al. 2011; Guerra et al. 2010; Vanterpool et al. 2014; Mnatsakanyan et al. 2015). The elasticities of the films are considered similar (~1000-2000 kPa) with respect to a cell's ability to distinguish between values greater than 40 kPa (Llopis-Hernández et al 2016; Rico et al. 2016; Engler et al. 2006; González-García et al. 2012; Guerra et al. 2010).

It has been reported that FN adopts an interconnected network-like conformation with nano- to micron-sized fibres when adsorbed onto PEA, whereas on PMA FN has an isolated conformation on a similar scale (Guerra et al. 2010; Salmerón-Sánchez et al. 2011; González-García et al. 2012; Llopis-Hernández et al. 2013; Vanterpool et al. 2014; Rico et al. 2016; Moulisová et al.

2017) (Figure 1.3). It is reported that the availability of the FN RGD binding site is similar between PEA and PMA, whereas the PHSRN peptide is more available on PEA (Vanterpool et al. 2014). Further, the growth factor binding site was suggested to be more available on PEA (Moulisová et al. 2017; Llopis-Hernández et al. 2016). These results are believed to indicate that FN has an extended conformation with FN-FN interactions on PEA, similar to the result of cell-mediated FN fibrillogenesis, but not on PMA (Salmerón-Sánchez & Dalby 2016). BMP2 and VEGF bind to FN on PMA and PEA and their blank counterparts (Llopis-Hernández et al. 2016; Moulisová et al. 2017).

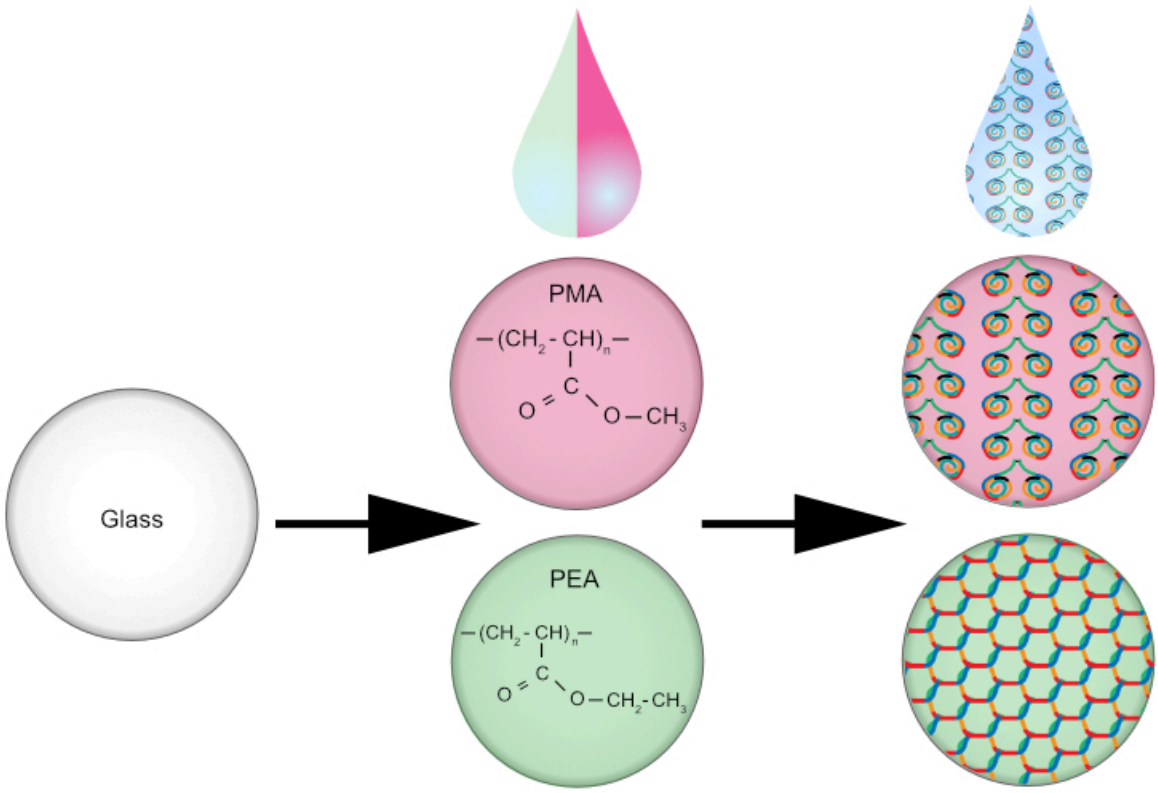

Figure 1.3 Assembling the artificial FN matrices. A solution of PEA (green) or PMA (pink) is used to produce a film of the polymer on a glass (white) substructure. The surfaces of the coated samples are then immersed in a solution of FN, which organises into either an interconnected network-like conformation (PEA) or a relatively isolated non-connected conformation (PMA).

The behaviour of several cell types on these artificial ECMs have been studied, and some are relevant to TERM having focused on differentiation. Myogenic differentiation of C2C12 cells into myoblasts was reportedly greater on PEA with FN substrates compared to those cultured on PMA with FN (Salmerón-Sánchez et al. 2011; Mnatsakanyan et al. 2015). It was communicated that human bone marrow STRO-1^{+} MSCs underwent greater osteogenic differentiation by day 21

(21 d) of culture on PMA with FN than on PEA with FN, but at day 1 there was no difference in pRunx2 expression (González-García et al. 2012). Further, it was reported that osteogenic and adipogenic differentiation of mouse C3H10T1/2 cells, an embryonic MSC cell line, was not different between the FN conformations, and that instead, PEA with FN maintains C3H10T1/2 multipotency (Rico et al. 2016). Some experiments have also used growth factors in combination with the different FN conformations due to the reported differential availability of the growth factor binding domains. *In vitro* human MSC osteogenesis and *in vivo* repair *via* an implant in a mouse bone defect model were enhanced when using the PEA-based FN network and adsorbed BMP2 compared to that without BMP2 (Llopis-Hernández et al. 2016). Using VEGF instead of BMP2, it was shown that the vasculogenic response of human umbilical vascular endothelial cells (HUVEC) was improved when presented using the FN network compared to the isolated conformation (Moulisová et al. 2017).

Some focal adhesion-related proteins have been investigated as part of mechanistic studies on how these artificial ECMs influence cell behaviour. Using vinculin as a proxy for total focal adhesions, the size of focal adhesions formed after 3 h on the artificial ECM was reportedly different between FN conformations, with those on the isolated FN conformation (PMA) smaller than those formed on the FN network (PEA) in mouse MC3T3 and C3H10T1/2 cells, and human L929 fibroblasts, suggesting a difference in downstream signalling (Guerra et al. 2010; Rico et al. 2016; Vanterpool et al. 2014). However, focal adhesions after 3 d culture in human bone marrow STRO-1$^+$ MSCs were reportedly not different in size (González-García et al. 2012). Further, the degree of focal adhesion kinase activation, an integrin-associated enzyme involved in regulating focal adhesion formation (Geiger et al. 2009), after 3 h appeared to be greater in cells cultured on PEA with FN than PMA with FN in C2C12 myoblasts and C3H10T1/2 cells (Rico et al. 2016; Mnatsakanyan et al. 2015). Signalling pathways downstream of VEGF signalling are reportedly more active when HUVECs are presented with VEGF bound to the FN network conformation compared to soluble VEGF or VEGF adsorbed onto the isolated FN conformation (Moulisová et al. 2017).

Figure 1.4 is an illustration of the model of how the artificial ECMs create a synergistic effect between growth factor receptors and integrins with substrate-bound FN and growth factors. In summary, the network conformation of FN on PEA is believed to result in greater exposure of the growth factor binding and cell binding domains of FN, in comparison to the isolated conformation on PMA. The increased growth factor binding domain exposure means that growth factors bind to the network conformation more so than to the isolated counterpart. Together, this means cellular growth factor receptors and integrins are engaged together on PEA to a larger extent than on PMA, resulting in differential effects on cultured cell phenotype.

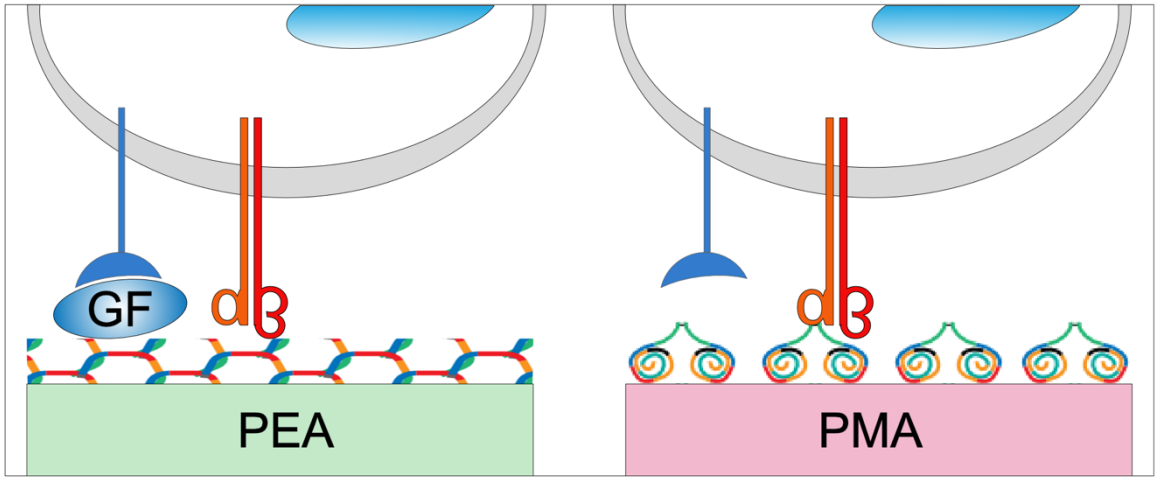

Figure 1.4 Model of synergistic binding on material-driven FN matrices. FN is believed to adopt an extended network-like conformation when adsorbed onto a film of PEA, which exposes integrin and growth factor binding sites otherwise hidden when in an isolated, globular conformation, as reported when adsorbed onto PMA. The exposure of the growth factor binding domains on PEA means growth factors (GF) can bind, and together with the also exposed cell binding sites, means both cellular integrins (orange and red alpha and beta sticks) and growth factor receptors (blue cup) adhere. Synergistic binding is lower on PMA because of the less expanded FN conformation.

1.4 Bioinformatics Data and Software

The field of bioinformatics can be defined as the amalgamation of biology, computer science, and mathematics (Bayat 2002; Hagen 2000). Technological advancements and their application to biological research have led to the production of large, complex datasets from biological experiments, the analyses of which require this combination of fields. This is perhaps most notably embodied by "omics" experiments, namely genomics, transcriptomics, proteomics, and metabolomics, that aim to analyse the whole complement of

genes, transcripts, proteins, and metabolites in a biological sample, respectively (Horgan & Kenny 2011). By studying these molecular classes at such scale, it is anticipated that a more holistic understanding of biology will be gained, in contrast to small-scale experiments that perhaps measure only a handful of molecules at a time. This way of measurement also means that targets do not need to be picked for study in advance.

These approaches are not without limitations, however. Raw data direct from the recording instruments are usually in an abstract form from which the biological context is unclear and which therefore must be pre-processed before useful information can be extracted. Selecting and using appropriate pre-processing steps usually requires the assistance of a bioinformatician or statistician (Horgan & Kenny 2011). These large-scale datasets are normally too large and the pre-processing too complex to be analysed using conventional graphical applications (for example, Microsoft Excel). The use of software programming languages is much more suitable as they are flexible, expressive, permit explicitly recorded analyses for reproducibility, and scale with dataset size. There are domain-specific programming languages like R or MATLAB that are built specifically for numeric data handling tasks, or general-purpose languages like Python that have specialist libraries for the same purpose. Indeed, toolkits and workflows that cover the range of pre-processing and analysis steps exist for each omics type, such as BioPython for handling sequencing data using Python (Cock et al. 2009) or XCMS for metabolomics data (Smith et al. 2006).

However, the ability to process a bioinformatics dataset from raw data to biological inference using appropriate tools does not necessarily offset its complexity. Ideally, an end user would be provided with a list of interesting molecules summarised from a bioinformatics experiment without needing to know any further details, but ultimately certain ones do need to be preserved. These are related to what the instrument actually records and how it is related back to molecular biology. Genomics and transcriptomics are based around the measurement of fragments of genes or transcripts, but the biological inference from the experiment is in the form of changes in whole genes or transcripts (Goodwin et al. 2016). Metabolomics and proteomics is the measurement of

fragments or derivatives of molecules with specific properties, but the biological inference is of changes in whole metabolites or proteins (Patti et al. 2012; Mallick & Kuster 2010). It is important to acknowledge the underlying structure of the data and pre-processing steps when using it to make biological inferences (Figure 1.5). This argument can be generalised to any data source, but it is particularly important for omics technologies because of the current state of bioinformatics software and the division between bioinformatics and end users of biological data.

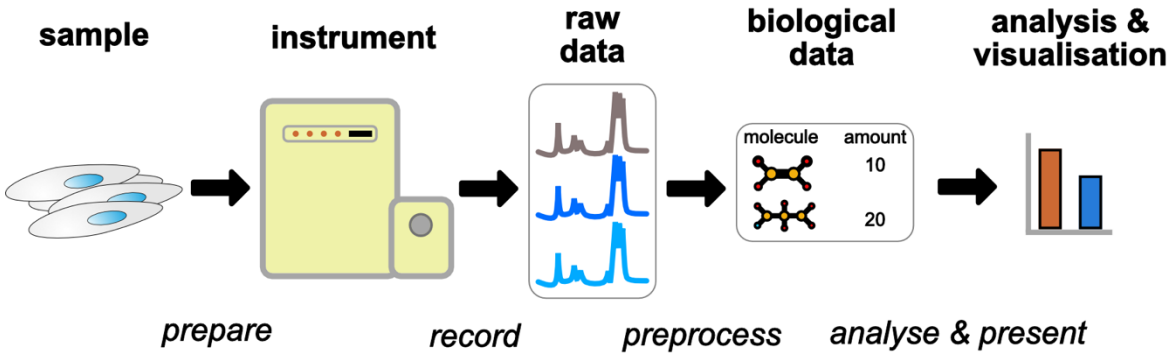

Figure 1.5 General pipeline for an omics experiment. A sample is prepared or extracted in a form suitable for measurement, which is then done using an instrument. The subsequent raw data is pre-processed to transform it into data that represents the biology of the sample, which then needs to be analysed and visualised for inferences to be made. When a list of biological data and analyses are interpreted it should be in the context of these previous stages.

There is no shortage of established and new bioinformatics software aimed at users without practical knowledge of software development, programming, or statistics to enable biological researchers to benefit from bioinformatics without needing significant training (Spicer et al. 2017; Goecks et al. 2010). Such software needs to balance the desired endpoint with respect to the researcher (a list of changing molecules, for example) and providing important other details about how the data was produced. This is a trade-off between complexity, because bioinformatics data is inherently complicated and some aspects of this complexity needs to be preserved while remaining clear and intuitive; and usability, because hard-to-use software is unlikely to be helpful or used in practice. Getting this balance wrong may mean the software does not support the user in performing bioinformatics analyses. Further, the software needs to be easy to maintain and expand by the developers (Carpenter et al. 2012).

1.4.1 Improving the Usability and Interpretation of Metabolomics Data with the Polyomics integrated Metabolomics Pipeline

1.4.1.1 Metabolomics

Metabolites are important in determining cell phenotype because they form a cell's biochemistry (Patti et al. 2012), the connection between genes and phenotype (Fiehn 2002). By measuring the metabolome, the complement of small molecules transformed during metabolism, there is potential for direct insight into cell status. The comprehensive measurement of the metabolome, metabolomics, has been used in the study of stem cells including MSCs (McNamara et al. 2012; McNamara et al. 2011; Surrati et al. 2016; Tsimbouri et al. 2014; Alakpa et al. 2016), neural stem and progenitor cells (Knobloch et al. 2013), HSCs (Oburoglu et al. 2014), and induced pluripotent and embryonic stem cells (Panopoulos et al. 2012). Untargeted metabolomics to measure as many metabolite classes as possible, in contrast to targeted metabolomics that focuses on specific metabolites (Zamboni et al. 2015), was used to compare the metabolomes of MSCs cultured on the artificial ECMs. In particular, liquid chromatography—mass spectrometry (LC-MS) was used to acquire the metabolomics data.

Figure 1.6 (a) outlines the acquisition of LC-MS metabolomics data. A sample (Figure 1.6 (a) sample stage), which is a solvent-based extract of metabolites, is passed through a chromatography column that separates metabolites based on their interaction with its solid phase. The time taken for a given metabolite to elute from the column is called retention time (RT) (Figure 1.6 (a) LC stage). The separated sample then gradually elutes into the mass spectrometer, which ionises and measures the relative abundance (or intensity) of the metabolites at each RT over a range of masses (Figure 1.6 (a) MS stage). The total ion current (TIC) chromatogram shows the sum of measured intensities over time (Figure 1.6 (a) TIC stage). At each RT, a range of masses is detected. These are processed into peaks — a mass (m/z) and RT pair — which represent metabolites (Smith et al. 2006; Scheltema et al. 2011). The peaks at a specific RT for two hypothetical samples are shown in Figure 1.6 (b). The abundance of metabolites between experimental conditions forms the data of interest (Figure 1.6 (b)). Annotating a peak with its cognate metabolite then allows biological inference in the context

of known biochemicals and their pathways. Some peaks can be confidently annotated by matching to authentic chemical standards measured alongside experimental samples — known as "identified" peaks — but for peaks that do not have corresponding standards, tentative metabolite identification is performed by matching against a pre-existing database (Sumner et al. 2007; Dunn et al. 2013; Aretz & Meierhofer 2016).

Database searching can produce ambiguous results: a metabolite can annotate multiple peaks at the same mass, which can be especially challenging if they have conflicting abundances (Figure 1.6 (c)). Further, a single peak can resolve to multiple metabolites that have the same mass (Figure 1.6 (d)). Using a list of metabolites without considering this so-called "peak misannotation" is a pitfall of untargeted LC-MS metabolomics (Lu et al. 2017). Adding to the complexity is that, with respect to metabolic pathway analysis, a metabolite can belong to multiple pathways. It is crucial for LC-MS metabolomics data interpretation that a user understands the potential complexity that underlies the detected pathways and metabolites.

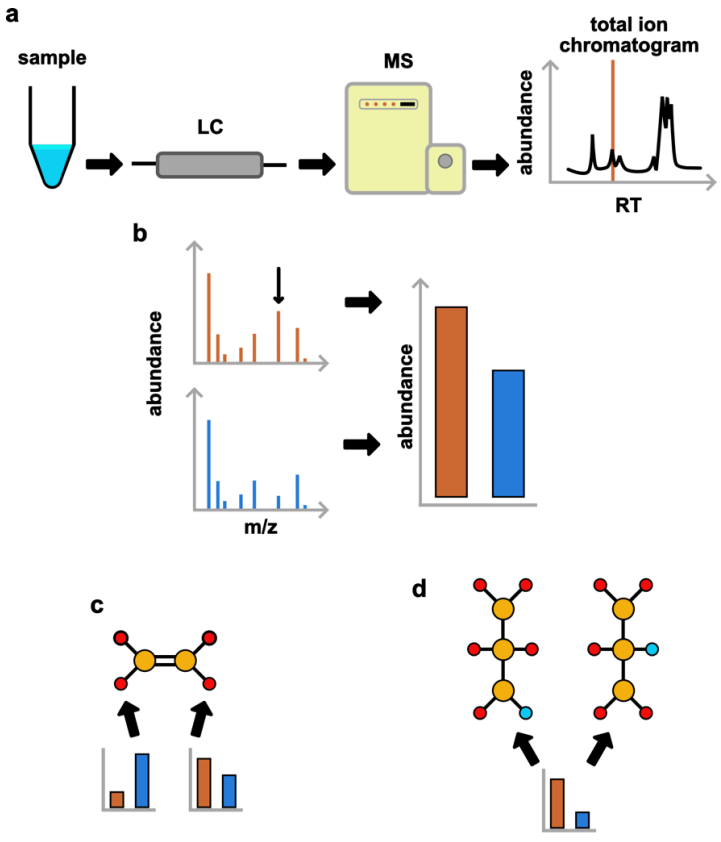

Figure 1.6 The acquisition and pre-processing of LC-MS metabolomics data. (a) A solvent extract (sample) is loaded into a liquid chromatography (LC) column and elutes into a mass

spectrometer (MS), which measures the mass (mass to charge ratio, m/z) of metabolites in the sample. Overall the result can be visualised as a TIC, the sum of intensities of detected metabolites at each RT. (b) Peaks, which are mass and RT pairs that represent a metabolite, are detected by pre-processing this raw signal. For example, peaks are detected at a specific RT (orange line on the TIC in (a)), shown in two example samples as orange and blue spectra, and the abundances of one of these in each spectrum (black vertical arrow) is compared in the bar chart. Peaks are assigned to metabolites by searching a database, leading to some ambiguity in matches such as multiple peaks annotated by one metabolite (c) or multiple metabolites matching to one peak (d).

1.4.1.2 Polyomics integrated Metabolomics Pipeline

The Polyomics integrated Metabolomics Pipeline (PiMP) is a freely available, open-source metabolomics pipeline which aims to be an easy-to-use metabolomics data analysis application (Gloaguen et al. 2017). It is developed by Glasgow Polyomics (University of Glasgow, UK), who also provide services for metabolomics data acquisition (among other omics technologies). PiMP fulfils an unmet need in metabolomics data processing as an integrated pipeline to take metabolomics data from its raw form to biological context, and it does this in such a way that enables biological researchers to do their own analysis. To use PiMP, a user enters details about the experimental design, provides the raw metabolomics data, optionally sets some pre-processing parameters, and PiMP returns the peaks, metabolites, and pathways found in the dataset and statistics on their comparisons between experimental groups.

PiMP is built as a web application so that it may be used by anyone with access to a modern web browser, and so works using a client-server architecture. This means no installation is required. The server code is built using the Django web framework, a Python library, which uses the common model-view-template (related to model-view-controller) design pattern for structuring and accessing (model), manipulating (view), and presenting (template) data. An R pipeline is controlled by Django to pre-process and analyse the data. The web page itself was built using modern web standards: HTML, CSS, JavaScript.

Each of the sets of peaks, metabolites, and pathways are shown in separate tabs in PiMP (Figure 1.7). There is some informational overlap between these tabs so that the relationships of the metabolomics data can be followed. For example, the metabolites are shown in a sidebar when a peak is selected in the peak comparison tab (Figure 1.7 (a)), and the pathways associated with a metabolite are shown in the metabolites tab (Figure 1.7 (b)). However, following the

relationships can be cumbersome because of the need to navigate these tabs. In the instance displayed in Figure 1.7 (b), after discovering 1,3-diaminopropane, it is not clear which, if any, other metabolites are associated with the same peaks that represent 1,3-diaminopropane. Further, a sidebar is not an efficient way of displaying the associated peaks, or for that matter, any of the relationships underlying the data. Selecting a pathway in the pathways tab does not provide a list of associated metabolites found in the dataset, and for large pathways like the ABC transporters map, an alternative to using the sidebar to present this information is needed anyway. Nevertheless, because of the way PiMP was architected it can be readily modified with new functionality to improve these issues.

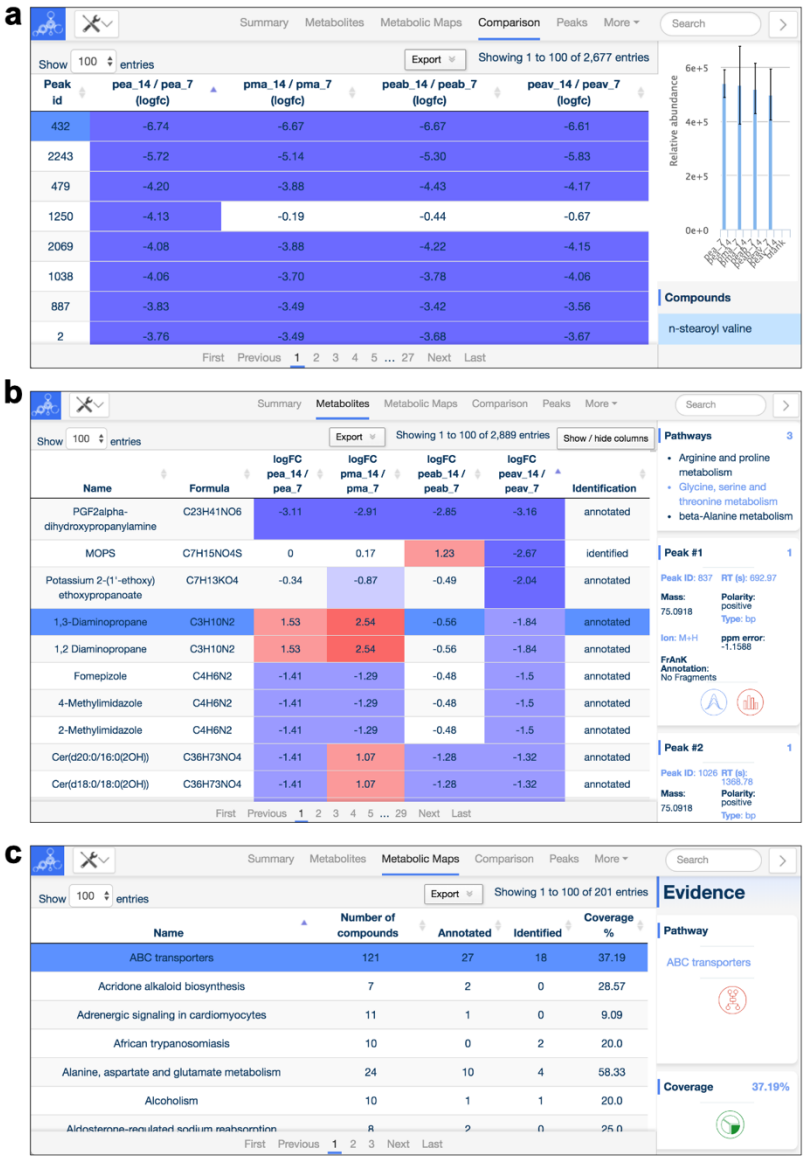

Figure 1.7 PiMP uses a tabbed-interface to display the peaks, metabolites, and pathways in a metabolomics dataset. (a) The comparisons tab shows a table of relative changes of

peaks by experimental group. Selecting one (in the example, peak ID 432) shows the related metabolites in a sidebar (for peak 432, there is one compound, n-stearoyl valine). (b) In the metabolites tab, a table of metabolites and their relative changes is shown. Selecting a metabolite (here, 1,3-Diaminopropane) shows its related pathways and peaks in a sidebar. (c) The pathway tab displays a list of pathways that contained metabolites found in the dataset. The ABC transporters pathway is selected in this example to show its coverage and a link to a pathway database. However, it does not show which metabolites are in each pathway.

1.4.2 Fluorescence Microscopy and High-Content Automated Image Analysis

1.4.2.1 Fluorescence Microscopy

Biological fluorescence microscopy refers to the imaging of organic samples stained with antibodies that are conjugated to fluorescent compounds, known as fluorophores, or dyes with inherent fluorescence activity, that recognise and bind to a molecule of interest. Fluorescence is the emission of light of a specific range of wavelengths (emitted light) from a fluorescent compound after it has absorbed a particular colour of light (excitation light), and a fluorescence microscope serves as a way to target the excitation light at a stained sample, and where to capture emitted light from, at high magnification (Figure 1.8). The wavelength regions corresponding to a particular colour of fluorescence are referred to as channels, and fluorescence microscopes are equipped with a mechanism for selecting which channel to image. Multiple fluorophores that target different molecules and have distinct emission and excitation profiles can be used to image multiple components at once.

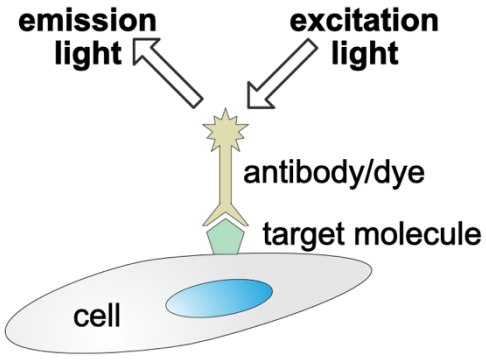

Figure 1.8 Fluorescent staining to visualise molecules of interest. A cell expressing a molecule of interest, either intracellularly or extracellularly, is stained with an antibody conjugated to a fluorophore or fluorescent dye that binds to that molecule. A microscope is used to focus light of a specific excitation wavelength onto the stained target molecule, which is absorbed by the fluorescent compound, and which then emits light at a specific wavelength that is captured by the microscope to form an image. A sample can be stained with multiple fluorophores against different target molecules.

Several of the aforementioned studies used fluorescence image processing to examine the influence of the FN conformations on cell morphometry and intracellular protein expression. L929 fibroblasts and C2C12 myoblasts cultured for 3 h on PEA with FN had a larger area than those on PMA with FN (Vanterpool et al. 2014; Mnatsakanyan et al. 2015). Further, C2C12 cells were reported to be more circular on PMA with FN than on PEA with FN (Mnatsakanyan et al. 2015). However, while the area of C3HT101/2 cells was not reportedly different across the FN conformations, they were more circular on PMA with FN (Rico et al. 2016). HUVECs were shown to have a larger area on PEA with FN with adsorbed VEGF in comparison with PEA with FN alone, but the presence of VEGF on PMA with FN appeared to make no difference (Moulisová et al. 2017). Image processing was also used to draw the aforementioned conclusions about focal adhesion size.

1.4.2.2 Automated Microscopy

These image analyses tended to be of small numbers ($n < 200$) of cells from manually recorded fluorescence images, which is typical for fluorescence microscopy analyses. Fluorescence microscopy can also be performed in an automated manner with additional hardware, known as automated fluorescence microscopy. The bulk of fluorescence imaging in this thesis used an automated microscope system comprised of a motorised sample platform and channel selector (a light filter) with a fluorescence microscope, which was controlled using a computer and software package. The software controlling the set-up can be instructed to take many images (tiles) over a defined area in a number of fluorescence channels, that can then be stitched together to form a large multichannel mosaic image (Figure 1.9). This was used to capture images of multiple entire samples at the same time, meaning that every single cell stained by the fluorescent markers could be analysed.

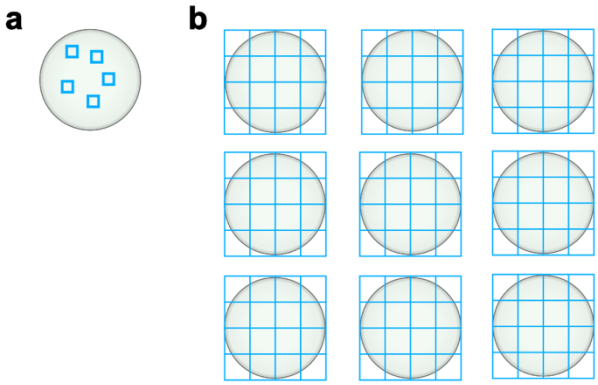

Figure 1.9 Unautomated and automated microscopy. Using unautomated microscopy (a) requires manual selection of regions (blue squares) to image, and is done for each sample (green circles) in turn, whereas automated microscopy (b) means whole samples and defined regions (blue grids) can be imaged together automatically.

1.4.2.3 High-Content Image Analysis

If the target molecule in a fluorescence image can be used to identify individual cells then they can be computationally segmented from the image into distinct regions. This allows each cell on a substrate to be measured to understand the variation of such a molecule on a single cell level. Further, while other molecules of interest might not entirely delimit a cell, they can be imaged in a channel separate to the one used for segmentation to allow for its measurement as well (Figure 1.10).

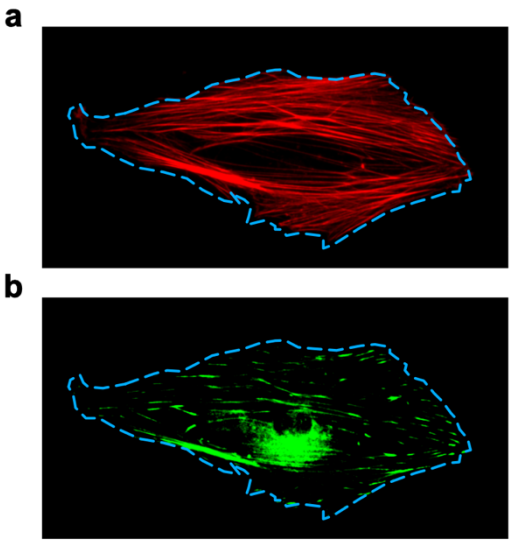

Figure 1.10 Cell segmentation. If an imaged fluorescently-labelled molecule delimits a cell body it could be used to computationally segment the cell from the rest of the image for analysis. For example, the actin cytoskeleton can be used (a). Another molecule labelled with a different fluorophore that may not demarcate a cell, but lies within its bounds, can be measured within the segmented region, for example vinculin (b). Red, actin; green, vinculin; blue, example segmentation of cell.

Various measurements about the molecules of interest can be determined by quantifying a fluorescence image. The segmented raw pixels, represented as intensity values, can be summed to produce an estimate of the total amount of marker. The intensity values can also be converted into various forms for further processing to calculate features related to their boundary and internal qualities (Gonzalez et al. 2009; Gonzalez & Woods 2010a). High-content image analysis is the name given to this process, the extraction of information from images using computational approaches to reveal patterns that may have remained otherwise unnoticed (Shamir et al. 2010; Sommer & Gerlich 2013). Internal qualities of a segmented region are properties like texture, which is the relationship between intensity and position. Textural features of biological molecules as calculated from fluorescence molecules are a common way of comparing them between experimental conditions (Caicedo et al. 2017). Using a toy greyscale image as an example (Figure 1.11 (a)), different textural statistics can be calculated from the histogram of its pixel intensities (Figure 1.11 (b)) or grey-level co-localisation matrix (GLCM) (Figure 1.11 (b)). Whereas a pixel intensity histogram ignores the spatial relationship between pixels, a GLCM represents a combination of which pixel intensities are adjacent to others (Haralick et al. 1973). In addition to these pixel-based features, geometric measurements of the segmented region, known as morphometrics, can be calculated, such as area or perimeter length.

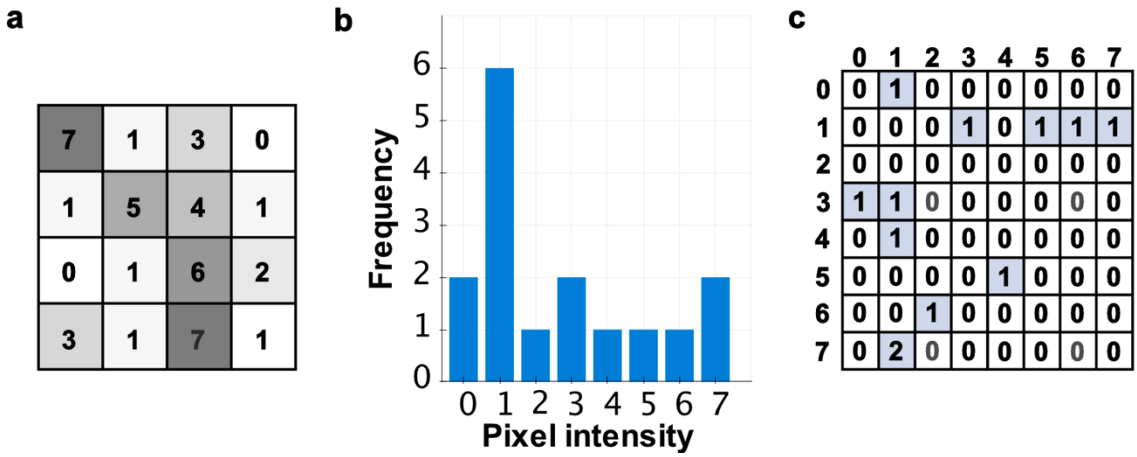

Figure 1.11 High-content image analysis. (a) Example greyscale image. Each pixel in an image is represented as a number that is proportional to the intensity of fluorescence recorded at that location. The calculation of various features relating to the intensity and relative locations of the pixels using computational tools can be referred to as high-content image analysis. For example, statistics about the histogram of pixel intensities (b) or statistics about the GLCM, which summarises the mapping between pixel intensities and locations in an image (c). The example here shows counts of pixel intensities directly to the

right of a given pixel for image (a). For example, there is one pixel with intensity 1 to the right of a pixel with intensity 0 (1st row, 2nd column), and there are no other pixels adjacent to a pixel with intensity 0 (1st row); there are two pixels with intensity 1 that are to the right of a pixel with intensity 7 (8th row, 2nd column).

Because of the large number of measurements that can be made about each cell in an image, and the large number of cells that can be imaged, the process becomes data-intensive and relies on the same fields that underlie the processing of omics datasets. For this reason, this domain is sometimes referred to as bioimage informatics (Peng 2008).

Since target molecules for fluorescent staining are not limited to those that demarcate cell-surface boundaries, intracellular organelles like nuclei can also be segmented with a sufficient marker. Together with a multichannel image of different markers that delineate the borders between various organelles, this means that multiple features of several cellular and subcellular biomolecules can be determined from a single image. The combination of automated fluorescence microscopy and high-content image analysis therefore enables the broad measurement and comparison of larger numbers of subcellular and cellular aspects and samples for less effort than manual fluorescence microscopy (Ljosa & Carpenter 2009). High-content analysis can be applied to small numbers of manually selected images as well, but conclusions are more robust when drawn from the large numbers of cells gained from automated microscopy.

1.4.2.4 High-Content Automated Fluorescence Microscopy Analysis of Biomaterials Control of MSCs

Several studies have used high-content automated fluorescence image analysis to profile stem cell behaviour on various biomaterials. The ALP expression of MSCs cultured on a variety of topographical features made of different geometric shapes was imaged to determine the most osteogenic combination, and various cell morphometric features were measured to determine which were associated with the highest ALP expression (Unadkat et al. 2011). Further, descriptors based on the actin cytoskeleton were used to establish what morphometric features were associated with topographically enhanced osteogenesis (Hulshof et al. 2017). Using cell shape and nuclei or actin cytoskeleton texture features, surfaces with different roughness or topography were compared for their capacity to induce osteogenesis in MSCs (Unadkat et al.

2013; Vega et al. 2015). The effect of a substrate with a hydrophobicity gradient on MSC differentiation in adipogenic media was measured using high-content image analysis, which revealed that MSC position on the gradient could be determined from nuclear protein descriptors that were correlated with degree of differentiation (Vega et al. 2012). High-content imaging of multipotency-associated proteins has also been used to select combinations of polymers that can maintain MSC multipotency from a number of candidates (Duffy et al. 2014). More generally, cytoskeletal and nuclear protein features derived from fluorescence images can be used to predict differentiated phenotypes at an early time point (Liu et al. 2010; Treiser et al. 2010; Marklein et al. 2016; Vega et al. 2012; Vega et al. 2015).

1.5 Project Aims

The aim of this project was to further the aforementioned works on using the artificial ECMs to control MSC behaviour. In particular, the study of human bone marrow STRO-1$^+$ MSCs on the FN conformations (González-García et al. 2012) will be expanded to include those with additional growth factors BMP2 (Llopis-Hernández et al. 2016) and VEGF (Moulisová et al. 2017). This was to take advantage of the potential synergistic effect of co-presenting growth factors and biomaterial-induced FN domain exposure to regulate cell behaviour (Salmerón-Sánchez & Dalby 2016). In doing so, several techniques (high-content image processing and metabolomics), which have not been used before to study these substrates and cells, produced sizeable and complex datasets. To analyse these, bespoke and general software for visualisation and analysis was therefore developed. In all, several pieces of software were built in order to extend the analysis of STRO-1$^+$ MSC behaviour on the artificial ECMs, or to generally improve the state of analysis of the associated data.

The objectives of this project were:

1. Characterise the growth and osteogenic differentiation of human bone marrow STRO-1$^+$ MSCs on the artificial ECMs with BMP2 or VEGF.

2. Develop and use a high-content image processing pipeline to analyse automated microscopy images of MSCs cultured on the artificial ECMs in

order to study substrate-induced differences in cell morphology, and cytoskeleton and nucleus features.

3. Make improvements to PiMP by developing a metabolomics data visualisation and analysis tool and compare the metabolomes of MSCs cultured on the artificial ECMs.

2 Surface and Cell Culture Characterisation

2.1 Introduction

The introduction described artificial ECMs made of FN and growth factors, and studies of MSC osteogenesis cultured with them, including commercially-sourced human bone marrow MSCs (Llopis-Hernández et al. 2016), primary human bone marrow STRO-1$^+$ MSCs (González-García et al. 2012), and mouse C3H10T1 MSCs (Rico et al. 2016). Primary human bone marrow STRO-1$^+$ MSCs (referred to as MSCs hereon) have not been characterised on the artificial ECMs with growth factors. This chapter extends the previous work by presenting experiments of MSCs cultured on the full range of artificial ECMs: PMA or PEA with FN (PMA + FN, PEA + FN), and PEA with FN and BMP2 or VEGF (PEA + FN + BMP2, PEA + FN + VEGF). Using STRO-1$^+$ MSCs instead of commercially-sourced cells is preferable as they are provided at passage zero, thus having undergone minimal and transparent preparation (Williams et al. 2013). Further, the STRO-1$^+$ fraction of human bone marrow is believed to be an enriched source of MSCs (Gronthos et al. 2003; Gronthos et al. 1994).

In this chapter, the binding and organisation of FN and growth factors to the artificial ECMs were explored using customary and new approaches, for example infrared (IR) spectroscopy. Based on these data, the substrates were reliably reproduced. MSCs were cultured on the artificial ECMs with the aim of characterising their survival and differentiation, tested by measuring the expression of relevant proteins and their global biochemical profile. MSCs adhered and spread quickly, forming actin and tubulin cytoskeletons and focal adhesions. OPN expression was greater on the network conformation of FN in comparison to its isolated counterpart, and IR analysis indicated changes in lipid, protein, and nucleic acids between the two conformations. No discernible differences were observed when adding BMP2 or VEGF. Finally, different culture conditions were tested to improve differentiation. Serum-free adhesion conditions used in prior studies (Rico et al. 2016; Llopis-Hernández et al. 2016; González-García et al. 2012) caused notable cell death and therefore were unsuitable for use with these MSCs.

The aims of this chapter were to: (1) compare the production of the artificial ECMs with previous data to ensure reproducibility and extend their characterisation using IR spectroscopy; (2) characterise MSC behaviour across the different artificial ECMs; (3) optimise culture conditions to affect MSC differentiation.

2.2 Materials and Methods

2.2.1 Tables of Reagents

Table 2.1 Fabrication Materials

Item	Source
Toluene	Alfa Aesar
PMA	UPV, Spain
PEA	UPV, Spain
Ethanol	VWR Chemicals
12 mm diameter glass coverslips	VWR
13 mm diameter Raman-grade calcium fluoride CaF_2 windows	Crystran
Human plasma fibronectin	Sigma-Aldrich or R&D Systems
Recombinant human BMP2	R&D Systems
Recombinant human VEGF-165 (VEGF)	R&D Systems
Phosphate buffered saline (PBS)	Sigma-Aldrich

Table 2.2 Cell Culture Reagents

Item	Source
PBS	Sigma-Aldrich
Foetal bovine serum (FBS)	Sigma-Aldrich
4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid (HEPES)	Fisher Scientific
Dulbecco's Modified Eagle's medium (DMEM)	Sigma-Aldrich
Trypsin	Sigma-Aldrich
Versene	In-house
Phenol red 0.5 %	Sigma-Aldrich
Penicillin-streptomycin	Sigma-Aldrich
MEM non-essential amino acids	Sigma-Aldrich
Sodium pyruvate	Life Technologies
L-glutamine 200 mM	Invitrogen
Glucose	Fisher Scientific
Medium 199	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Ascorbate-2-phosphate	Sigma-Aldrich
Dexamethasone	Sigma-Aldrich
Insulin	Sigma-Aldrich
Isobutylmethylxanthine	Sigma-Aldrich

Table 2.3 Immunocytochemistry, histochemistry, and fluorescence labelling reagents

Item (and dilution if appropriate)	Source
Anti-β-tubulin (1:100)	Sigma-Aldrich
Anti-OCN (1:100)	Santa Cruz Biotechnology
Anti-VCAM1 (1:100)	Santa Cruz Biotechnology
Anti-STRO-1 (1:100)	Santa Cruz Biotechnology
Anti-ALP (1:100)	Santa Cruz Biotechnology
Anti-FN (1:100)	Sigma-Aldrich
Anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (1:100)	Sigma-Aldrich
Anti-OPN (1:100)	Santa Cruz Biotechnology
Anti-vinculin (1:100)	Sigma-Aldrich
Rhodamine phalloidin (1:500)	Invitrogen
Fluorescein streptavidin (1:50)	Vector Laboratories
Biotinylated immunoglobulin G (IgG) (1:50)	Vector Laboratories
IRdye 680 (1:800)	Li-Cor
IRdye 800 (1:800)	Li-Cor
Vectashield mountant with 4',6-diamidino-2-phenylindole (DAPI)	Vector Laboratories
Bovine serum albumin (BSA)	Sigma-Aldrich
Tween 20	Sigma-Aldrich
PBS	Sigma-Aldrich
Triton-X100	Sigma-Aldrich
$MgCl_2$ hexahydrate	VWR Chemicals
NaCl	VWR Chemicals
Sucrose	Fisher Scientific
Formaldehyde (37-41%)	Fisher Scientific
Li-Cor blocking buffer	Li-Cor
CellTag 700 (1:500)	Li-Cor
Methanol	VWR Chemicals
LIVE/DEAD Viability/Cytotoxicity Kit	Invitrogen
Oil red O	Sigma-Aldrich
Isopropanol	Sigma-Aldrich

2.2.2 Recipes

Versene

Water	1000 mL
NaCl	8 g
KCl	0.4 g
Glucose	1 g
HEPES	2.38 g
EDTA	0.2 g
0.5 % Phenol red	2 mL
Adjusted to pH	7.5

2.5 % trypsin solution

Trypsin	0.5 mL
Versene	20 mL

Basal media (cell lines)

DMEM	400 mL
FBS	50 mL
Medium 199	100 mL
Penicillin-streptomycin	10 mL
L-glutamine 200 mM	5 mL
Sodium pyruvate	5 mL

Basal media (MSC)

DMEM	500 mL
FBS	0 (0 mL), 2 (10 mL), 5 (25 mL), or 10 (50 mL) %
Penicillin-streptomycin	10 mL
MEM non-essential amino acids	5 mL
Sodium pyruvate	5 mL

Osteogenic differentiation media

Basal media	500 mL
Ascorbate-2-phosphate	350 µM
Dexamethasone	0.1 µM

Adipogenic differentiation media

Basal media	500 mL
Insulin	1.7 nM
Indomethacin	200 µM
Isobutylmethylxanthine	500 µM
Dexamethasone	1 µM

Fixation solution

PBS	90 mL
Formaldehyde (37-41 %)	10 mL
Sucrose	2 g

Permeabilisation buffer

PBS	100 mL
Sucrose	10.3 g
NaCl	0.292 g
$MgCl_2$ hexahydrate	0.06 g
HEPES	0.476 g
Adjusted to pH	7.2
Triton X	0.5 mL

0.5 % Tween 20 solution (wash buffer)

PBS	100 mL
Tween 20	0.5 mL

Blocking solution for immunofluorescence microscopy

PBS	100 mL
BSA	1 g

Oil red O staining stock solution

Oil red O	150 µL
Isopropanol	5 mL

Oil red O staining solution

Oil red O staining stock solution	3 mL
Deionised water	2 mL
Syringe filtered	

2.3 Cells

2.3.1 Mesenchymal Stem Cells

MSCs were provided by Bone and Joint Research Group, Institute of Developmental Sciences, University of Southampton, UK. MSCs were isolated by STRO-1 (in-house antibody) magnetic cell separation from bone marrow of haematologically-normal consenting anonymised patients undergoing orthopaedic surgery at Southampton General Hospital, and were transported as growing cells under chilled conditions. They were then maintained at 37 °C in stem cell basal media (BM) with 10 % FBS, which was changed twice weekly. Cells were split 1:3 before confluence was reached. No MSCs beyond passage 5 were used.

For dissociation, cells were rinsed in 37 °C HEPES before incubation at 37 °C in 2.5 % trypsin solution for 5-10 mins with regular agitation. Trypsin was inactivated with excess BM and cells were pelleted by centrifugation at 1400 RPM for 5 mins. After discarding supernatant, the cell pellet was re-suspended immediately in the appropriate medium for seeding.

MSCs used in the IR study of differentiation in chemical induction media were cultured by Dr Lesley-Anne Turner, University of Glasgow.

2.3.2 Saos-2 Cells

Saos-2 cells were cultured by incubation at 37 °C in BM, which was changed twice weekly. Cells were split 1:10 before confluence was reached. Dissociation was by the same method as MSCs.

2.3.3 Fabrication of Artificial ECMs

Bulk PEA with 1 % benzoin or PMA with 0.35 % benzoin was provided by the Universitat Politecnica de Valencia (UPV), Spain, or produced in-house (Mark Sprott, Division of Biomedical Engineering, University of Glasgow, UK). 2.5 % or 4 % PEA and 6 % PMA w/v solution was made by dissolving the corresponding amount of bulk polymer in toluene. 100 µL of polymer solution was spun coat at 3000 RPM at 3000 RPM s^{-1} for 30 s onto 12 mm diameter ethanol-cleaned circular glass coverslips or 13 mm diameter circular CaF_2 windows. Samples were then dried at 60 °C in a vacuum for 2 h and stored at RTP. Then, samples were immersed in 200 µL of a 20 µg/mL human plasma FN in PBS for 1 h at room temperature. Growth factors were then adsorbed in the same way using 25 or 50 ng mL^{-1} VEGF-165 (VEGF) or BMP2; 25 ng mL^{-1} was used for experiments up to optimisation, and 50 ng mL^{-1} used thereafter, unless noted.

2.3.4 Atomic Force Microscopy

FN conformation images were produced using atomic force microscopy (AFM) (JPK Nanowizard 3; Zeiss Axio Observer A1; Accurion Halcyonics_i4 balance table) in tapping mode on dry FN-coated samples (section 2.3.3) washed several times in deionised water. A pyramidal-tip probe (MPP-21220, Bruker) was used. Force spectroscopy was performed on submerged samples after immersion in deionised water overnight at room temperature, using 5 µm silica beads glued to tip-less cantilevers (Nanosensors TL-FM-10). AFM done with assistance from Eleni Grigoriou, Annie Cheng, Marco Cantini, and Virginia Llopis-Hernandez (Division of

Biomedical Engineering, University of Glasgow). A Kruskal-Wallis test was used to compare conditions.

2.3.5 Culture on Artificial ECMs

Following protein adsorption but prior to cell seeding, samples were stored in PBS. Cells were seeded at specific densities in various culture media and cultured for up to 21 d. Except where mentioned, cells were seeded at $1x10^3$ cells cm^{-2}. Cells were seeded in serum-free BM for 2 h before culture in serum-containing media where noted.

2.3.6 Fluorescence Microscopy

Samples were washed with warm PBS then immersed in 37 °C fixing solution for 15 min, 4 °C permeabilisation solution for 5 min, and 37 °C blocking buffer for 5 min. Cells were incubated in primary antibody diluted in blocking buffer and optionally rhodamine phalloidin, then washed in wash buffer before repeating the process with secondary (biotinylated IgG) and tertiary (fluorescein streptavidin) stains without phalloidin. Samples were mounted in VECTASHIELD mounting medium with 4',6-diamidino-2-phenylindole (DAPI) on a microscope slide. Prepared samples were stored at 4 °C. Mosaic images were taken using an Olympus BX51 coupled to a Prior L200/D automated stage, controlled using Surveyor software and an OASIS controller (Objective Imaging). Vinculin images were taken using a Zeiss Axiophot. Exposure, gain, and offset are constant for all images of an experiment. Contrast was adjusted for display purposes and images were cropped using Fiji (Schindelin et al. 2012).

2.3.7 In-Cell Western

In-Cell Western (ICW) is a high-throughput dual-channel near-infrared fluorescence assay (Aguilar et al. 2010). A representative ICW image of samples in a well plate is shown in Figure 2.1. The red 700 nm channel (Figure 2.1 (a) top) is fluorescence from CellTag, a whole-cell stain, or a dye binding to an anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) antibody, and the green 800 nm (Figure 2.1 (a) bottom) represents expression of the protein of interest.

Negative controls are samples without the normalisation stain or antibody against the protein of interest (Figure 2.1 (b)). A ratio of the green-to-red fluorescence is calculated to measure protein expression normalised to total cell volume or control protein expression. Single channel fluorescence can be reported as well. Compared to whole-sample immunofluorescence microscopy, ICW is higher throughput but has a lower resolution.

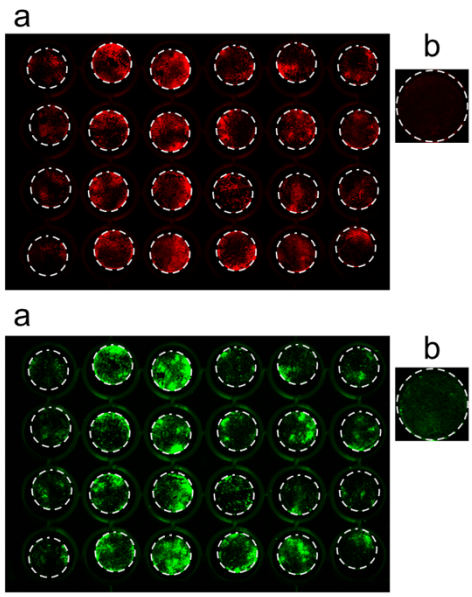

Figure 2.1 ICW example of samples in a 24 well plate (a) Example of ICW scanned image in 700 (top) and 800 (bottom) nm channels. Each well is a separate sample, allowing for high-throughput multiple comparisons. (b) Negative control showing low red (top) and green (bottom) background.

Samples were prepared and stained according to section 2.3.6 using Li-Cor ICW blocking buffer, CellTag, and secondary antibodies. ICW analyses of serum-free and low-serum conditions used GAPDH as the normalisation factor; all other ICW assays used CellTag. For cell-based ICW assays with more than one donor, a mixed effects linear model which controlled for the within-donor variation was used (using the formula syntax in R: protein expression ~ artificial ECM + (1 | donor)) (Krzywinski et al. 2014; Aarts et al. 2014). All others were analysed using a permutation analysis of variance (ANOVA) with pairwise comparisons. Holm p-value adjustment was used for multiple hypothesis test correction. A Kruskal-Wallis test followed by pairwise comparisons using Dunn's-test was used to analyse the FN and BMP2 ICW assays.

2.3.8 Cell Viability Assays

Samples were stained using the LIVE/DEAD Cell Viability/Cytotoxicity kit as per the manufacturer's instructions. Briefly, live samples were washed and incubated in calcein AM/ethidium homodimer staining solution under culture conditions for 30 min, and then washed, immersed in PBS, and immediately imaged using a Zeiss Axiovert 200m.

2.3.9 Infrared Spectroscopy

Fourier transform infrared spectroscopy using synchrotron radiation (SR-FTIR) is a label-free approach to chemically profile a substance. Different biological molecules with an electric dipole moment have distinct infrared signatures (Baker et al. 2014), so by comparing the absorbance of infrared wavelengths between samples inferences about substrate or cell biochemistry can be made.

SR-FTIR was performed at Diamond Light Source (Oxford, UK), using a Bruker Hyperion 3000 microscope with a Bruker Vertex 80 V Fourier Transform IR Interferometer, a 36 x IR objective, and mercury cadmium telluride detector, recording 256 co-scans at 4 cm^{-1} resolution. Measurement coordinates were chosen using 20x magnification brightfield microscopy. Data was pre-processed by removing CO_2 absorbance, correcting for negative absorbance (an artefact of SR-FTIR), correcting for scatter using RMieS-EMSC (Bassan et al. 2010) (only for cell spectra), removing wavenumbers outside the range 3700-1000 cm^{-1}, and vector normalising (Trevisan et al. 2012; Baker et al. 2014). Assignments of wavenumbers to chemical groups was based on matches to the literature on biochemical spectra (Movasaghi et al. 2008) and so were considered tentative. SR-FTIR was performed with the assistance of Dr Lesley-Anne Turner (Centre for Cell Engineering, University of Glasgow, UK); Dr Matthew Baker and Dr Holly Butler (University of Strathclyde, UK); and Dr Chris Kelley, Dr Katia Wehbe, Dr Ann Fitzpatrick, and Dr Gianfelice Cinque (B22, Diamond Light Source, Oxford, UK).

2.3.10 Statistical Analysis Tools

FTIR data pre-processing was performed using a pipeline built using MATLAB. All statistical analyses were performed using R with the following packages: ggplot2 (Wickham 2016), dplyr, plyr (Wickham 2011), reshape2, PMCMR, lmTest, lme4 (Bates et al. 2015), MASS, XLConnect, tidyr, JADE, lmperm, lsmeans, and ggsignif.

2.3.11 Oil Red O staining

Samples were washed with warm PBS and then immersed in 37 °C fixing solution for 15 min. Samples were quickly immersed in 60 % isopropanol, which was immediately removed, and then allowed to air dry. Oil red O solution was added to cover samples for 15 min at room temperature. The staining solution was then removed and samples were washed with water until clean before imaging using a Zeiss Axiovert 25 and a QImaging Micropublisher 3.3 RTV colour camera.

2.4 Results

2.4.1 Characterisation of the Artificial ECMs

AFM was used to show the distribution of FN adsorbed to PMA and PEA on glass and calcium fluoride (CaF_2). Glass is the usual substructure for assembling the artificial ECMs, and CaF_2, needed for IR analysis, was novel. Figure 2.2 shows the results on glass, and Figure 2.3 on CaF_2. On both substructures, FN had an isolated arrangement on PMA, whereas on PEA it formed a network. These results matched previous data (Llopis-Hernández et al. 2013; Rico et al. 2016; Guerra et al. 2010; Salmerón-Sánchez et al. 2011; Vanterpool et al. 2014; González-García et al. 2012; Llopis-Hernández et al. 2016; Moulisová et al. 2017) to demonstrate reproducibility, and also showed that the artificial ECMs can be made on CaF_2.

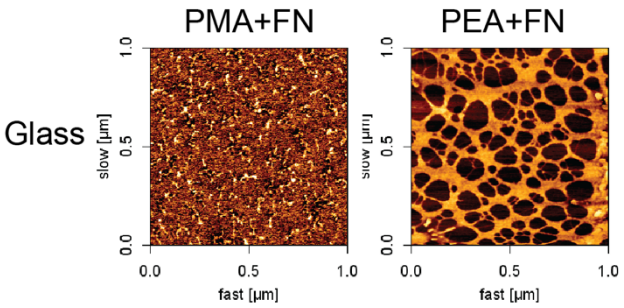

Figure 2.2 AFM images of glass-based FN matrices. Note the isolated conformation on PMA and the connected network-like conformation on PEA.

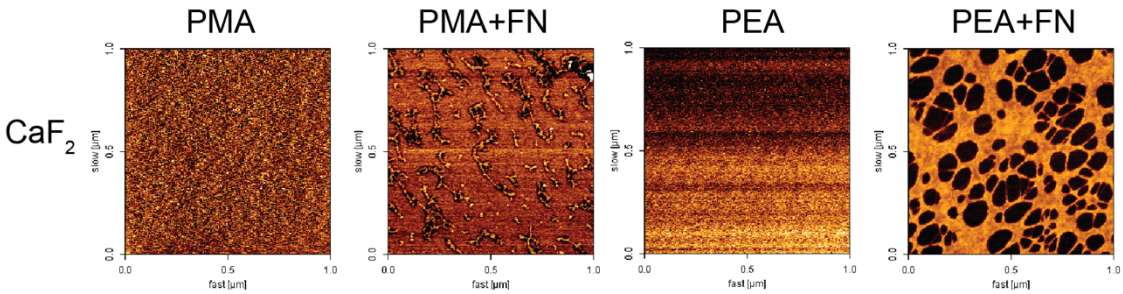

Figure 2.3 AFM images of CaF_2-based FN matrices. Polymer-only substrates (PMA, PEA) are shown for comparison. Note the isolated conformation on PMA and the connected network-like conformation on PEA.

Substrate elasticity can regulate MSC differentiation (Engler et al. 2006). The elastic modulus of each surface was calculated using force spectroscopy, which was not significantly different between PMA and PEA (Figure 2.4 (a)). The amount of FN adsorbed to the polymers did not differ, as indicated by ICW assay using a polyclonal antibody (Figure 2.4 (b)), suggesting the primary difference in adsorbed FN between PMA and PEA was its conformation. BMP2 binding to FN-coated surfaces was also examined by ICW (Figure 2.4 (c)); more BMP2 was observed on PMA with FN than the other two conditions. Note that BMP2 binds directly to the polymer film (Llopis-Hernández et al. 2016) and this assay did not specifically measure the FN-bound fraction that may have changed with FN conformation on PEA or PMA.

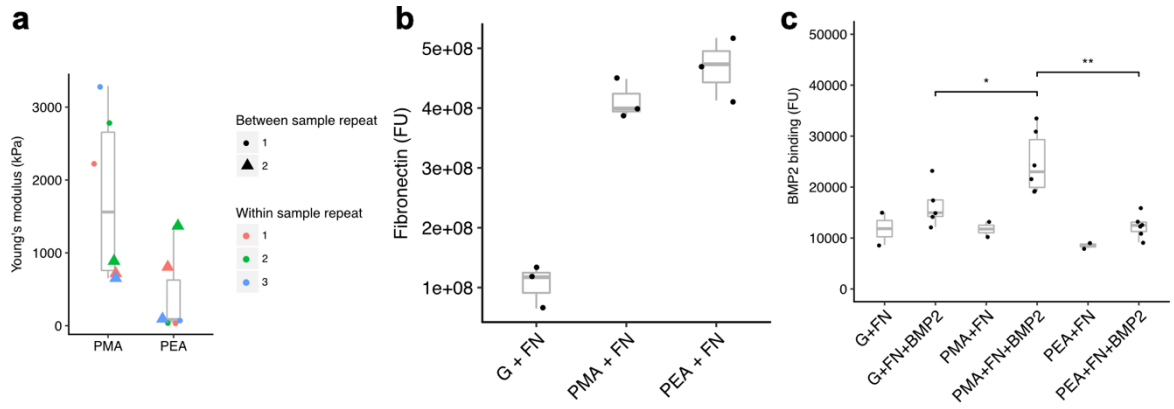

Figure 2.4 Characterising the artificial ECMs: (a) Force spectroscopy of spun-coat PMA and PEA (n = 6); (b) ICW quantification of FN adsorption (n = 3); and (c) ICW quantification of 50 ng mL^{-1} BMP2 binding to FN adsorbed to glass, PMA, and PEA (n = 5 (glass) or 6); conditions without BMP2 were negative controls for anti-BMP2 antibody binding (n = 2). G = glass. Compared by Kruskal-Wallis rank sum test followed by a pairwise test for multiple comparisons of mean rank sums (Dunn's-test). * = p < 0.05; ** = p < 0.01.

SR-FTIR was used to compare the chemistry of the blank polymers, elucidate any differences in FN conformation across the polymers, and study the addition of BMP2 to the network conformation (VEGF was not measured due to sample size limitations) (Baujard-Lamotte et al. 2008; Barth 2007; Steiner et al. 2007). Establishing the chemical signature of the cell-free substrates was also important as a background measurement for later cell IR analysis (section 2.4.5). Figure 2.5 shows the absorbance of the artificial ECMs. The polymers were distinguished by a number of regions: 3000-2900 cm^{-1} (stretching of methyl [CH_3] and methylene [CH_2] groups); 1500-1300 cm^{-1} (various CH, CH_2, and CH_3 deformations); 1300-1000 cm^{-1} (vibrations unique to the whole molecule). There is a common strong absorbance at 1800-1700 cm^{-1} (C=O). Tentative assignments of wavenumbers to functional groups was by comparison to a biological FTIR database (Movasaghi et al. 2008). The PEA and PMA spectra were also similar to NIST reference FTIR spectra for ethyl and methyl acrylate monomers (NIST Mass Spec Data Center 2017). Of note was the variation of absorbance within samples at these distinguishing bands, which may be due to uneven polymer distribution. Adsorbed FN was not detected on either polymer; no features corresponding to biological spectra (beyond the underlying polymer fingerprint) were observed (Baker et al. 2014). Similarly, BMP2 was not detected, suggesting that the amount of FN and BMP2 was too low for comparison by SR-FTIR, at least relative to the polymer IR signature.

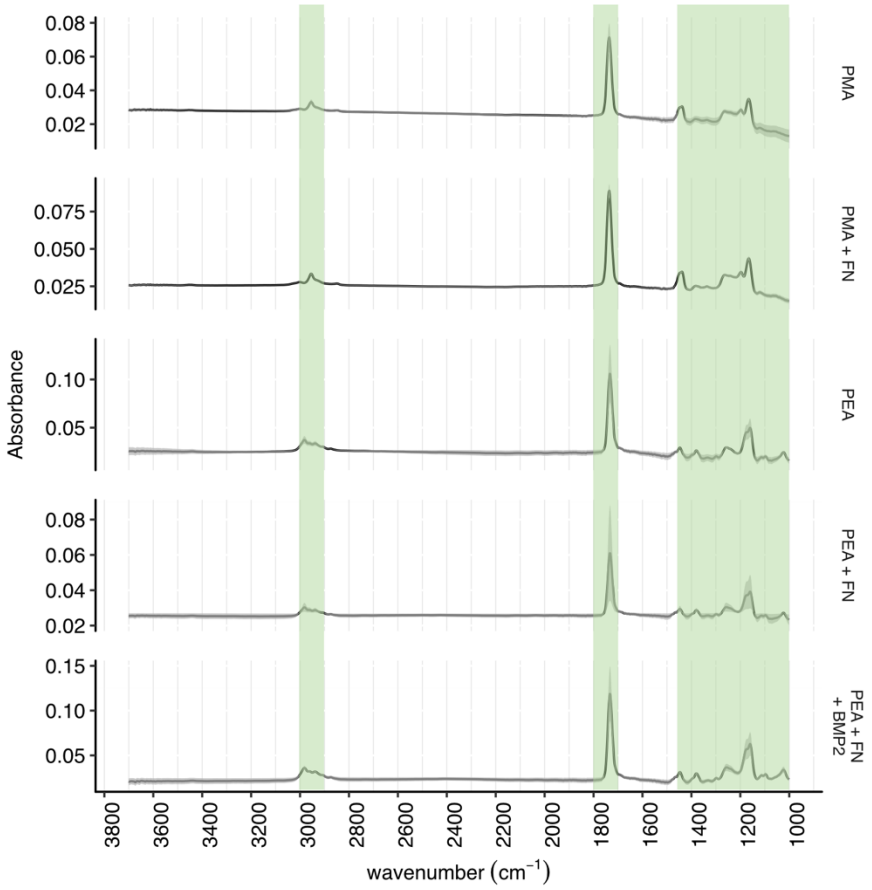

Figure 2.5 SR-FTIR absorbance of cell-free artificial ECMs. Mean (black line) ± standard deviation (grey shadow) of multiple spectra (n > 30) taken on one sample of each type. Green regions indicate notable wavenumbers.

2.4.2 Characterising Cell Culture on the Artificial ECMs

2.4.2.1 Saos-2 Cells on Surfaces

Saos-2 cells (Saos, hereafter) are used as a model for osteogenesis (Schröder et al. 2005; McQuillan et al. 1995; Lian & Stein 1992; Pautke et al. 2004; Stein & Lian 1993), and they proliferate as an adherent monolayer in the same conditions as MSCs (with a slight difference in the BM recipe) (McQuillan et al. 1995), and have been used in biomaterial studies (Okumura et al. 2001; Liu et al. 2009). These attributes made them ideal for testing cell adhesion, growth, and osteogenesis on the artificial ECMs prior to using less readily-available MSCs.

After 21 d, immunofluorescent mosaic images of Saos stained for actin, nuclei, and OCN, a marker for osteogenesis (Pittenger et al. 1999; Kulterer et al. 2007; Lian & Stein 1992; Stein & Lian 1993), were taken. In general, Saos formed a confluent monolayer with some regions of lower density, demonstrated by a representative image in Figure 2.6. The variation was likely due to a

combination of cell seeding, behaviour, proliferation and sample immersion throughout culture — samples would occasionally float and partially surface from their culture media. The actin cytoskeletons were clearly formed, and OCN was expressed across all surfaces (Figure 2.7). These results show that the artificial ECMs reproduced for this work maintain cells for 21 d and support osteogenic differentiation of osteogenic cells.

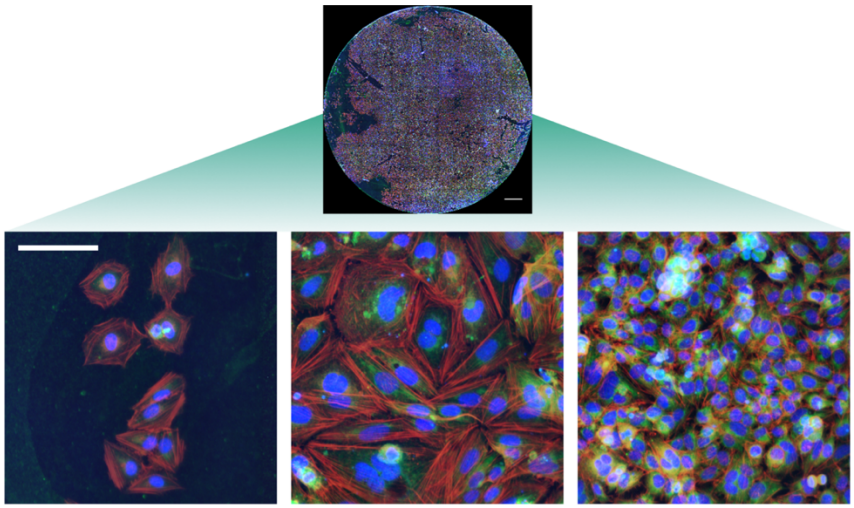

Figure 2.6 Representative image of Saos cultured for 21 d, stained for actin, nuclei, and OCN. Highlighted regions demonstrate the range of cell density. Whole sample (top) scale bar 1 mm. Lower images scale bar 100 µm. Red, F-actin; blue, nuclei; green, OCN.

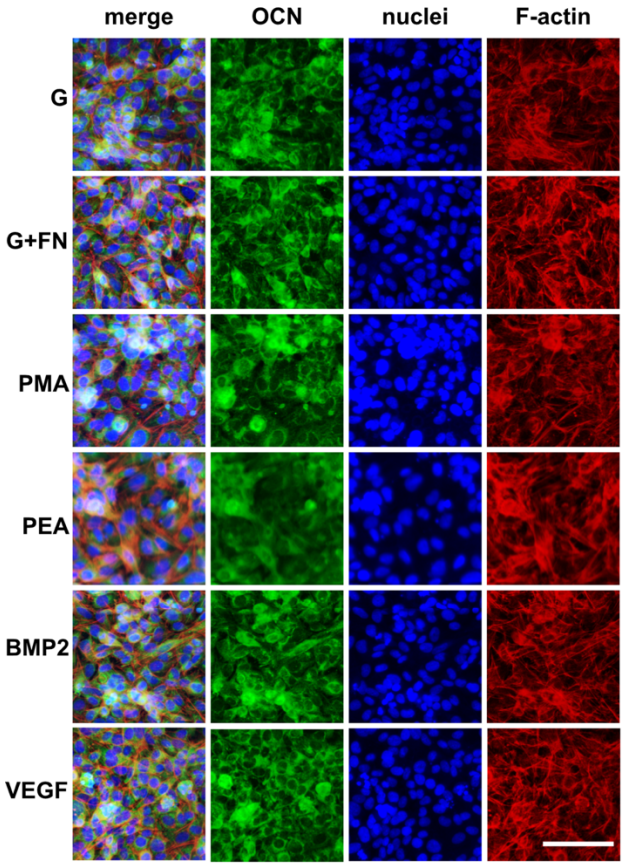

Figure 2.7 OCN, DAPI, and actin in Saos over all conditions after 21 d. All artificial ECMs supported Saos adhesion, growth, and osteogenic marker expression for 21 d. G = glass,

PMA/PEA = PMA/PEA + FN, BMP2/VEGF = PMA/PEA + FN + BMP2/VEGF. Scale bar 100 µm. Red, F-actin; blue, nuclei; green, OCN. n = 2.

2.4.2.2 MSC Characterisation

Before evaluating MSC behaviour on the artificial ECMs they were characterised and tested for their ability to differentiate — a criteria for validating MSCs (Dominici et al. 2006). They readily adhered to glass and adopted a typical flat, spread, and stellate morphology (Figure 2.8) (Baksh et al. 2004), dividing and spreading over 21 d (Figure 2.10 top left). Within 24 h MSCs had developed clear actin cytoskeletons with various morphologies as indicated by immunofluorescence microscopy (Figure 2.9).

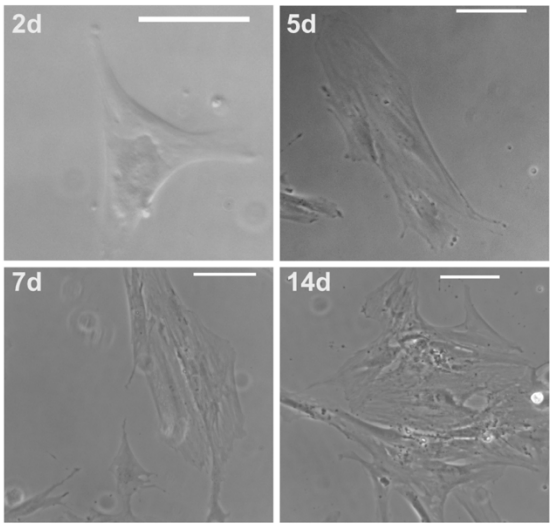

Figure 2.8 Phase contrast microscopy of MSCs cultured on glass over time in BM. MSCs readily adhered, spread and grew on glass when cultured in BM. Scale bar 100 µm.

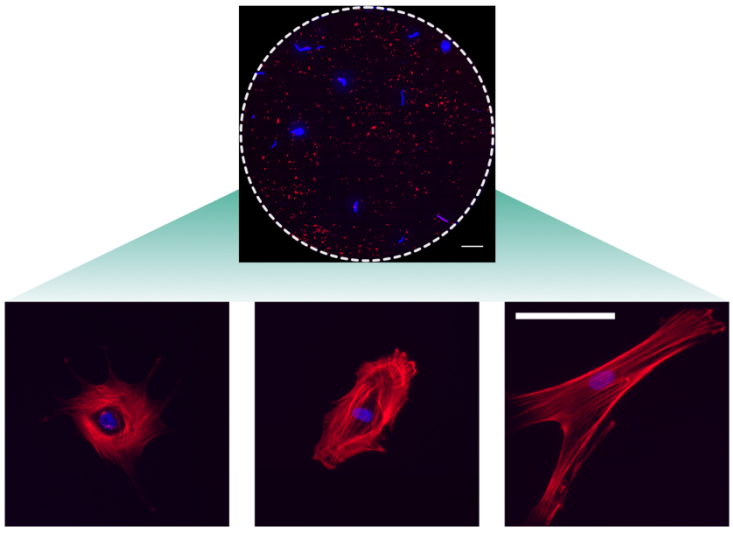

Figure 2.9 Representative image of MSCs cultured for 24h in BM, stained for actin and nuclei. The actin cytoskeletons of MSCs formed within 24 h of adhesion highlighted a

variety of morphologies. White dashed circle highlights the edge of the sample. Whole sample (top) scale bar 1 mm. Lower images scale bar 100 µm. Red, F-actin; blue, nuclei.

The multipotency of these MSCs was tested by culturing in adipogenic (AM) or osteogenic (OM) induction media; MSCs are expected to have lineage-specific features following 21 d culture under these conditions (Pittenger et al. 1999; Lian & Stein 1992; Kulterer et al. 2007; Dominici et al. 2006; Stein & Lian 1993). After 21 d, MSCs cultured in BM formed a dense and flat monolayer (Figure 2.10 top left), whereas in OM they additionally formed several nodule-like features (Figure 2.10 top centre). These nodules stained positively for OCN (Figure 2.10 bottom left, bottom centre). Lipid droplets had accumulated in several cells cultured in AM indicating that adipogenesis had occurred (Pittenger et al. 1999). These differentiation events — lipid accumulation and nodule-like formations — were infrequent as many cells retained a similar morphology as those cultured in BM, similar to other reports of STRO-1$^+$ cells (Gronthos et al. 1994; Simmons & Torok-Storb 1991; Pittenger et al. 1999).

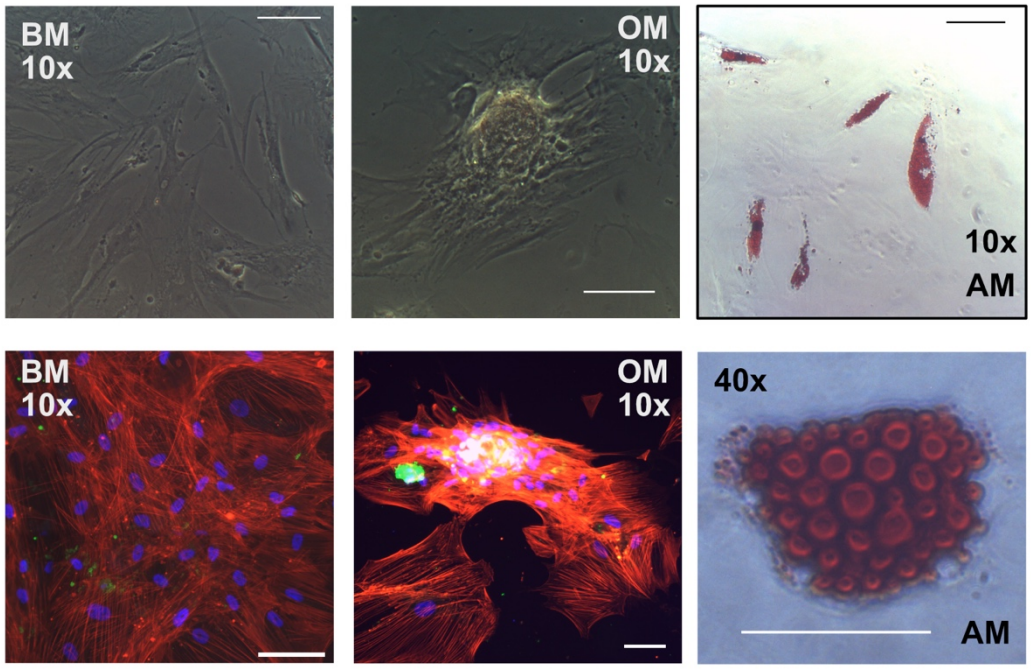

Figure 2.10 MSC differentiation after 21 d culture on glass in BM, AM, or OM. In BM, MSCs proliferated and spread (top left), whereas in OM (top right) they also formed nodule-like features, as indicated by phase contrast microscopy. High density of MSCs in BM (bottom left) and OM (bottom centre) was also indicated by fluorescent microscopy of actin (red) and nuclei (blue), and MSCs in the nodule-like features in OM also expressed OCN (green). Regions of MSCs cultured in AM (top right, bottom right) stained positively for indicated lipid droplets (oil red O). 10x objective scale bar 100 µm. 40x objective scale bar 50 µm. n = 2.

SR-FTIR was used to compare the biochemical profiles of differentiating MSCs (Figure 2.11). While the average spectra were indistinguishable, the variance did discriminate conditions: lipid-associated wavenumbers (3000-2800 cm^{-1}) identified MSCs cultured in AM, and the fingerprint region (< 1450 cm^{-1}) (Baker et al. 2014) was different for each culture. For later studies of MSCs cultured across the artificial ECMs, these biochemical profiles would have been a useful baseline for comparison but the differences between differentiation conditions were not obvious enough. Altogether, these data are coincident with and indicate that these were typical MSCs (Baksh et al. 2004; Dominici et al. 2006).

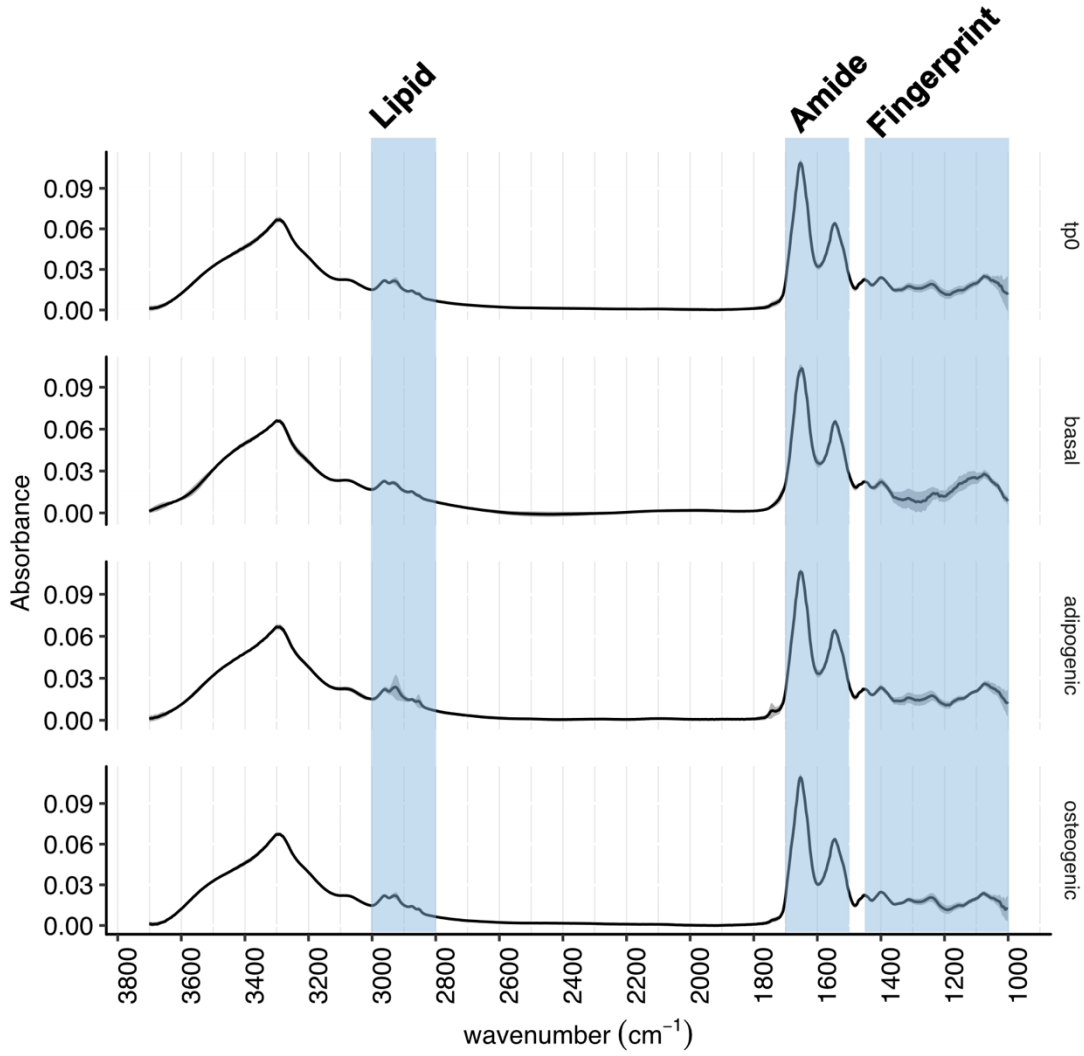

Figure 2.11 SR-FTIR absorbance of MSCs at time-point 0 (tp0) or cultured in differentiation media for 21 days. MSCs cultured in AM (adipogenic) had variable absorbance in the lipid region. Variation in the fingerprint region also distinguished differentiation conditions, which indicated different phenotypes. Biologically relevant absorbance regions are highlighted by the blue labelled boxes. Mean (black line) ± standard deviation (grey shadow) of multiple spectra (n = 2 biological replicates, n = 1036 spectra in total).

2.4.3 MSC Culture on the Artificial ECMs

In the previous section the MSCs were characterised. In this section, their adhesion, growth, and differentiation across the artificial ECMs are compared. The effect of varying culture conditions is also studied.

After 24 h culture on the artificial matrices the actin cytoskeletons of MSCs were imaged (Figure 2.12). Across all surface combinations, including control glass and glass with FN, actin cytoskeletons formed, and cells spread to adopt a polygonal morphology.

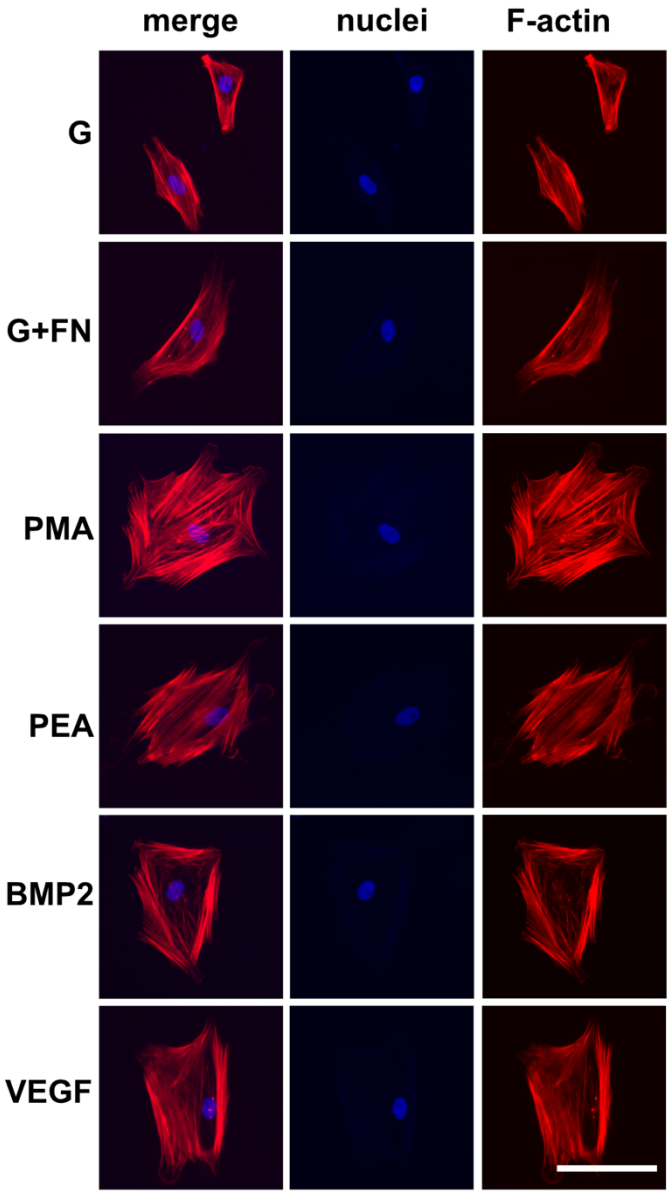

Figure 2.12 MSCs cultured for 24 h in BM on the artificial ECMs, stained for actin and nuclei. A range of morphologies can be seen. The MSCs were spread and had clear actin cytoskeletons. Scale bar 100 μm. Red, F-Actin; blue, nuclei. n = 2.

ß-Tubulin is a cytoskeletal protein with diverse roles including organelle localisation, cell division, and metabolism (Cassimeris et al. 2012; Dutcher 2001; Oakley 2000). After 3 d culture, cells remained spread and networks of ß-tubulin were evident across all surface combinations (Figure 2.13).

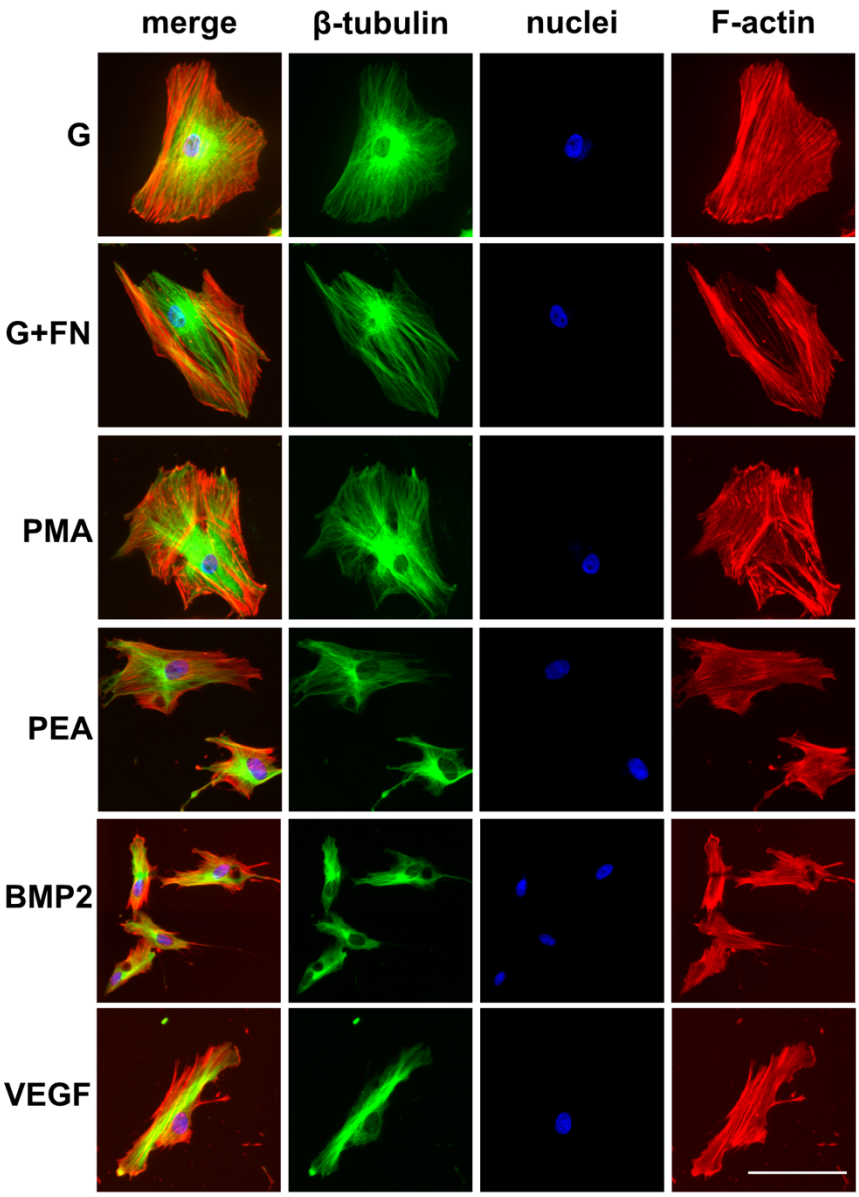

Figure 2.13 MSCs cultured for 3 d on the artificial ECMs, imaged for actin, nuclei, and β-tubulin. Microtubule and actin cytoskeletons were formed at 3 d culture on all artificial ECMs. Scale bar 100 µm. Red, F-actin; blue, nuclei; green, β-tubulin. n = 2.

Adhesions of MSCs to the surfaces were compared using vinculin staining. Vinculin is a focal adhesion protein that is central to recruiting various cell signalling components (Carisey & Ballestrem 2011; Geiger et al. 2001). As demarcated by vinculin, focal adhesions were found in cells on all surfaces (Figure 2.14).

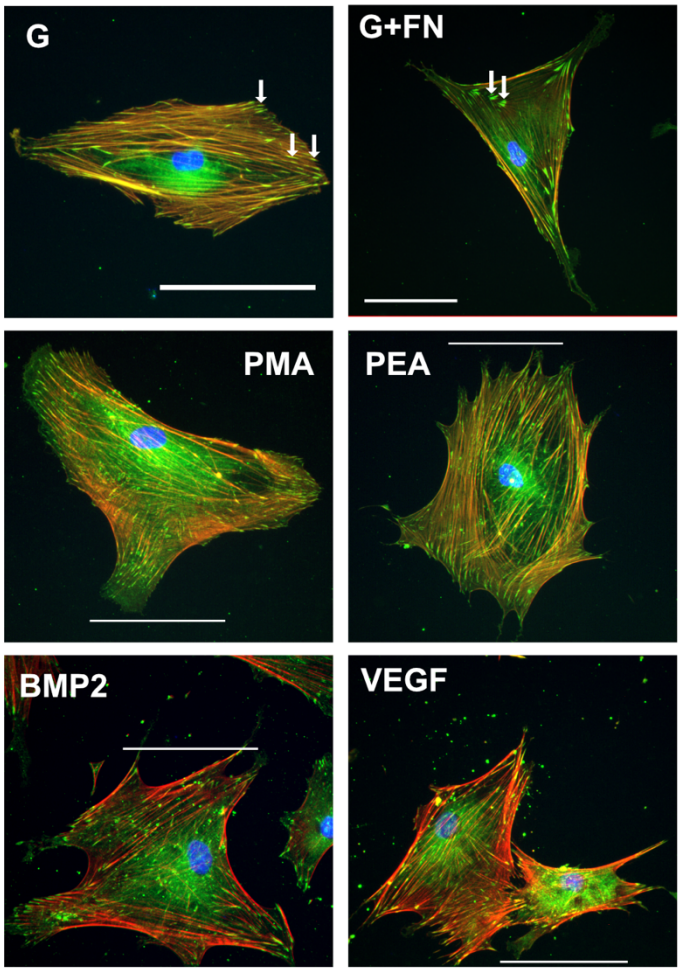

Figure 2.14 Vinculin distribution in MSCs cultured for 3 d on artificial ECMs. Focal adhesion formation occurs in MSCs adhered to all the artificial ECMs. White arrows indicate example focal adhesions, regions of bright vinculin staining at the terminals of actin fibres. Scale bar 100 μm. Red, F-actin; blue, nuclei; green, vinculin. n = 2.

At these early time points — 24 h and 3 d, early relative to the duration of culture for differentiation (21 d) — there is little cell-to-cell contact. After 21 d, cells had proliferated resulting in cell-to-cell contact with some regions of indistinguishable cell boundaries, as visualised by immunofluorescence microscopy of their actin cytoskeletons (Figure 2.15). There was a range of cell densities, and punctate OCN staining was observed on all matrices. Although a dense region of OCN expression was observed in MSCs cultured on PMA with FN, its pattern was not similar to the nodule-like formations observed in OM (Figure 2.10). Note also that this only occurred on a small region on one of two samples.

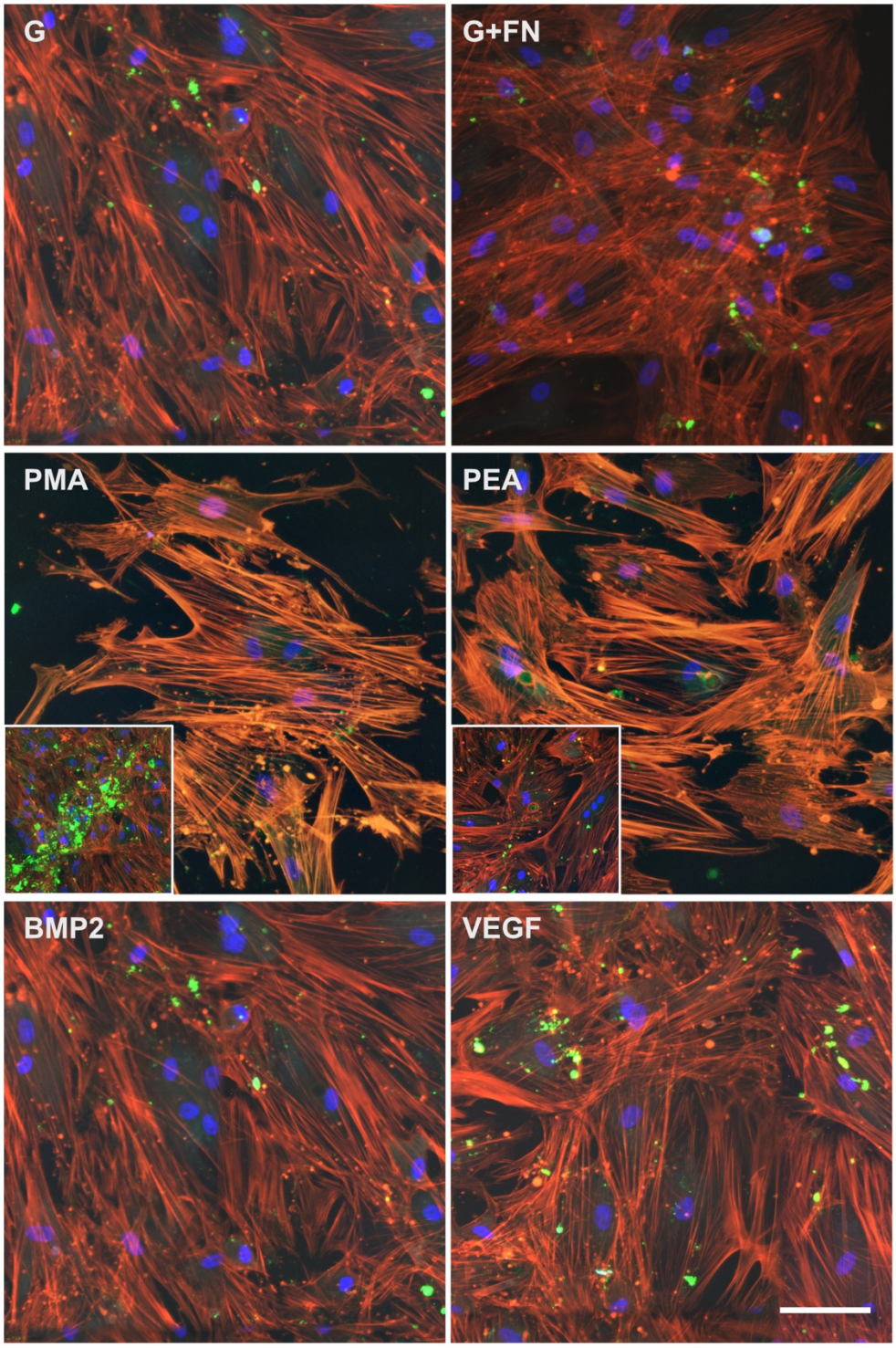

Figure 2.15 OCN expression in MSCs cultured for 21 d in BM on the artificial ECMs. MSCs have low amounts or negative OCN expression after 21 d culture on the artificial ECMs. Note there were no obvious nodule-like features similar to that associated with late-stage osteogenesis. PMA inset: region of relatively greater OCN expression and cell density. PEA inset: region with higher cell density. Red, F-actin; blue, nuclei. Scale bar 100 μm. n = 2.

Given that MSCs on the artificial ECMs did not produce significant bone matrix (Figure 2.15), and yet they have the capacity to do so (Figure 2.10), their ability to undergo osteogenesis on the surfaces in OM was tested. MSCs were cultured for 21 d in OM on the artificial ECMs and their OCN expression visualised with

immunofluorescence microscopy. MSCs on control glass and glass with FN displayed nodule-like features that stained with OCN, but this did not occur on the artificial ECMs (Figure 2.16). Note again that punctate OCN expression was observed on the artificial ECMs.

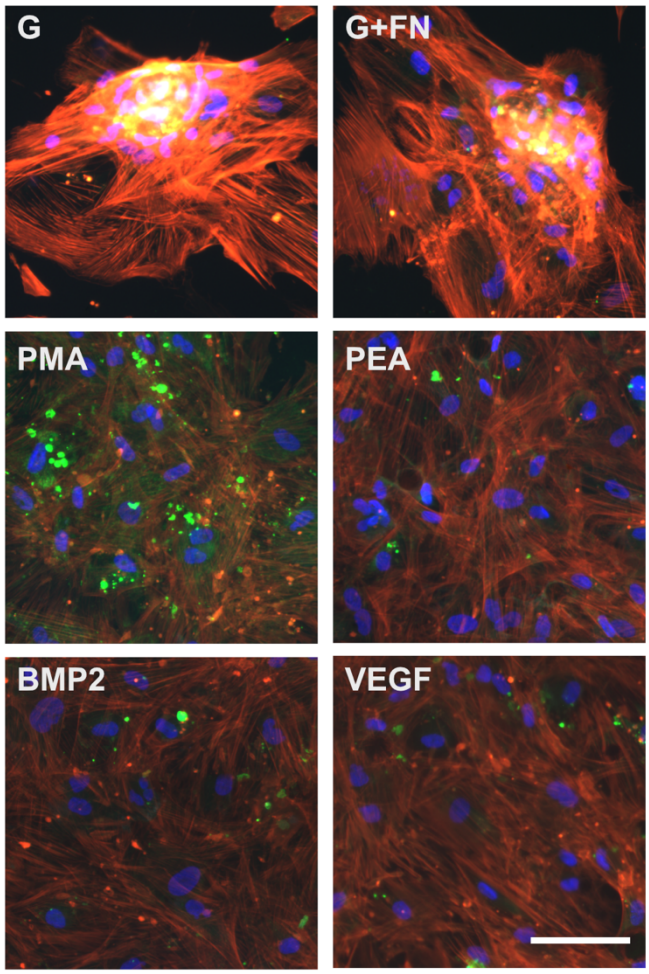

Figure 2.16 OCN staining of MSCs cultured on the artificial matrices in OM. MSCs on glass and glass with FN displayed nodule-like features that stained positively for OCN, indicating osteogenesis had occurred. Low or no expression of OCN occurred on the polymer-based artificial ECMs. Scale bar 100 µm. Red, F-actin; blue, nuclei; green, osteocalcin. n = 2.

To further evaluate differentiation, the expression of osteogenic — OCN and OPN, another osteogenic marker — and stemness — VCAM1 (Stewart et al. 1999; Gronthos et al. 2003) and NESTIN (Pinho et al. 2013; Méndez-Ferrer et al. 2010; Isern et al. 2013; Ding et al. 2012) — proteins were quantified by ICW after 21 d in culture on the artificial matrices. This was repeated for 3 different donors (Figure 2.17). Note that this experiment was missing the G with FN substrate because it was started before it was considered as a useful control. Both OPN and NESTIN had different expression in at least some of the comparisons (Figure 2.18). NESTIN expression was estimated to be higher on glass in comparison to

the other surfaces. PMA with FN had a lower expression of OPN than PEA with FN and glass. No differences were estimated with respect to OCN and VCAM1 expression. These results provided some evidence that MSCs were differentiating to an osteoblastic lineage on PEA with FN, and perhaps to other lineages on the other matrices as judged by the decrease in NESTIN compared to the control glass matrix. However, while the role of NESTIN and VCAM1 in MSC stemness is unclear (Chan et al. 2015), both OCN and OPN are expected to increase during osteogenesis (Kulterer et al. 2007; Lian & Stein 1992) and yet only OPN is shown to be differentially expressed.

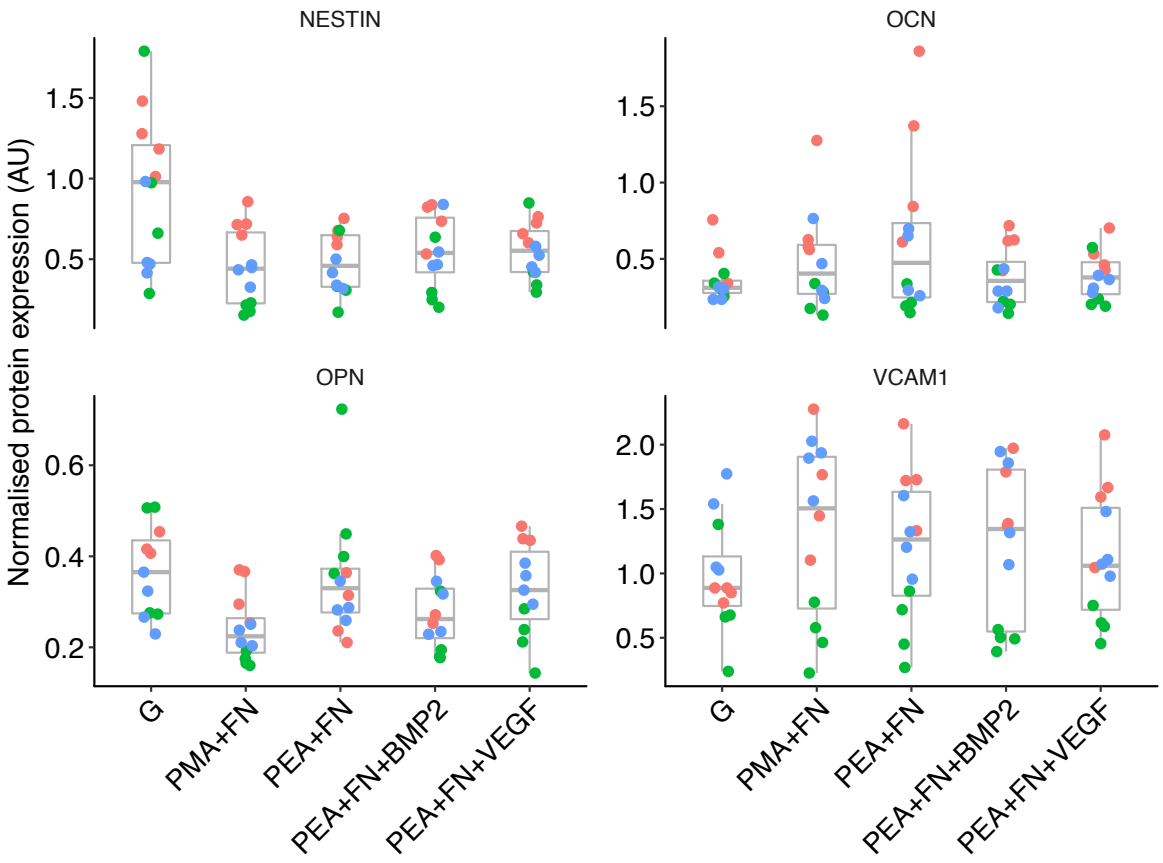

Figure 2.17 ICW analysis of osteogenic (OCN, OPN) and stemness (VCAM1, NESTIN) marker expression normalised to total cell volume by MSCs after 21 d culture on artificial ECMs. Note relatively large intra- and inter- donor variation. n = 3 biological replicates (donors) distinguished by colour; n = 3 or 4 repeats per donor. AU, arbitrary units.

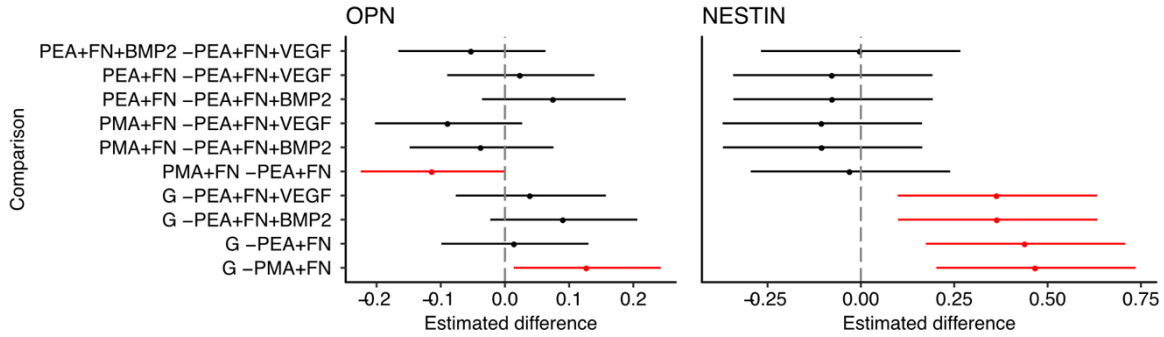

Figure 2.18 Comparison of the difference between protein expression in MSCs on the artificial ECMs (mean ± 95 % confidence interval) from Figure 2.17. MSCs cultured on PMA had a lower expression of OPN than on PEA. MSCs on glass had a higher expression of NESTIN compared to the other artificial ECMs. Estimates are calculated from a mixed effects linear model that controlled for within-donor variation. Red, $p < 0.05$.

2.4.4 Optimising Culture Conditions to Improve Differentiation

It is posited that FN from serum contained in culture media masks the FN conformations of the artificial ECMs to interfere with the interaction between cells and the underlying FN pattern (González-García et al. 2012; Rico et al. 2016), so previous experiments using osteogenic cells or MSCs did not provide serum for initial — 2 or 3 h — adhesion (Llopis-Hernández et al. 2016; González-García et al. 2012; Rico et al. 2016). Culture following the serum-free adhesion phase included varying amounts of serum, and differences in osteogenesis and cell adhesion between the different artificial ECMs were observed, including increased OCN expression on PEA with FN and BMP2 compared to PEA and PMA (Llopis-Hernández et al. 2016), increased OCN and OPN expression on PEA with FN compared to PMA (González-García et al. 2012), and increased FAK activation on PEA compared to PMA (Rico et al. 2016). This serum-free regime may enhance the effects of the artificial ECMs to bolster and/or consolidate the differentiation of MSCs in the above experiments.

After adhesion for 2 h in serum-free conditions and culture in 1 % serum BM for 21 d on the various substrates with doubled growth factor concentration, the expression of OPN and actin was visualised by immunofluorescence microscopy (Figure 2.19). Many cells appeared shrivelled with atypically shaped nuclei and negative actin staining on all artificial ECMs (Figure 2.19 (a)). Differential interference contrast microscopy visualised the contrast between the well spread and misshapen cells (Figure 2.19 (b)). Further, these misshapen cells

stained positively for OPN yet did not have well established actin cytoskeletons (Figure 2.19 (c)). These misshapen cells had possibly died because the culture conditions were insufficient.

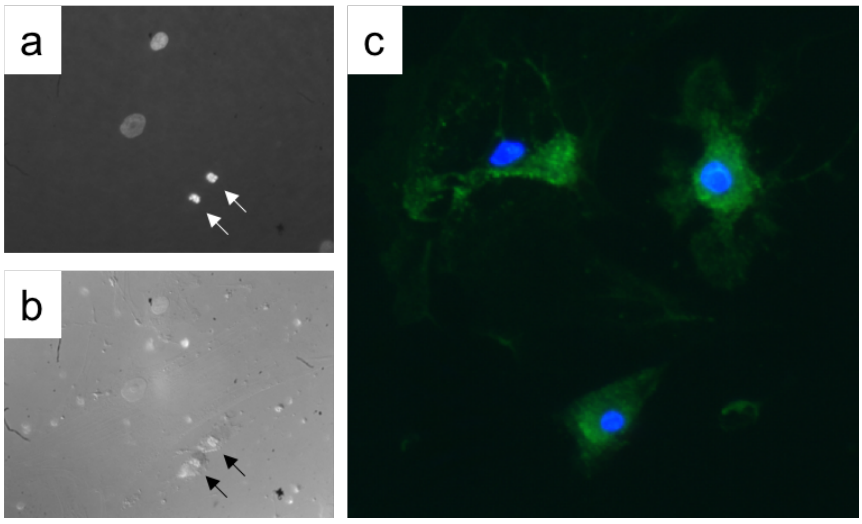

Figure 2.19 MSCs after 2 h serum-free adhesion followed by 1 % serum culture for 21 d Many MSCs had a misshapen morphology as determined by actin and nucleus staining, an indicator of cell death. These MSCs also appeared to express OPN, but this was likely to have been false positive staining. (a) DAPI image of normal and misshapen nuclei (arrows). (b) Matching differential interference contrast microscopy image to (a). Misshapen cells stained positively for OPN but did not appear to have well-formed actin cytoskeletons (c). Red, F-actin; blue, DAPI (nuclei); green, OPN. n = 2.

Since 1 % serum after serum-free adhesion appears to lower cell viability, the effect of a range of cell densities (2-6x10^3 cells cm^{-2}) and serum concentrations (1-10 %) was tested to find a compromise between serum concentration and MSC viability, focusing on glass or PEA with FN and BMP2 matrices (Figure 2.20). With increasing cell density and serum concentration, more cells appeared with an expected morphology. At higher densities, it was difficult to establish whether the effect was still present. At the lower density and serum concentration conditions, there are fewer misshapen and actin-negative cells on glass compared to the FN substrate. This is evidenced in more detail in Figure 2.21 of cells cultured in 1 % serum at 2x10^3 cells cm^{-2}.

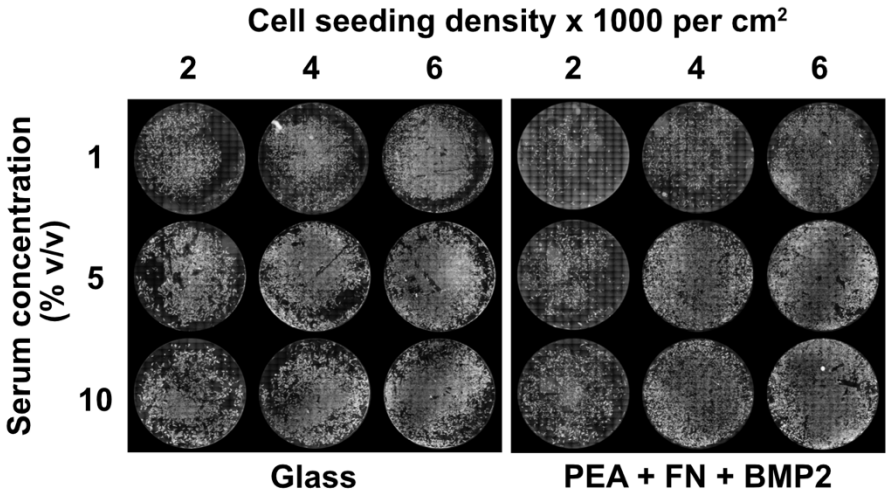

Figure 2.20 Serum concentration and cell density optimisation. MSCs were cultured for 21 d on glass or PEA with FN and BMP2. At lower serum concentration and cell density, fewer cells appeared well formed on PEA with FN and BMP2 in comparison to glass. Actin staining is shown to indicate the degree of adhesion and cell viability. n = 2.

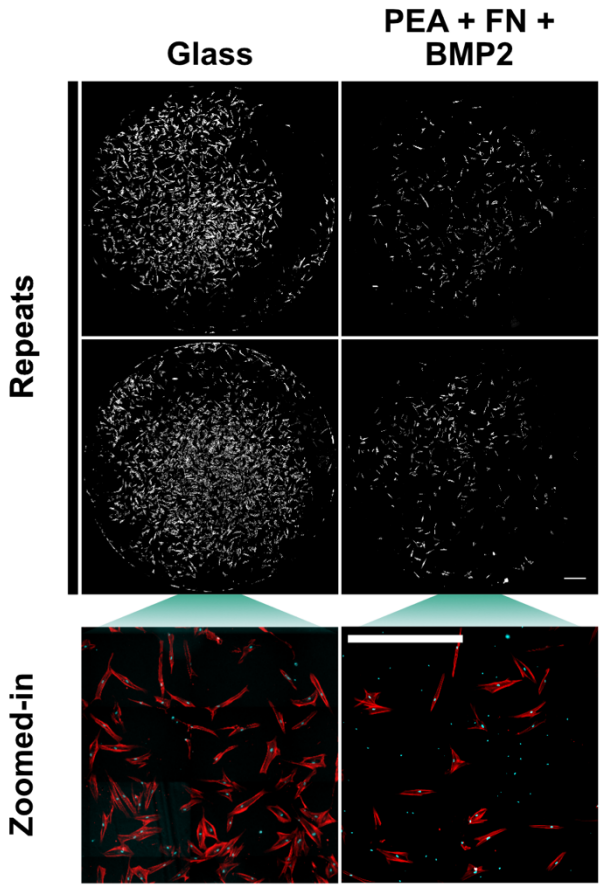

Figure 2.21 MSCs cultured for 21 d in 1 % serum at $2x10^3$ cells cm^{-2} on glass or PEA with FN and BMP2. Many more cells stained positively with actin on glass as evidenced by the zoomed in images: few nuclei were missing a cytoskeleton on glass, whereas many were on PEA. Many nuclei on PEA with FN and BMP2 appeared misshapen in comparison to those on glass. Whole sample scale bar 1 mm. Zoomed-in region scale bar 1 mm. Red, F-actin; blue, nuclei.

To confirm cell death, a live/dead cell stain was performed. After 30 min in culture in low serum conditions — 2 or 5 % serum — no cell death was observed (Figure 2.22). More cell death was observed after 2 h in serum-free conditions than in 5 % serum (Figure 2.23).

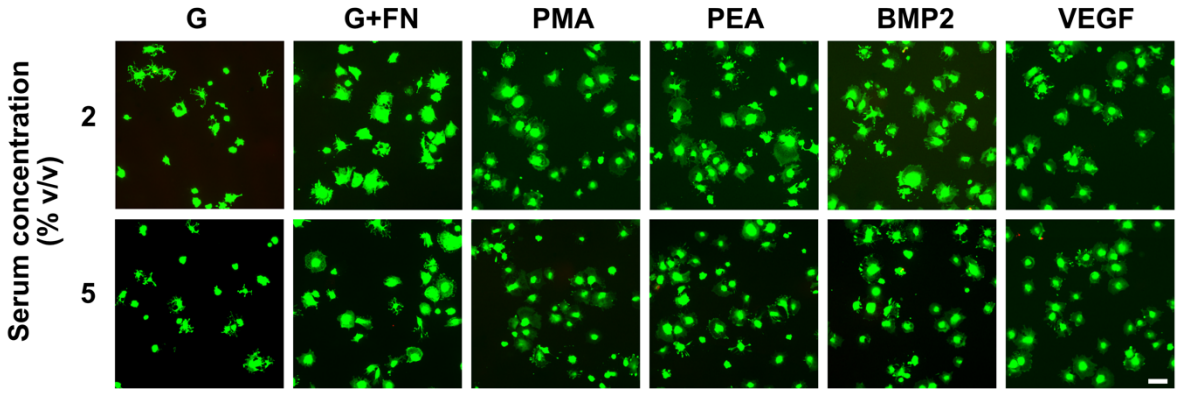

Figure 2.22 Live/dead stain of MSCs cultured in 2 or 5 % serum for 30 min. Cell death, indicated by ethidium homodimer staining (red) was undetected in either serum concentration and substrate. Green, calcein AM, indicated live cells. Scale bar 100 µm. n = 2.

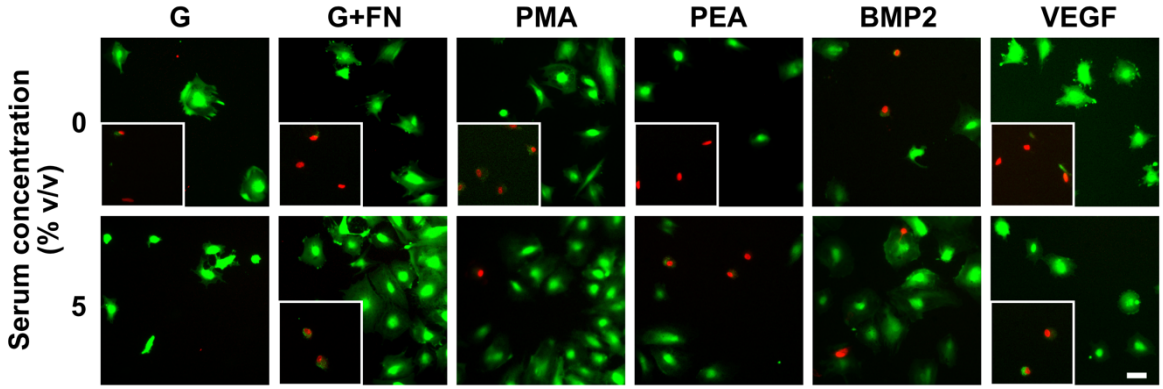

Figure 2.23 Live/dead stain of MSCs cultured for 2 h in 0 or 5 % serum. Dead cells were detected on all the substrates, and there appeared to be more in the 0 % serum condition compared to the 5 % serum condition. Inset images highlight dead cells. Red, ethidium homodimer; green, calcein AM. Scale bar 100 µm. n = 2.

Following the observations that some cells were viable at low and 0 % serum conditions, and that previous experiments have used these conditions (Llopis-Hernández et al. 2016; González-García et al. 2012; Rico et al. 2016), MSCs were cultured and their osteogenic and stemness protein expression measured as in section 2.4.3 with varying culture conditions. Figure 2.24 shows phase contrast microscopy images of MSCs cultured for 21 d in 5 % serum after 2 h of serum-free adhesion on the artificial ECMs. These conditions were chosen as the compromise between the number of healthy cells and low serum concentration

(Figure 2.20). On glass and glass with FN, cells had usual morphology and density. Many small and misshapen cells were seen on the artificial ECMs, with few normal MSCs. A positive control, MSCs cultured in osteogenic media with no serum-free step, demonstrated that these cells survived under regular serum concentrations. Figure 2.25 presents the ICW measurement of stemness and osteogenic proteins, and the estimated differences between matrices are in Figure 2.26. Expression of VCAM1 was higher on PEA with FN and VEGF compared to glass, but no other comparisons showed differences. Both PEA with FN and with BMP2 cells had higher average expression of OCN compared to glass controls, and PMA with FN had a higher expression than glass with FN. Finally, MSCs cultured in 2 % serum for 21 d were compared, but no differences were observed within the single donor that was tested (Figure 2.27).

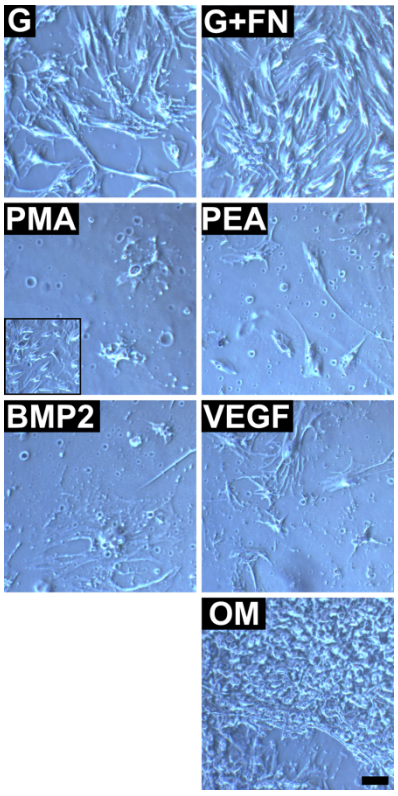

Figure 2.24 Phase contrast microscopy of MSCs cultured for 21 d in 5 % serum following 2 h serum-free adhesion. MSCs cultured on the glass controls appeared to have proliferated and spread to a greater extent than the polymer-based substrates. Many MSCs had misshapen morphology on the polymer-based substrates. MSCs cultured in OM for the entire culture with no serum-free adhesion stage. PMA inset: another region with denser cells. Scale bar 100 µm. n = 4 (n = 2 OM).

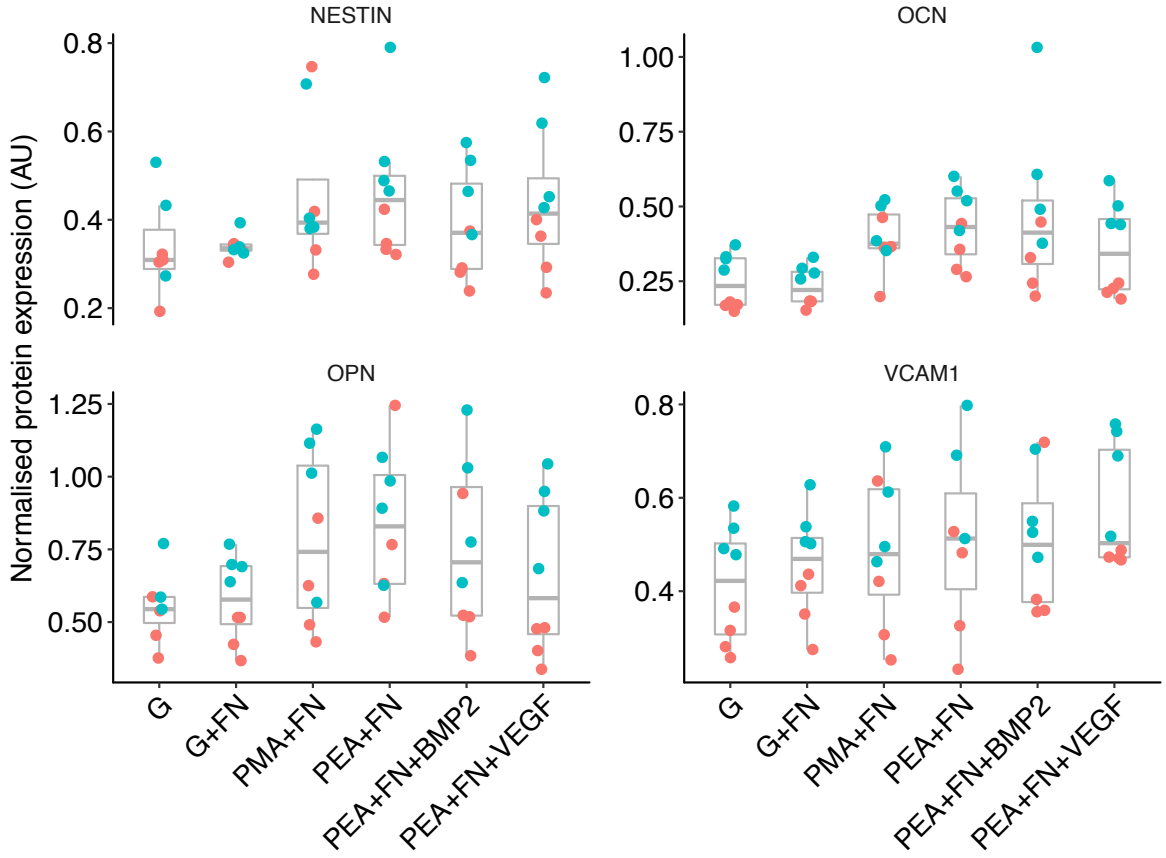

Figure 2.25 ICW analysis of stemness and osteogenic protein expression normalised to GAPDH in MSCs cultured for 21 d in 5 % serum following 2 h serum-free adhesion. AU, arbitrary units. n = 2 biological replicates (donors) distinguished by colour; n = 3 or 4 repeats per donor.

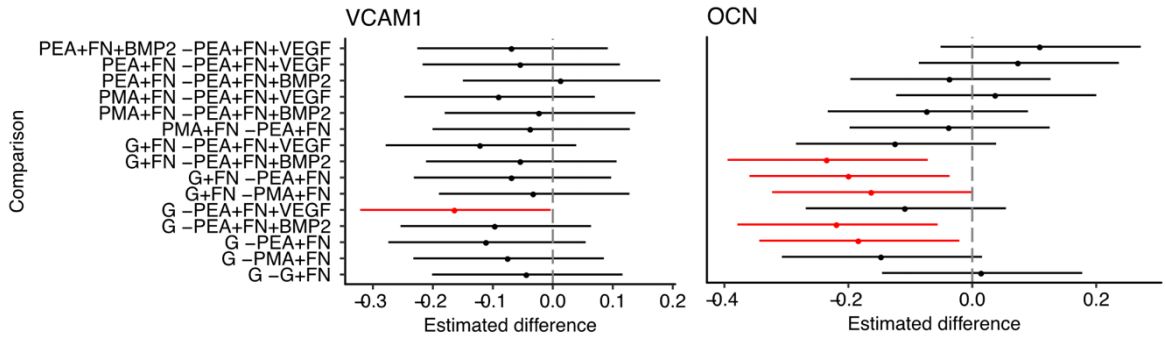

Figure 2.26 Comparison of differences between protein expression on the artificial ECMs (mean ± 95 % confidence interval) from Figure 2.25. VCAM1 expression was higher in MSCs cultured on PEA with FN and VEGF in comparison to glass. OCN expression was lower in MSCs cultured on glass and glass with FN in comparison to several of the polymer-based ECMs. Estimates are calculated from a mixed effects linear model that controlled for within-donor variation. Red, $p < 0.05$.

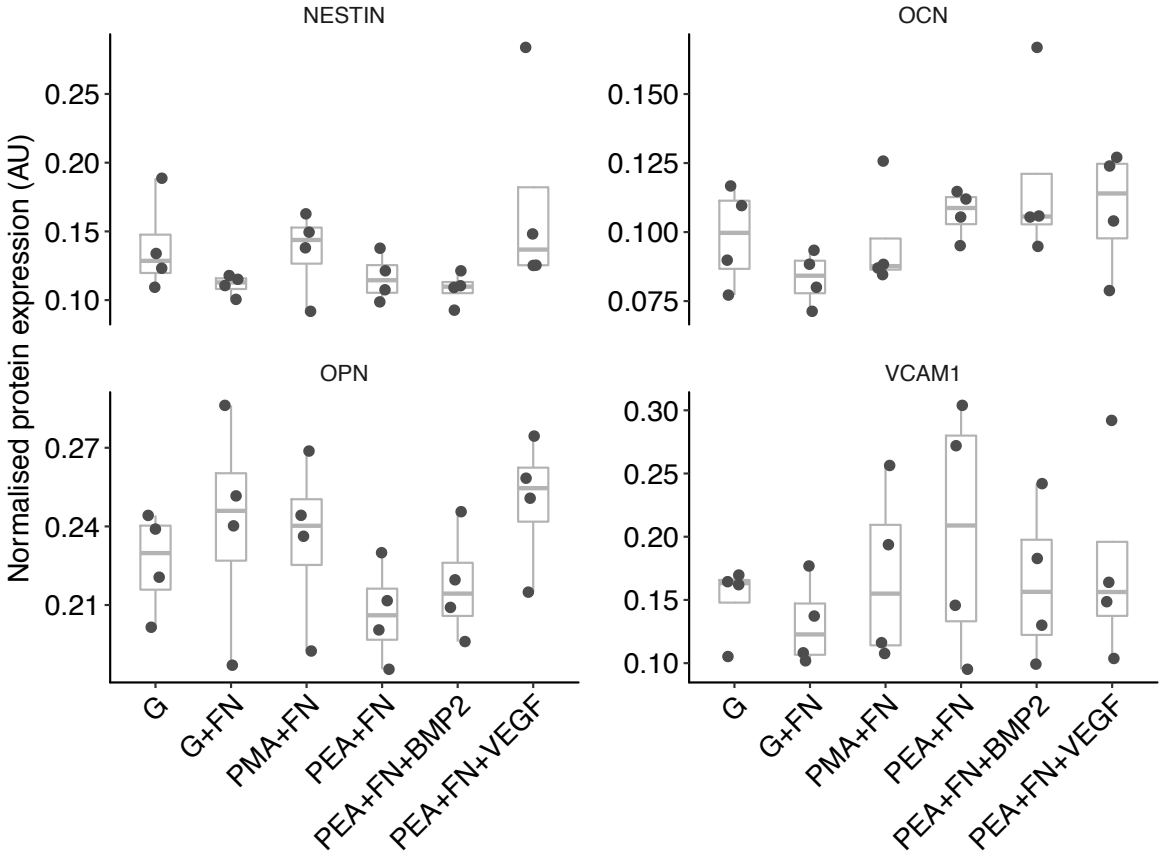

Figure 2.27 Osteogenic and stemness protein expression normalised to GAPDH by MSCs cultured in 2 % serum for 21 d. AU, arbitrary units. n = 4.

Although differences in the expression of some relevant proteins are observed using these low serum conditions, and previous published data also shows changes in osteogenesis (Llopis-Hernández et al. 2016; González-García et al. 2012; Rico et al. 2016), only the full 10 % serum condition was taken forward. This is because the same artefact of staining dead cells by immunofluorescence microscopy (by the protocols used here) (Figure 2.19) may occur using ICW, and these results were inferred from such an assay, confounding any conclusions. Nevertheless, the outcomes of experiments using 10 % serum culture conditions indicated that the substrates were still able to influence MSC behaviour to a degree.

2.4.5 Whole Biochemical Analysis

The measurement of individual protein expression provided some insight into the state of MSCs on the substrates. In contrast, a broader and coarser understanding of the MSC phenotype can be obtained using IR spectroscopy.

MSCs were cultured for 21 d on the various artificial ECMs and their biochemical phenotypes measured using SR-FTIR (Figure 2.28). The previous analysis of cell-free substrates revealed that there was variation in the polymer-defining bands (Figure 2.5). These overlapped some important biochemical wavenumbers including those for lipids (3000-2900 cm^{-1}) and in the fingerprint region (< 1450 cm^{-1}). To compare cellular spectra independent components analysis (ICA) was used to estimate and separate the background polymer signal from that of the cells (Figure 2.29) (Rutledge & Jouan-Rimbaud Bouveresse 2013). The estimated polymer signal contained some biological features, suggesting incomplete separation, but many important ones were retained in the biological signal, for example the amide region (1700-1500 cm^{-1}).

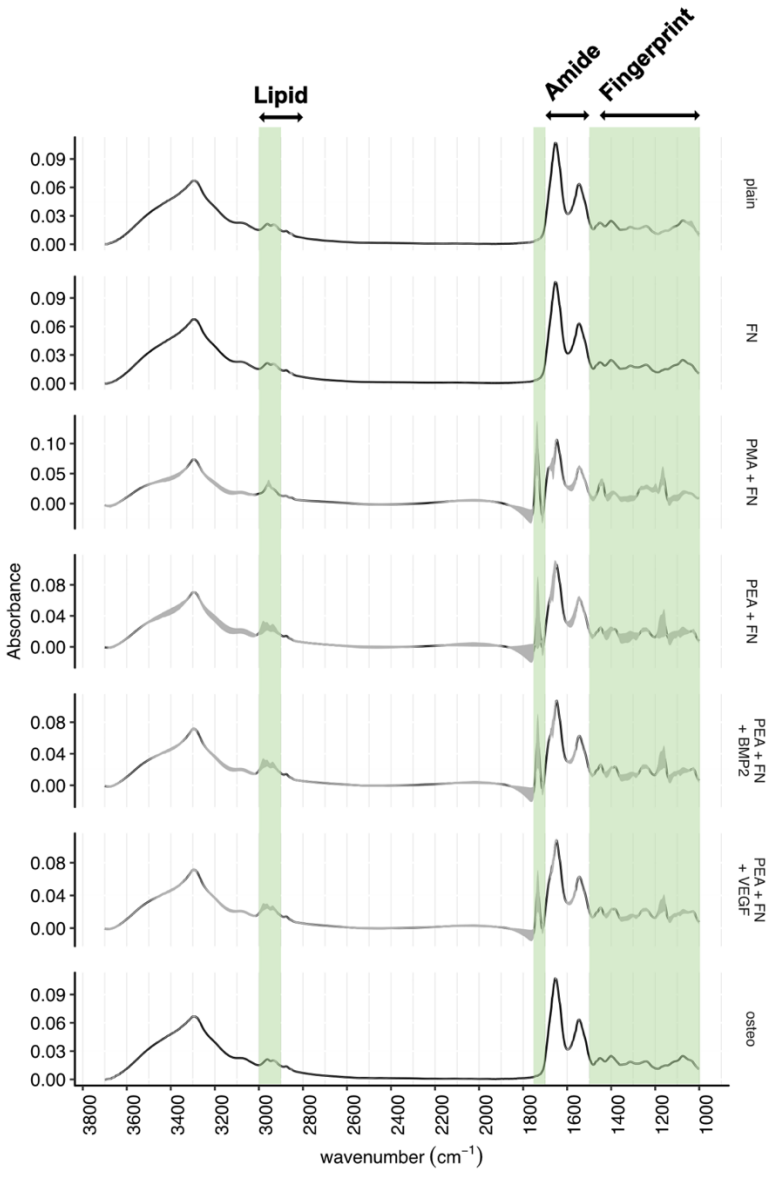

Figure 2.28 SR-FTIR absorbance of MSCs cultured on the artificial ECMs for 21 days. Mean (black line) ± standard deviation (grey shadow) of multiple spectra (n = 2 or 3 biological

replicates, summing to n = 1416 spectra). Green bars indicate the notable wavenumbers of the corresponding blank polymers. Note that there was overlap between the wavenumbers associated with biochemical features (black arrows) and the blank polymer wavenumbers, for example, in the lipid and fingerprint regions. The amide region did not appear to be obstructed by the background polymer signal.

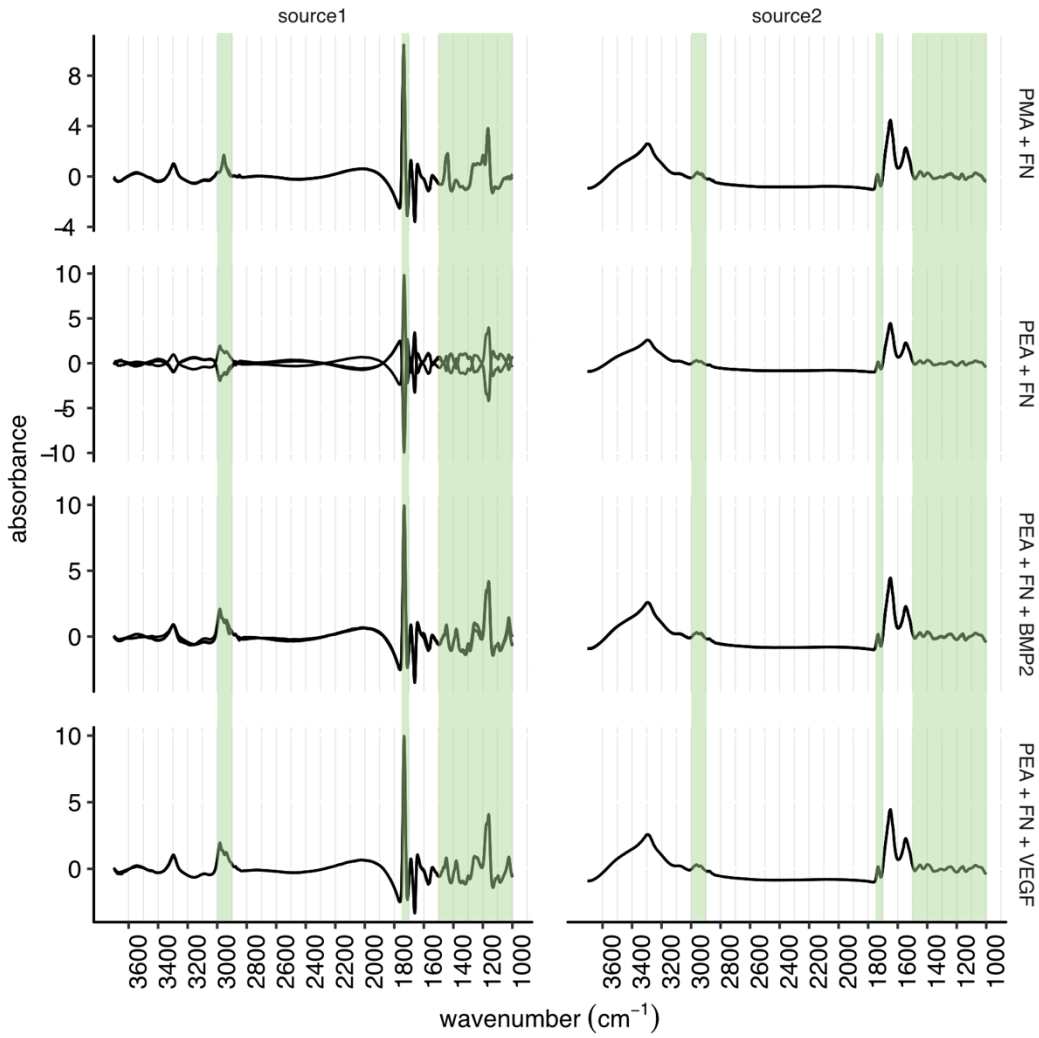

Figure 2.29 ICA separation of SR-FTIR measurements of MSCs cultured on artificial ECMs into two sources, tentatively the polymer signal (source1) and cellular biochemical signal (source2). The green bars highlight the wavenumbers associated with the blank polymers. Note that source1 contains signal within these regions, indicating that the blank polymer signal has been at least moderately isolated from source2, that is mostly cellular signal. from the Each black line in the subplot is a replicate. n = 2 or 3.

PCA analysis of the resultant biological signals showed separation between PMA and PEA cultures (Figure 2.30). Visualisation of the loadings of the first principal component, while ignoring the main blank polymer bands (3000-2900 cm^{-1}; 1400-1100 cm^{-1}; 1800-1700 cm^{-1}) — in case of incomplete signal separation by ICA — revealed several points of interest, including the amide I and II regions, the lipid region (~2900 cm^{-1}), and others including nucleic acids, lipids, and proteins (1700-1300 cm^{-1}). The SR-FTIR data showed that there was a broad difference in

the biochemical phenotype of MSCs on the FN network and isolated conformations.

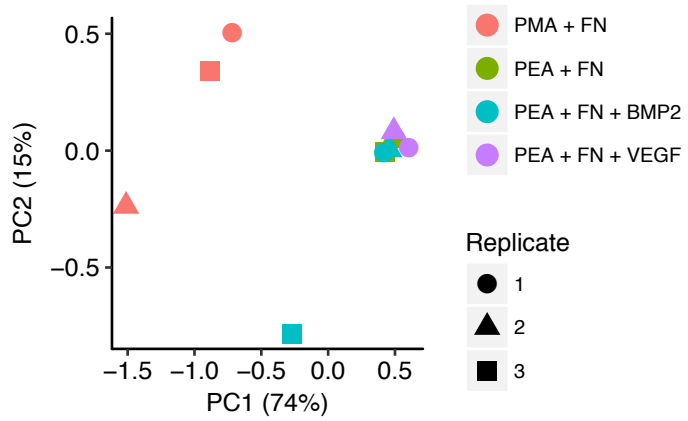

Figure 2.30 PCA of biological SR-FTIR signals after ICA from Figure 2.29. Note separation between FN conformations (PMA + FN and PEA + FN), whereas there is clustering between PEA + FN and the growth factor substrate variants.

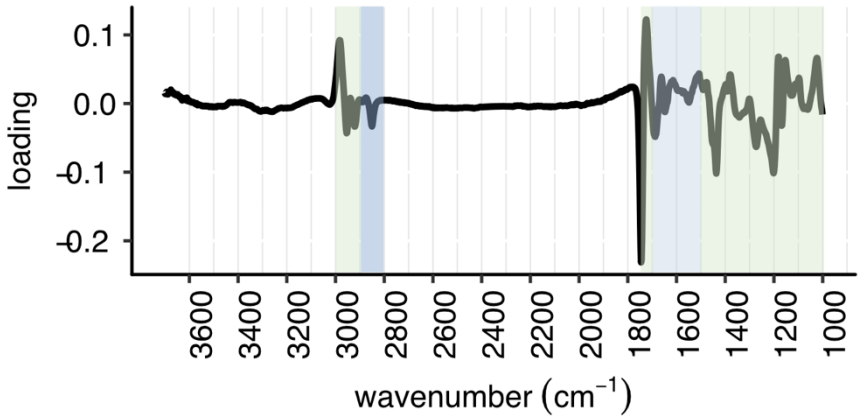

Figure 2.31 Loadings of PC1 from Figure 2.30. Loadings with large magnitudes highlighted by green bars were possibly due to background polymer signal. Loadings denoted by the blue bars were considered more reliable indicators of wavenumber differences between conditions.

2.5 Discussion

This chapter described the characterisation of a series of artificial ECMs, defined by their pattern of adsorbed fibronectin — network (PEA) or isolated (PMA) — and addition of growth factors BMP2 or VEGF, and the culture of primary human bone marrow STRO-1$^+$ MSCs on these. Whereas previous work has studied these cells on the network and isolated conformations (González-García et al. 2012), the characterisation with the addition of BMP2 and VEGF is novel.

Substrate characterisation shows that the artificial ECMs were reproducibly made from published methods (Moulisová et al. 2017; Llopis-Hernández et al. 2016; González-García et al. 2012; Guerra et al. 2010), and that CaF_2 was a viable substrate for producing these biomaterials, opening up the range of possible characterisation techniques to include IR analysis (Baker et al. 2014). IR analysis is a common approach to characterising protein conformation at the biomaterial surface (Glassford et al. 2013), and while no conclusions about FN conformation could be drawn here — likely because the thickness (~ 1µm (Vanterpool et al. 2014)) of the polymer overwhelms the nanometer high FN signal — FTIR can be used to compare adsorbed FN conformation across substrates (Baujard-Lamotte et al. 2008). Future FTIR work could use thinner films. Nanoscale chemical imaging such as near-field IR spectroscopy (Donaldson et al. 2016; Cinque et al. 2016) or nanoscale secondary ionisation mass spectroscopy (Paine et al. 2017) could also be used to explore the binding of growth factors to the FN network and compare FN conformations.

Many cell types have been studied on subsets of these artificial ECMs before including mouse C3H10T1/2 MSCs (Rico et al. 2016), mouse MC3T3 osteoblasts (Guerra et al. 2010), human osteoblasts (Rico et al. 2009), C2C12 myoblasts (Mnatsakanyan et al. 2015; Salmerón-Sánchez et al. 2011), primary human STRO-1^+ bone marrow MSCs (González-García et al. 2012), commercially-sourced bone marrow MSCs (Llopis-Hernández et al. 2016), NIH3T3 fibroblasts (Llopis-Hernández et al. 2013), L929 fibroblasts (Vanterpool et al. 2014), and HUVECs (Moulisová et al. 2017). This work here added Saos-2 cells to the list. The study of primary human STRO-1^+ bone marrow MSCs over the full repertoire of published artificial ECMs is a relatively important contribution. It fits with the dogma of tissue engineering — using cells direct from a donor (Lanza et al. 2013) — avoids proprietary preparations, and meant the cells were at passage 0 on arrival. Furthermore, these points mean data are as relevant to *in vivo* behaviour as possible, without implantation. Primary human STRO-1^+ bone marrow MSCs have been used in the study of other osteogenic and multipotency-maintaining biomaterials (McNamara et al. 2011; Tsimbouri et al. 2014; Dalby et al. 2007; González-García et al. 2012; McMurray et al. 2011).

Over all surfaces in 10 % serum, MSCs adhered, proliferated, and developed actin and ß-tubulin cytoskeleton components and focal adhesions. OPN expression was higher on the network FN compared to the isolated organisation, the opposite to the previous study of these cells that used serum-free conditions and different quantification methods (González-García et al. 2012). NESTIN expression was lower on the artificial ECMs compared to control glass suggesting that differentiation is occurring on the artificial ECMs, but whether NESTIN truly represents an MSC-only population is contested (Xie et al. 2015). SR-FTIR analysis has been used to profile MSCs undergoing osteogenesis (Krafft et al. 2007), chondrogenesis (differentiation to cartilage lineage cells, namely chondrocytes) (Chonanant et al. 2011), and adipogenesis (differentiation to fat lineage cells, namely adipocytes) (Liu et al. 2016). Analysis here suggested that there are differences in lipid, protein, and/or nucleic acid composition between MSCs cultured on the networks and those on the isolated organisation. Together, these results suggested differences in degree of osteogenesis between MSCs cultured on the PMA and PEA-based biomaterials. The addition of growth factors at the given concentrations did not lead to discernible differences, and this may be specific to primary human bone marrow STRO-1^{+} MSCs.

Altering culture conditions to match previous work (González-García et al. 2012; Llopis-Hernández et al. 2016) lead to notable MSC death, which made whole-surface measurements such as ICW unreliable as cell debris and false positive immunofluorescence staining could not be ignored. Nevertheless, when cultured in 5 % serum after 2 h serum-free adhesion, MSC cultured on FN networks with VEGF have a greater expression of VCAM1, and OCN expression is greater on the networks with/without BMP2 and on the isolated network compared to controls. More patients need to be measured for inferences to be drawn about 2 % serum conditions without serum-free adhesion.

3 High-Content Automated Microscopy Analysis of MSCs on the Artificial ECMs

3.1 Introduction

In the previous chapter, MSCs were characterised for their growth, viability, biochemistry, and differentiation on the artificial ECMs. Results suggested that the FN conformation affected the degree of osteogenesis and general protein, lipid, and nucleic acid content. Representative cells from whole-sample mosaic images of nuclei and actin and tubulin networks at various time-points were used as evidence.

Hand-picking individual representative cells from images is a time-consuming process that is susceptible to bias, and there may be more information to gain by looking at the whole population of cells (Figure 3.1) (Caicedo et al. 2017). Another problem with this approach is that there may be seemingly imperceptible differences, due to subtlety and in obscure patterns of image features (Ljosa & Carpenter 2009). These issues can be offset by using a biological image analysis suite, for example CellProfiler (Carpenter et al. 2006; Kamentsky et al 2011) or ImageJ/Fiji (Schneider et al. 2012; Schindelin et al. 2012; Schindelin et al. 2015), which provides an interface to automate and objectify comparisons between cells and measure more subtle features. These tools are convenient for researchers without programming experience but greater flexibility and performance can come from using a scripting language, for example, Python and its scikit-learn package (van der Walt et al. 2014) or MATLAB and its Image Processing Toolbox (Gonzalez et al. 2009; Gonzalez & Woods 2010b; Schindelin et al. 2012; Carpenter et al. 2006), both of which have been used as part of CellProfiler (Carpenter et al. 2006; Kamentsky et al. 2011). Furthermore, a script forms an explicitly recorded and automated approach meaning that methods are fully reproducible and tractable.

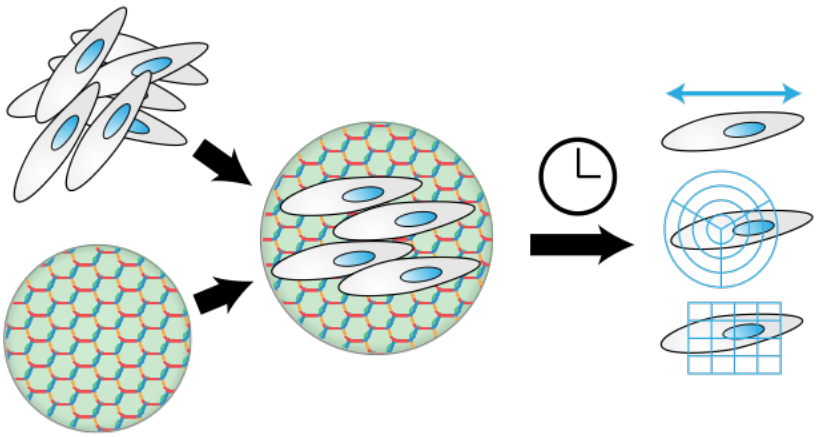

Figure 3.1 Quantification of cell morphology over whole samples. Cells are seeded on the artificial ECMs for a given length of time, after which a whole-sample image is taken and the morphology of each cell is measured and compared to other conditions.

In this chapter, an image processing pipeline was developed in MATLAB and used to profile and then compare the morphology of cells across the artificial ECMs. These were used to gain more insight from the previous chapter's image data and analyse some newer datasets of pRunx2, an osteogenic transcription factor (Stein et al. 2004), and ALP, an enzyme involved in bone matrix mineralisation (Sharma et al. 2014; Pittenger et al. 1999), expression in MSCs. It will be shown that, generally, cells on PMA with FN or PEA with FN and VEGF had the most idiosyncratic morphometric features.

The aims of this chapter were to: (1) develop an image processing pipeline to detect and measure cell morphometrics from whole-sample multichannel automated fluorescence images, and (2) use it to compare nucleus, cytoskeletal, and lineage-specific protein morphometrics between MSCs cultured on the artificial ECMs.

3.2 Methods

3.2.1 Tables of Reagents

Table 3.1 Immunofluorescence Microscopy Reagents (in addition to those used in chapter 2)

Item (and dilution)	Source
Anti-phosphoS465-Runx2 (1:100)	Abgent
Anti-ALP (1:100)	Santa Cruz Biotechnology

3.2.2 MG63 Cell Culture

MG63 were cultured for 2 h on glass using the same media without serum as Saos-2 cells in the previous chapter. MG63 were used instead of Saos-2 cells because they were available at the time of the experiment.

3.2.3 Immunofluorescence Microscopy

As chapter 2. Montages of segmentations were made using ImageMagick.

3.2.4 Image Processing Pipeline

The image pipeline was written in MATLAB 2016a using the Image Processing Toolbox and Parallel Computing Toolbox. Since nuclei-only image analysis was performed on dense 21 d samples an extra pre-filter step was used: any segmented region with an eccentricity > 0.9 and an area > 2000 or < 275 pixels was removed.

3.2.5 Data Analysis

The image labeller was written in Python using tkinter and pillow, and data analysis was performed in R using ggplot2 (Wickham 2016), plyr (Wickham 2011), dplyr, tidyr, caret (Kuhn 2008), nnet (Venables & Ripley 2002), multcomp (Hothorn et al. 2008), lme4 (Bates et al. 2015), lmTest, MASS (Venables & Ripley 2002), and glht (Venables & Ripley 2002). For experiments where all artificial ECMs had replicates a mixed effects linear model using sample replicate as the random effect (using R formula syntax: image feature ~ artificial ECM + (1 | sample)) was used to estimate differences. Logistic regressions for predicting the artificial ECM were run on scaled and centred data using 10-fold cross validation repeated 5 times with down-sampling during resampling. Type III analysis of variance tables were used to compare the predictor importance in fitted logistic regression models. The per-class average performance metrics are averages of the respective metric over all cross validations and repeats.

3.3 Results

3.3.1 Developing the Image Processing and Data Analysis Pipeline

To measure each cell, they were first segmented from their images using the image processing pipeline (Figure 3.2). The implemented approach, which relies on a registered nuclei and actin image, used marker-based watershed segmentation (Meyer & Beucher 1990) with nuclei as internal markers, a model-based segmentation approach (Caicedo et al. 2017) that is one of the more common cell segmentation methods (Meijering 2012) and is generally similar to that used by CellProfiler (Carpenter et al. 2006; Kamentsky et al. 2011). Nuclei were detected from a DAPI (Chazotte 2011; Kapuscinski 1995) channel image using several steps: a Gaussian filter ($\sigma = 5$) that removed small noise elements and smoothed the image; a white top-hat transform using a disk-shaped structuring element with a radius of 20 px, which removed large noise elements (e.g. contamination by fibres during staining) and evened illumination; Otsu thresholding (Otsu 1979), which calculated a threshold level that was used to select nuclei pixels; a closing operation, which filled small holes; and border clearing, which removed any nuclei on the image borders that may have been incomplete (Gonzalez & Woods 2010a; Gonzalez et al. 2009) (Figure 3.3 left).

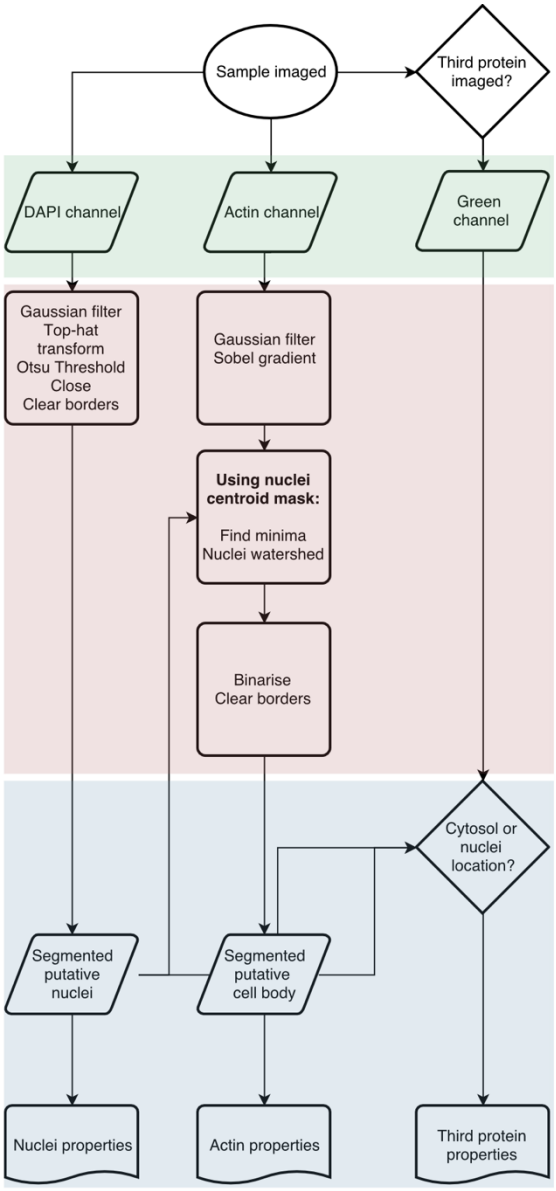

Figure 3.2 Flowchart of the image segmentation pipeline. The green box highlights the input fluorescence microscopy images; red, the segmentation process; blue, the measurement of the optional third channel and production of the final data.

Cell bodies were detected following nuclei segmentation using the actin (tetramethylrhodamine (TRITC)) channel. Images were smoothed using a Gaussian filter ($\sigma = 3$) to obfuscate neighbouring actin fibres so that a subsequent Sobel gradient filter detected the very edges of the cells. Minima corresponding to points halfway between neighbouring nuclei centroids — computed by a watershed distance transform — and the nuclei centroids themselves were then imposed on the gradient image, known as minima imposition, and watershed segmentation of the resultant images was used to segment the cell bodies. Finally, borders were cleared to remove any

incomplete cell segmentations (Gonzalez et al. 2009; Gonzalez & Woods 2010a; Soille 2013) (Figure 3.3 right).

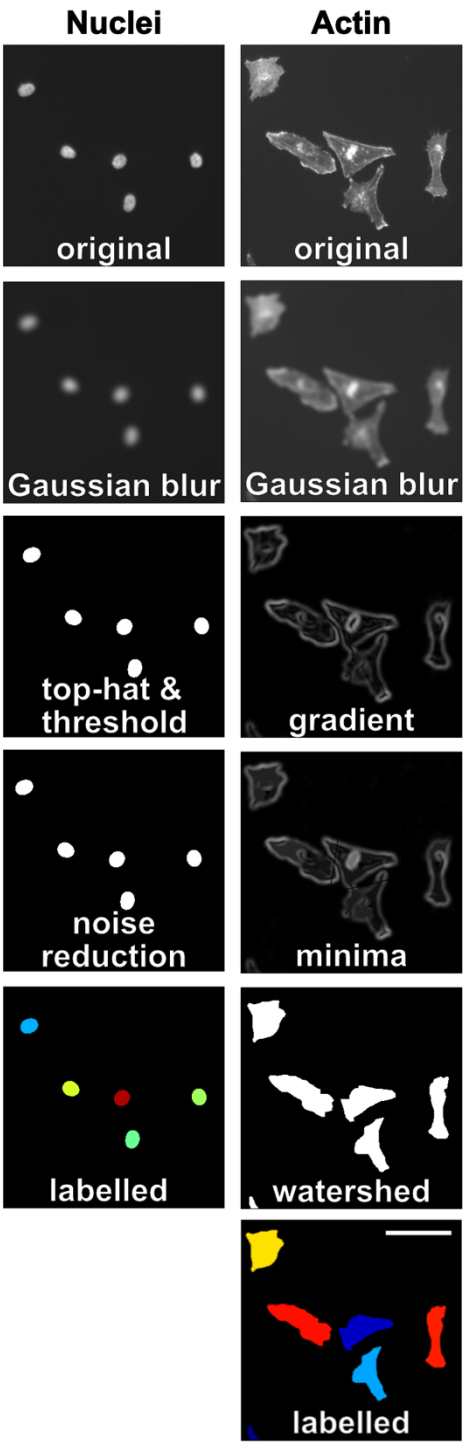

Figure 3.3 Nuclei and actin image processing steps, in order from top-to-bottom, using an example image. Left, DAPI image processing to segment nuclei: the noise reduction step represents close and border clearing operations. Right, TRITC image processing to segment cell bodies/actin cytoskeletons: the minima step represents minima imposition using information about nuclei locations. The colours of the segmentations in the labelled images were arbitrary. Scale bar 50 µm.

The MATLAB image processing pipeline was developed to allow different combinations of channels and regions to be analysed. At minimum, a blue DAPI nuclei image must be provided, but the red TRITC actin and a third green fluorescein isothiocyanate (FITC) channel are optional. With all combinations, measurements for each of the entities — nuclei, actin cytoskeleton, third green channel — and a combined RGB image of the segmented entities are returned by the pipeline. The third channel is treated differently to the other two. It undergoes no image processing per se, but instead, measurements are taken from it in regions corresponding to the actin or nuclei segmented regions, according to user preference (Figure 3.2). For example, with an interest in osteogenesis, a third channel image of phospho-Runx2 (pRunx2) would be measured at locations coincident with nuclei segmentations since it is an osteogenic transcription factor (Stein et al. 2004). If an actin image is provided, a final step removes all but the first nucleus in an actin segmentation, if multiple were found, and any nucleus segmentation that is not within the boundary of an actin segmentation.

The shape descriptors: area, eccentricity, perimeter length, major axis length, minor axis length, orientation, maximum intensity, mean intensity, and minimum intensity; histogram-based texture descriptors: standard deviation, skew, roughness, uniformity, and entropy; and grey-level co-occurrence matrix descriptors for horizontally adjacent pixels: contrast, correlation, energy, and homogeneity (Gonzalez et al. 2009; Gonzalez & Woods 2010a; Haralick et al. 1973) are calculated for the actin and nuclei segments. For the third channel, only the histogram-based texture and grey-level co-occurrence matrix descriptors are calculated. Recording many diverse features is a common approach in cell profiling (Caicedo et al. 2017).

Figure 3.4 shows the result of cell segmentation using the image processing pipeline on a test image of MG63 cells stained for actin and nuclei. MG63 cells, like Saos-2 cells in the previous chapter, are used as a model for osteogenesis, grow as an adherent monolayer, and their culture uses the same methods as Saos-2 cells (Pautke et al. 2004). Most nuclei were segmented well but several actin cytoskeletons, particularly of closely neighbouring cells, were incompletely segmented. Further, some nuclei remained after removing their corresponding

actin segmentations that were touching the image border. The image processing pipeline was then applied to a dataset from the previous chapter, where MSCs were cultured for 24 h and stained for actin and nuclei, to evaluate a larger sample size closer to the problem domain. Figure 3.5 (left) shows a random sample of the resultant segmentations. Artefacts in the DAPI channel image from the staining process were segmented, as were some incomplete cells.

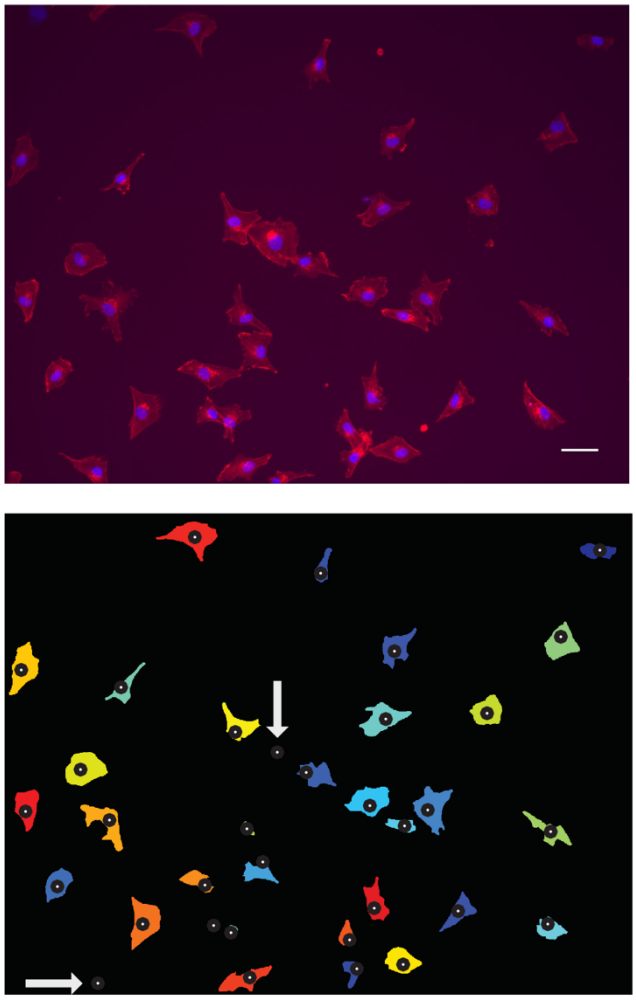

Figure 3.4 Segmentation of MG63 cultured on glass for 2 h. Top: original image; bottom: labelled segmentation image, where the circles (black outline, white centre) indicate the position of detected nuclei centroids and the variably coloured regions delimit detected cell boundaries. Note generally favourable segmentation with some incomplete segmentation of touching cells (vertical arrow), and some false positive nuclei identification (horizontal arrow). 10x objective. Top: Red, actin; blue, nuclei. Bottom: The colours of the segmentations are arbitrary. Scale bar 50 µm.

A cell-level quality control process (Caicedo et al. 2017) was developed to remove bad segmentations, which required a ground truth dataset. All 12,475 putative segmented cells were classified by hand as “keep” or “reject” — good or bad segmentation — in a decision to discard them before analysis (Figure 3.5).

Segmentations that were mostly incomplete, not in focus, and clearly not cellular material, such as contaminating fibres, were rejected. Note that some "good" segmentations were not particularly so due to mistakes during the manual classification. 8,132 segmentations were kept and 3,343 were rejected. A comparison of the annotated good and bad segmentations highlighted several distinguishing features, particularly actin perimeter, area, and intensity standard deviation (Figure 3.6), suggesting that bad segmentations tended to be smaller with a more homogeneous actin stain. Several features did not appear to have a normal distribution, so they were log-transformed to have a more normal distribution. These were the nucleus and actin grey level co-occurrence matrix contrast, pixel intensity histogram uniformity, aspect ratio, area, and perimeter; actin eccentricity, and grey level co-occurrence matrix energy; and nucleus pixel intensity roughness and mean. Features that would be expected to contain negative values, such as orientation (the angle between the major axis and the x-axis of the image, ranging from -90 to 90 degrees) or correlation were exempt from log-transformation. If transformation created infinite or null values for a particular segmentation it was removed from further analysis. The reported number of segmentations produced for each subsequently described set of images will be the number remaining after those with invalid feature quantifications have been removed.

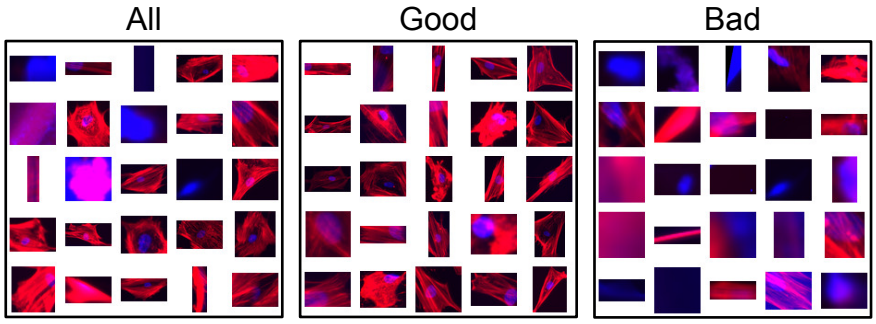

Figure 3.5 Generation of a ground truth dataset for developing a segmentation quality control process. Left: A random selection of segmentations produced by the image processing pipeline from images of MSCs cultured for 24 h on the artificial ECMs, showing segmented cells and non-cell matter. These were manually annotated as good (centre) or bad (right) segmentations. While some bad segmentations were mistakenly labelled as good, for example, centre panel 4th column 4th row, the majority of bad segmentations are genuinely so. Red, actin; blue, nuclei.

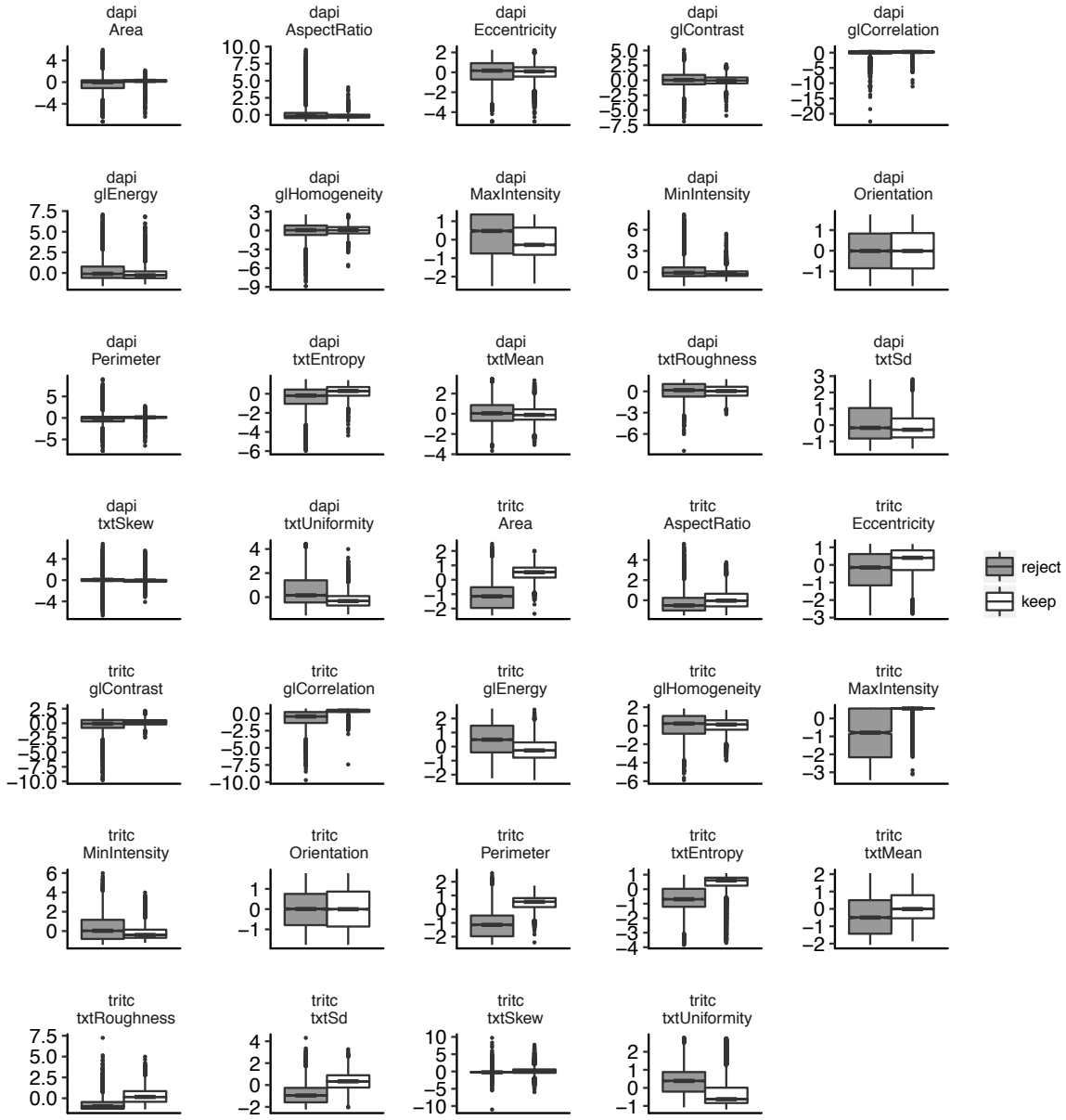

Figure 3.6 Distributions of segmentation features by quality classification. Note that some features were log10-transformed. Actin perimeter (tritc Perimeter), area (tritc Area), and intensity histogram standard deviation (tritc txtSd), were particularly different between classes. tritc = actin, dapi = nuclei, "txt" prefix = intensity histogram feature, "gl" prefix = grey-level co-occurrence matrix feature, reject = bad segmentation, keep = good segmentation. 8132 keep segmentations, 3343 reject segmentations.

Given that it was time-consuming and error-prone to classify each segmentation by hand, a logistic regression classifier, was tested for its utility as a semi-automatic filter using the annotated segmentations. A classifier learns from a set of training data, a subset of the total dataset, about the relationship between a class (which can be referred to as an outcome or a category) and features (otherwise known as predictors), so that it may predict the class of future test samples based only on the measured features. With respect to this specific task, the class was whether the segmentation was good or bad, and the features were

the morphometrics and textures of the segmentation. A logistic regression is the relationship, or function, between segmentation quality class and pixel features. Over several test subsets the logistic regression was accurate in predicting good and bad segmentations (Figure 3.7 (a)), with the probability of most predictions being near 1 or 0 (Figure 3.7 (b)). False positive segmentations, those incorrectly predicted as good, were usually predicted as such for all repeats (Figure 3.7 (c)), and these tended to be damaged or rounded cells (Figure 3.7 (d)). False negative segmentations (Figure 3.7 (e)) were predicted similarly and were usually incomplete segmentations (Figure 3.7 (f)). The distribution of actin area, one of the more discriminating features, for segmentations that had both true positive and false negative classifications over the test repeats is shown in Figure 3.7 (g), showing overlap with the other events. A random sample of these revealed a mixture of larger incomplete segmentations and non-cellular fluorescence (Figure 3.7 (h)). These results suggested that a logistic regression was adequate for filtering out most poor segmentations prior to analysis.

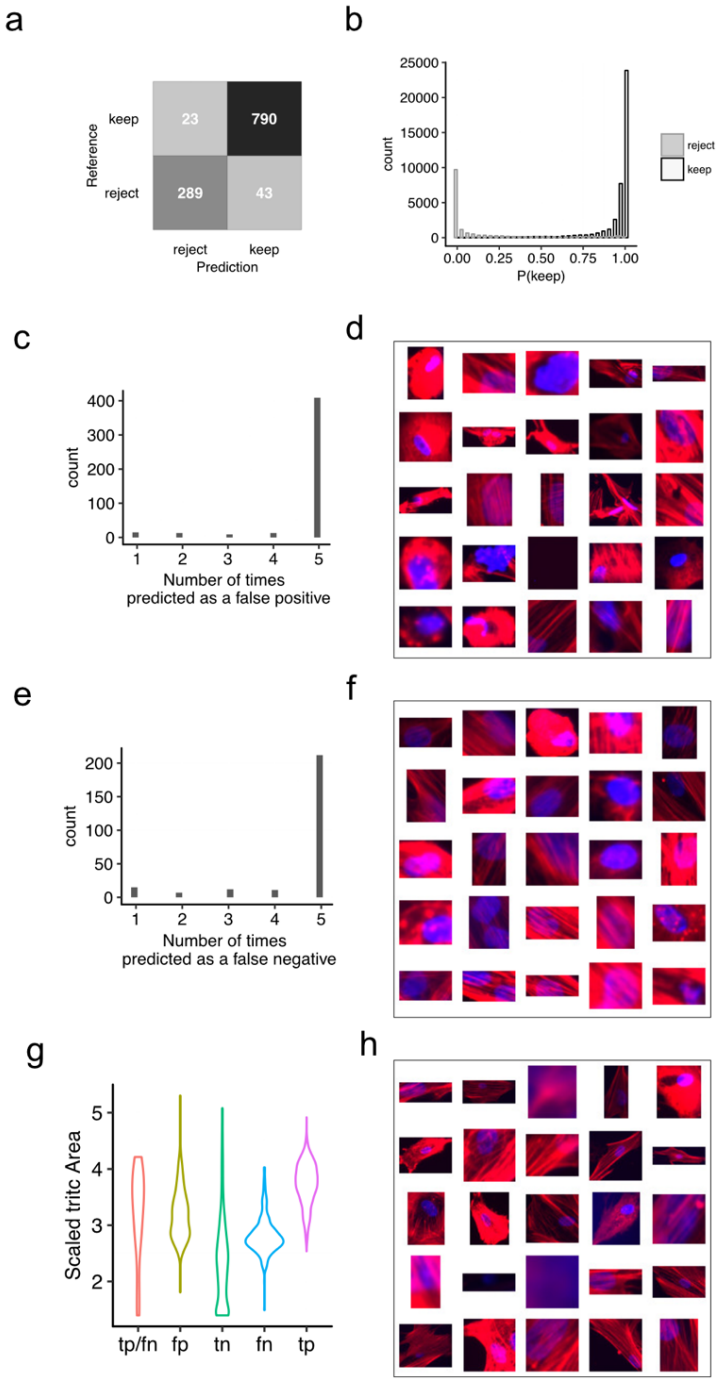

Figure 3.7 Segmentation quality control performance. (a) Average confusion matrix of predicted and reference classes over 5 repeats of testing/training. The majority of segmentations were correctly classified. (b) Distribution of the probabilities of keep classification, the positive class, (P(keep)) by reference class. The probability for most good quality segmentations was high and vice versa for bad segmentations. (c) Distribution of the number of times a false positive image was classified as such. Most false positive segmentations were always predicted as false positives over the 5 repeats. (d) Random sample of false positive segmentations, which tended to be damaged or rounded. (e) Distribution of the number of times a false negative image was classified as such. Most false negative segmentations were always predicted as false negatives over the 5 repeats. (f) Random sample of false negative segmentations, which were usually incomplete segmentations. (g) Distribution of actin area by classification type. Actin area is mostly distinct between prediction classes. (h) Random sample of segmentations that were classified as false positive or true negative over the repeats, which tended to be large incomplete segmentations or non-cellular fluorescence. tp = true positive, fp = false positive, tn = true negative, fn = false negative, "keep" = positive class.

The optimum sample size for fitting the logistic regression was determined by fitting the model on progressively larger portions of a training set from 1-99%, and testing on a constant test set (Figure 3.8 (a)). Learning curve plots in Figure 3.8 (b) suggested this was about 10 %. The final data curation and analysis pipeline is shown in Figure 3.9 (left), with manual curation of 10 % of the total dataset needed to filter the rest before analysis. A graphical Python application was written to expedite the curation step: this lets the user view each segmentation in a given list and classify each as “keep” or “reject” using keyboard shortcuts — classifying an image automatically moves to the next — and save the list of classifications and corresponding image names in a comma-delimited file (Figure 3.9 (right)). This image processing and data curation pipeline was used to generate the datasets analysed in the rest of this chapter.

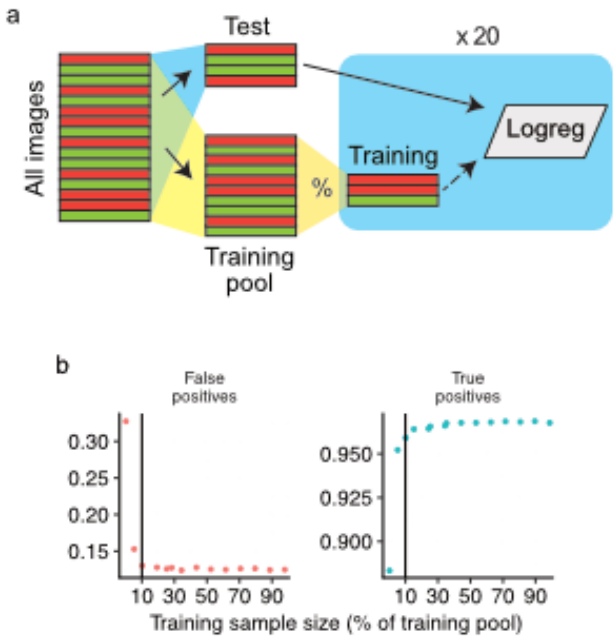

Figure 3.8 Optimal sample size for the quality control process. (a) Schematic of optimisation approach. A logistic regression model was fitted on a different sized random sample of training data and then tested on a constant dataset, twenty times. Red/green colouring represents segmentation quality. The blue region indicates repeated model fitting/testing. (b) The average true positive and false positive rates of the fitted model for each training sample size, showing that approximately 10 % of the total dataset is an optimal size for training. Logreg = logistic regression model fitting.

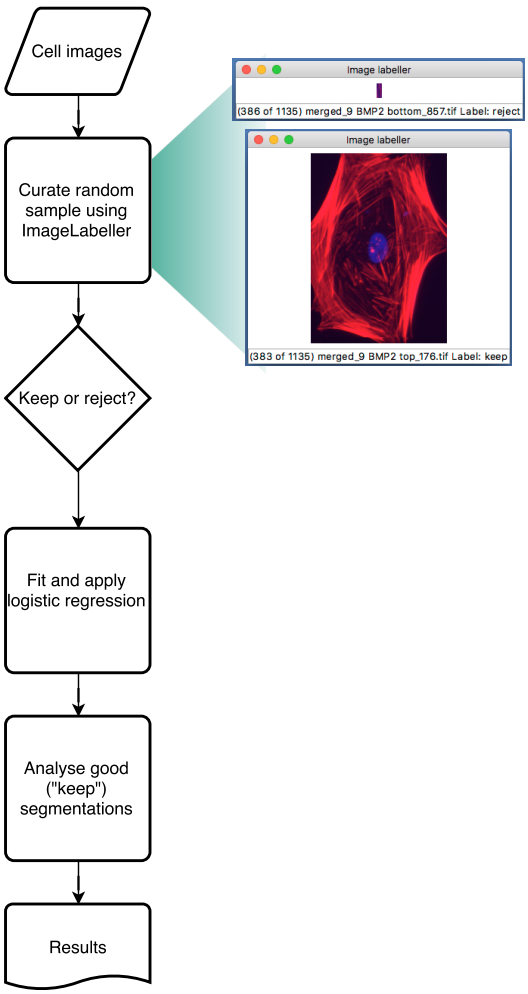

Figure 3.9 Data curation and analysis pipeline. (left) From top to bottom, of the segmentations from the image processing pipeline, 10 % are manually annotated to be kept or discarded using the image labeller (right). The labeller displays each segmentation in turn to be classified by the user, and shown is an example of a labelled bad (right, top) and a labelled good segmentation (right, bottom). The logistic regression is then fit to the curated sample and used to filter out bad segmentations from the rest of the unlabelled dataset before analysis.

3.3.2 Image Analysis of MSCs Cultured on the Artificial ECMs

3.3.2.1 Analysis of Actin and Nuclei After 24 h Culture

In the previous chapter, MSCs were cultured for 24 h and stained for actin and nuclei, which demonstrated adhesion, spreading, and the formation of actin cytoskeletons. Here, the image processing (Figure 3.2) and data analysis (Figure 3.9) pipeline was used to analyse, in more detail, the whole images from which these conclusions were drawn. A sample of the resulting good quality segmentations is shown in Figure 3.10 (a), which were evenly distributed across the different artificial ECMs and replicates suggesting that initial seeding density was similar (Figure 3.10 (b)). The results of a logistic regression used to predict

which substrate an MSC was cultured on, using all of the measured features, revealed (Figure 3.10 (c)) that the majority of cells cultured on the control substrates were distinguishable from those cultured on the polymer-based ECMs, and that those on PMA with FN, and to a lesser extent, those on PEA with FN and VEGF were also distinct. MSCs cultured on PEA with FN and with BMP2 were largely misclassified. Cells on the control substrates were also misclassified as one another more often than the other classes. A PCA plot revealed little clustering of conditions (Figure 3.10 (d)). The average accuracy, precision, recall, and F-1 score (Caicedo et al. 2017) were all approximately 50 %, with glass and particularly PMA with FN predictions faring relatively better (Table 3.2). The distribution of probabilities that each segmentation belonged to each class is shown in Figure 3.10 (e): correct glass, glass with FN, and PMA predictions tended to have higher probabilities; correct predictions for the other ECMs were less certain. The logistic regression coefficients of the top 10 most informative predictors are shown in Figure 3.10 (f), which suggested that nuclei-based features were more informative for discrimination. MSCs that were cultured on control substrates and correctly predicted appeared to have brighter actin staining (Figure 3.11).

Classification of the MSC images provided insight into how the actin and nucleus features varied by substrate (Figure 3.10 (f)) (Shmueli 2010). In contrast to classification, an inference task is used to infer the value of a property about a population from a sample of that population. Here, inference was used to compare the image features between the artificial ECMs. A mixed effects linear model was used to test for differences while adjusting for sample-specific variation (Gelman & Hill 2007; Krzywinski et al. 2014). Three actin-related features were significantly different in at least one of the non-control comparisons (Figure 3.12). MSCs cultured on PMA with FN had a greater aspect ratio, eccentricity, and perimeter than those on PEA with FN and with BMP2, and those on PEA with FN and VEGF had a greater perimeter than MSCs on PEA with FN and with BMP2.

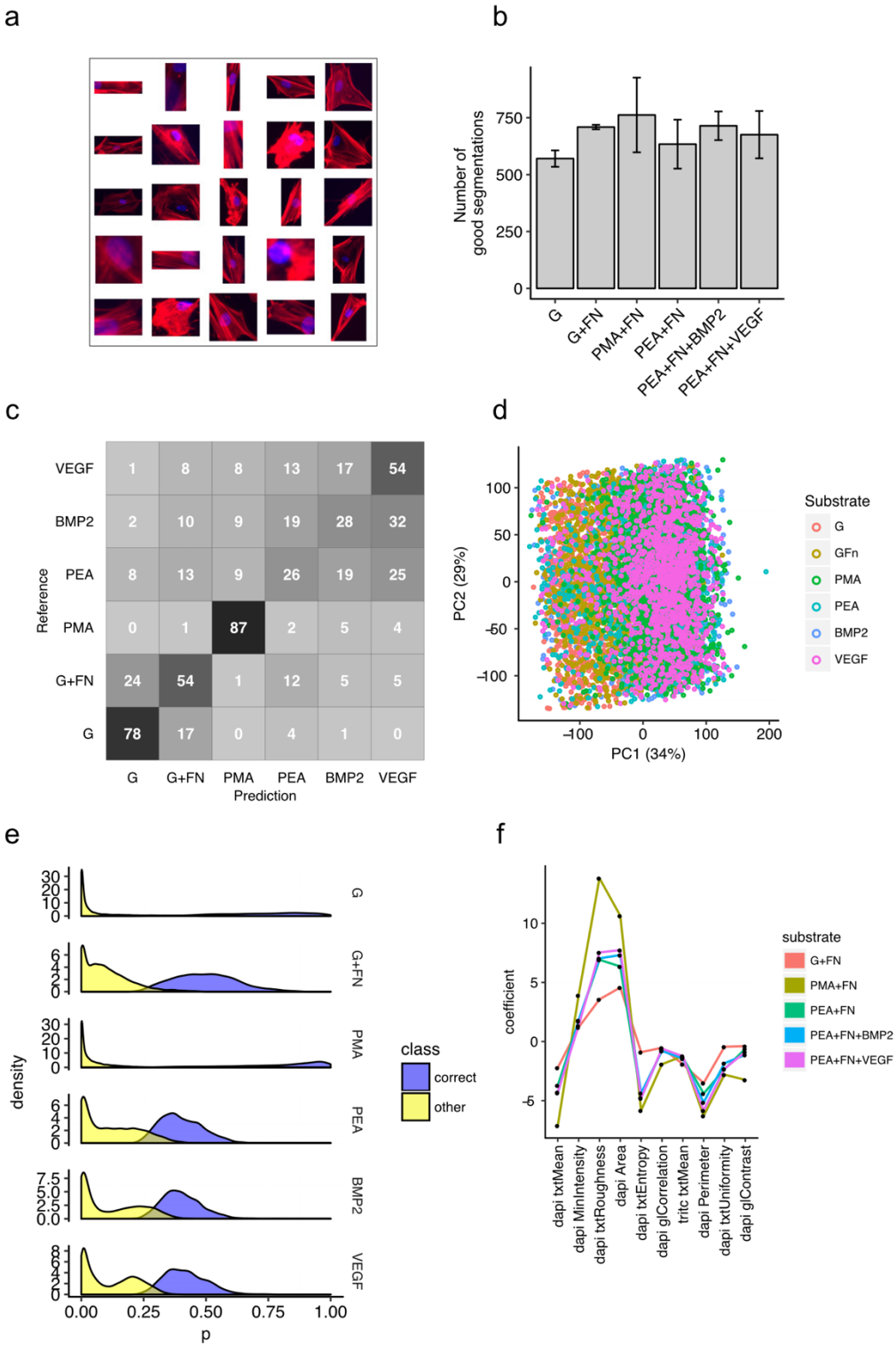

Figure 3.10 Results of a logistic regression used to predict which ECM a given MSC was cultured on after 24 h based on actin and nuclei image analysis. (a) Random sample of good segmentations. (b) Mean ± standard deviation of number of good segmentations (8,132 good segmentations in total, n = 2 sample replicates), showing small variations across replicates. (c) Confusion matrix showing average percentage of classifications for each substrate, showing high accuracy for MSCs cultured on PMA and VEGF compared to the other polymer-based ECMs. (d) PCA of all features. (e) Logistic regression probability of belonging to each class by correct or incorrect prediction, with predictions for MSCs cultured on glass and PMA having a high probability. (f) Top 10 most significant coefficients of the logistic regression fit for each artificial ECM, of which most are related to nuclei. tritc

= actin, dapi = nuclei, "txt" prefix = intensity histogram feature, "gl" prefix = grey-level co-occurrence matrix feature.

Table 3.2 Per-class performance for prediction of substrates at 24 h

	G	G+Fn	PMA+FN	PEA+FN	PEA+FN+BMP2	PEA+FN+VEGF	Mean
Precision	0.78	0.54	0.87	0.26	0.28	0.54	0.54
Recall	0.64	0.55	0.78	0.32	0.39	0.44	0.52
f-1 score	0.70	0.54	0.83	0.29	0.32	0.49	0.53

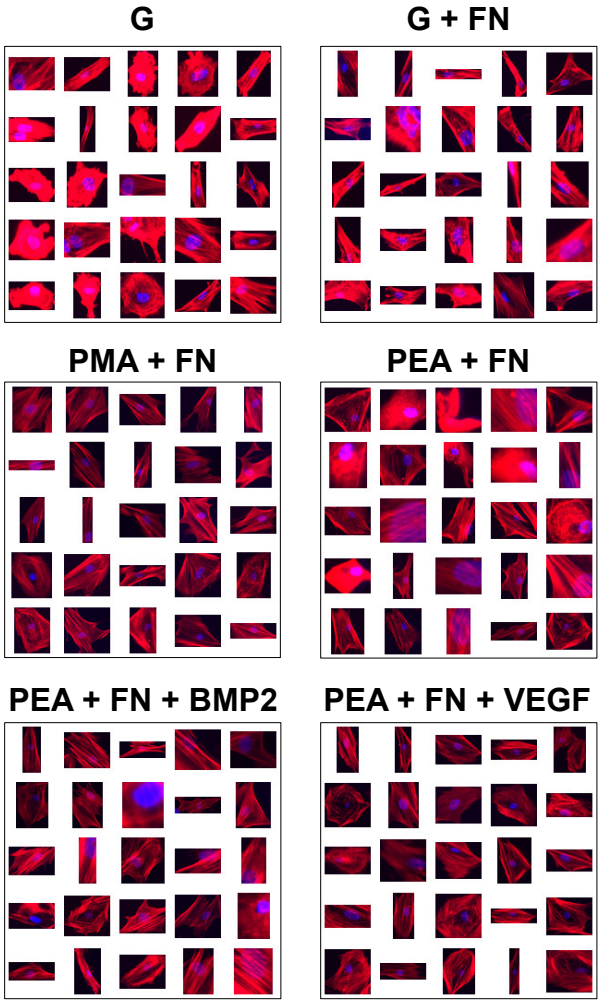

Figure 3.11 Random samples of correctly predicted MSCs after 24 h culture.

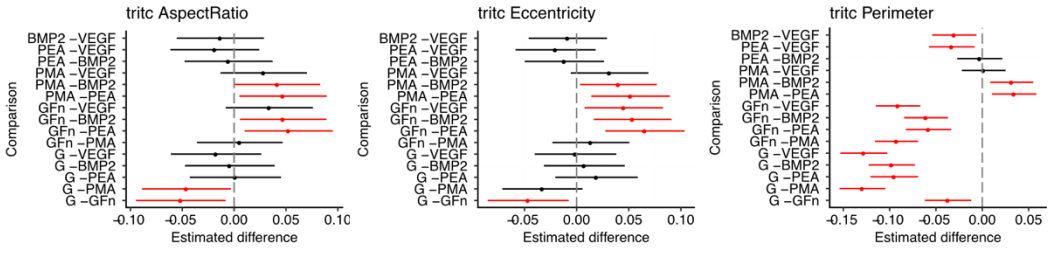

Figure 3.12 Features of MSCs cultured for 24 h on the artificial ECMs with significantly different estimated means ± 95 % confidence interval. MSCs on PMA had a greater aspect

ratio, eccentricity, and perimeter than those on PEA and BMP2. MSCs on VEGF had a greater perimeter than those on BMP2. Red, significant. tritc = actin.

The prediction of substrates using a logistic regression model was used to analyse the rest of the chapter's image datasets. Those with at least 2 replicates for each substrate were also analysed using the mixed effects model to control for replicate variance.

3.3.2.2 Analysis of Early Osteogenesis After 5 d Culture

pRunx2 is the activated form of an osteogenic transcription factor, localised to the nucleus in early osteogenesis (Stein et al. 2004; Vimalraj et al. 2015; Komori 2010). Images of MSCs cultured for 5 d on the artificial ECMs stained for pRunx2, actin, and nuclei were analysed as above. The inclusion of pRunx2 in the third (FITC/green) channel increased the number of features analysed. 58,636 segmentations were determined from the images, and 38,388 images were classified as good segmentations. pRunx2 staining occurred in nuclei (Figure 3.13 (a)). Fewer cells were observed on control substrates but seeding density was similar across replicates (Figure 3.13 (b)). Similar to the prediction performance for MSCs at 24 h (Figure 3.10 (c)), accuracy was high for cells on PMA with FN, PEA with FN and VEGF, and control substrates (Figure 3.13 (c)), with a small drop in overall performance (Table 3.3). No clustering was observed using PCA (Figure 3.13 (d)). The distribution of prediction probabilities reflects this reduction in confidence (Figure 3.13 (e)). The most informative predictors included the maximum and standard deviation of pRunx2 staining intensity (Figure 3.13 (f)). Differences between correctly predicted segmentations were not readily discernible by eye besides the seemingly brighter actin cytoskeleton staining on control substrates (Figure 3.14).

14 of 57 features were estimated to be significantly different between non-control substrates, none of which were based on pRunx2 (Figure 3.15, summarised in

Table 3.4). Many of the differences were between the different FN conformations. MSCs on PMA with FN had a more homogeneous actin cytoskeleton texture (higher intensity histogram uniformity, and higher grey-level co-occurrence matrix energy and lower entropy) and lower intensity (mean, maximum, and minimum) compared to those on PEA with FN. They also had a smaller nuclei aspect ratio and greater minimum intensity. Further, the growth factor-based ECMs also had differences compared to PMA; PEA with FN and BMP2 had similar differences to the comparison with PEA; and while PEA with FN and VEGF did not have a lower mean and maximum actin intensity, they did have a greater nuclei intensity (mean, maximum, and minimum). Comparing between the two growth factors on PEA with FN, MSCs on BMP2 had a smaller area and perimeter. Finally, MSCs on PEA with FN and VEGF had a smaller nuclear aspect ratio than those on the FN network on PEA alone.

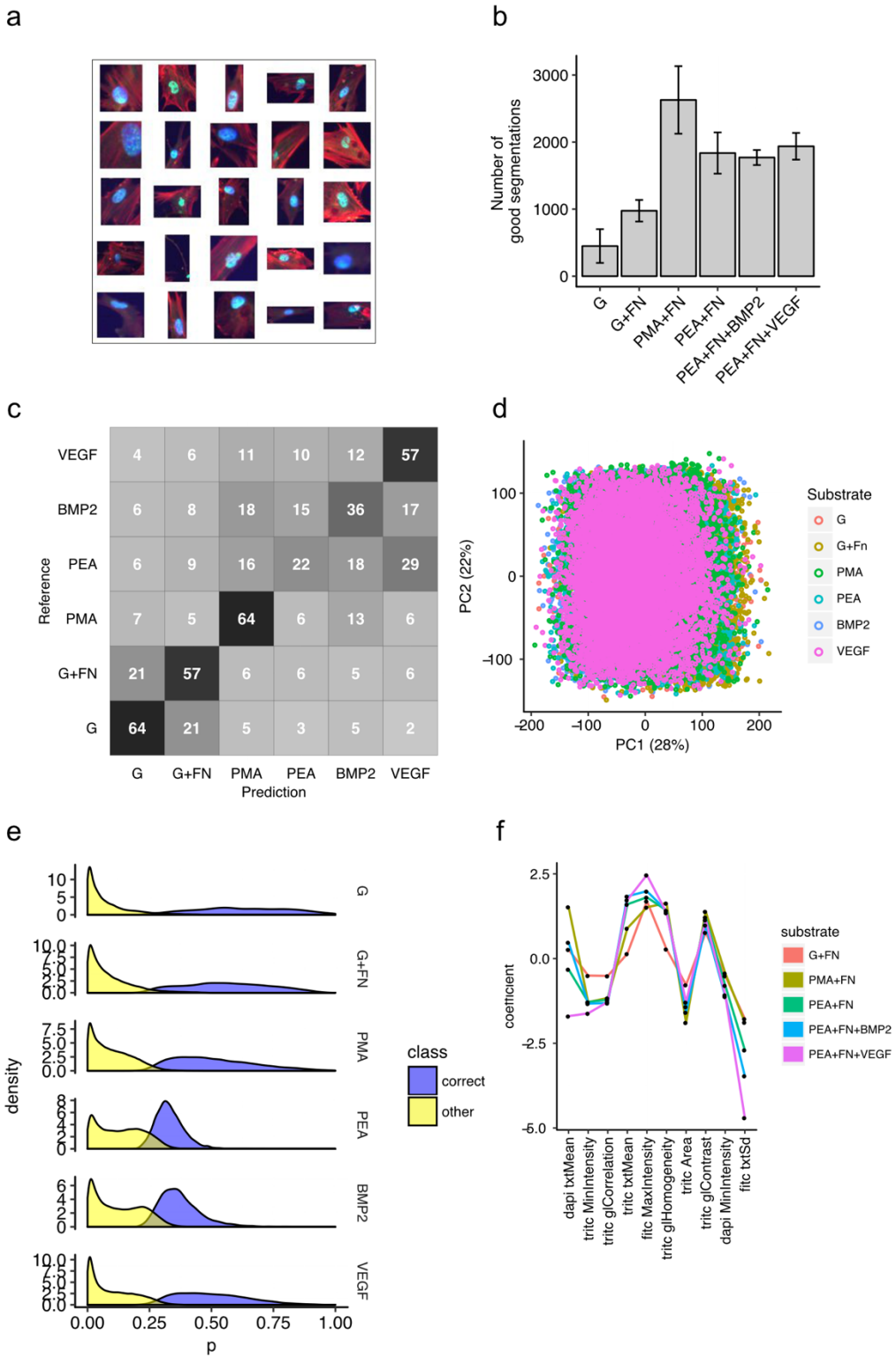

Figure 3.13 Results of a logistic regression used to predict which ECM a given MSC was cultured on after 5 d based on pRunx2, actin, and nuclei image analysis. (a) Random sample of good segmentations. (b) Mean ± standard deviation of number of segmentations (38,411 good segmentations, n = 4 sample replicates), showing similar numbers across replicates, but fewer segmentations on control substrates. (c) Confusion matrix showing average percentage of classifications for each substrate, with higher accuracy for MSCs cultured on control, PMA, and VEGF substrates. (d) PCA of all features. (e) Logistic regression probability of belonging to each class by correct or incorrect prediction. Predictions for MSCs on PEA and BMP2 had comparatively lower confidence, while the other classes were spread over a range of probabilities. (f) Top 10 most significant coefficients of the logistic regression fit for each ECM, which contains features from all three image channels. tritc =

actin, dapi = nuclei, fitc = pRunx2, "txt" prefix = intensity histogram feature, "gl" prefix = grey-level co-occurrence matrix feature.

Table 3.3 Per-class performance metrics for prediction of substrates at 5 d

	G	G+Fn	PMA+FN	PEA+FN	PEA+FN+BMP2	PEA+FN+VEGF	Mean
Precision	0.68	0.57	0.64	0.22	0.36	0.57	0.50
Recall	0.30	0.46	0.64	0.37	0.40	0.51	0.45
f-1 score	0.41	0.51	0.64	0.27	0.37	0.54	0.46

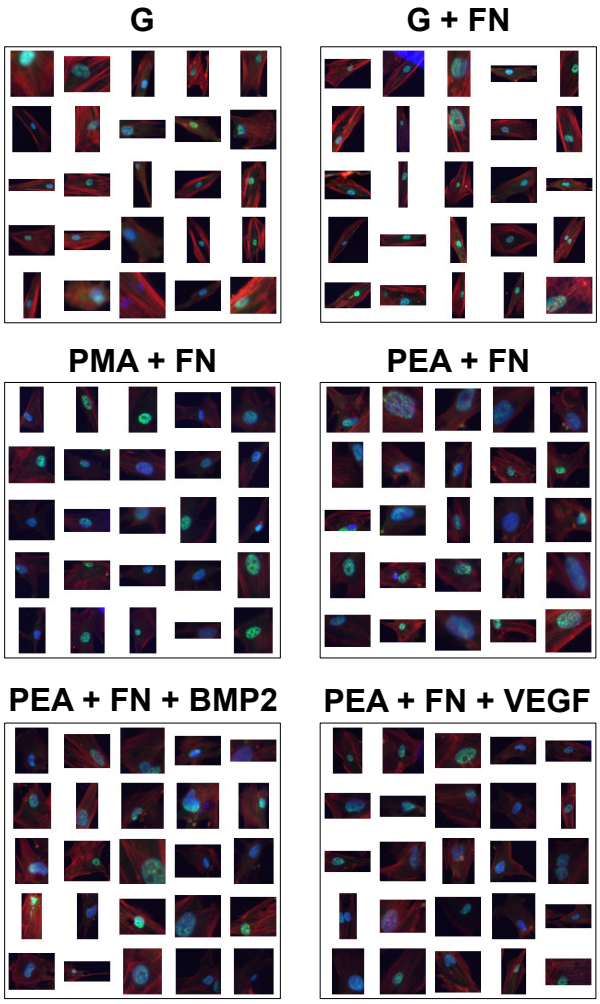

Figure 3.14 Random samples of correctly predicted MSCs after 5 d culture

Table 3.4 Summary of estimated differences from Figure 3.15. SD = standard deviation.

	BMP2-VEGF	PEA-VEGF	PEA-BMP2	PMA-VEGF	PMA-BMP2	PMA-PEA
Actin perimeter	↓					
Actin uniformity				↑	↑	↑
Actin entropy						↓
Actin energy				↑	↑	↑
Actin mean				↓	↓	↓
Actin max					↓	↓
Actin min					↓	↓
Actin area	↓					
Nucleus SD				↑		
Nucleus eccentricity				↓		
Nucleus aspect ratio		↓		↓	↓	↓
Nucleus mean				↑		
Nucleus max				↑		
Nucleus min				↑		↑

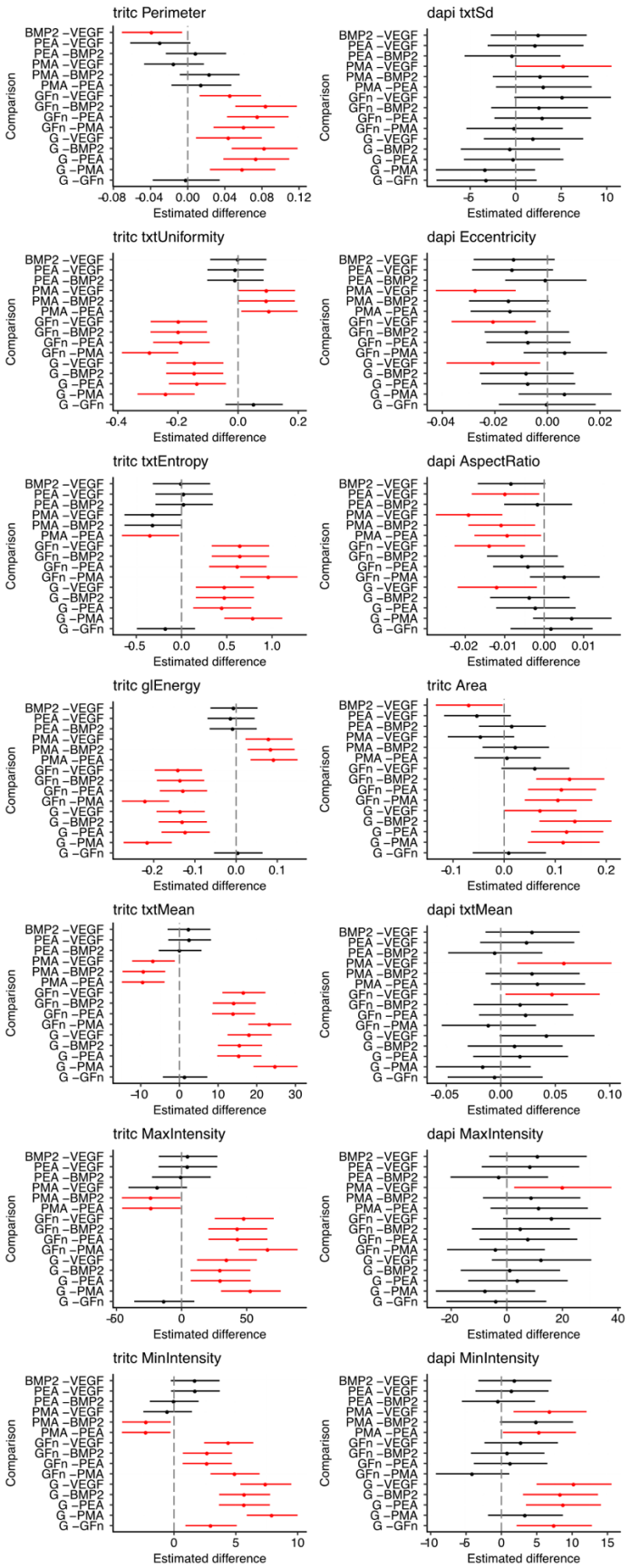

Figure 3.15 Features of MSCs cultured for 5 d on the artificial ECMs with significantly different estimated means ± 95 % confidence interval. MSCs on PMA tended to have differences to the PEA-based substrates in both actin and nuclei features. Red, significant (confidence interval does not include 0). tritc = actin, dapi = nucleus, "txt" prefix = intensity histogram feature, "gl" prefix = grey-level co-occurrence matrix feature.

3.3.2.3 Analysis of β-tubulin, Actin, and Nuclear Images After 3 d Culture

Figure 3.16 shows the results of the image analysis of ß-tubulin, actin, and nuclei of MSCs cultured for 3 d on the artificial ECMs. 13,263 segmentations were determined from the images, and 8,346 were predicted as good segmentations. Variation in seeding was greater in this experiment (Figure 3.16 (b)) partially because some samples partly surfaced during culture (the same problem occurred in the following ALP analysis, but it was corrected for all other experiments). Despite this, good quality segmentations were still obtained (Figure 3.16 (a)). In contrast to the previous analyses the correct prediction rate was roughly consistent across all substrates (Figure 3.16 (c), Table 3.5), which was supported by the distribution of prediction probabilities (Figure 3.16 (e)). No clustering of substrates was observed using PCA (Figure 3.16 (d)). Correct prediction rates for MSCs cultured on PEA with FN and VEGF appeared higher, but this condition only had one replicate (Table 3.5). The top features contained a mix of those for nuclei, actin, and ß-tubulin (Figure 3.16 (f)). Figure 3.17 shows samples of cell images that were correctly predicted, and by eye there was no discernible difference between them.

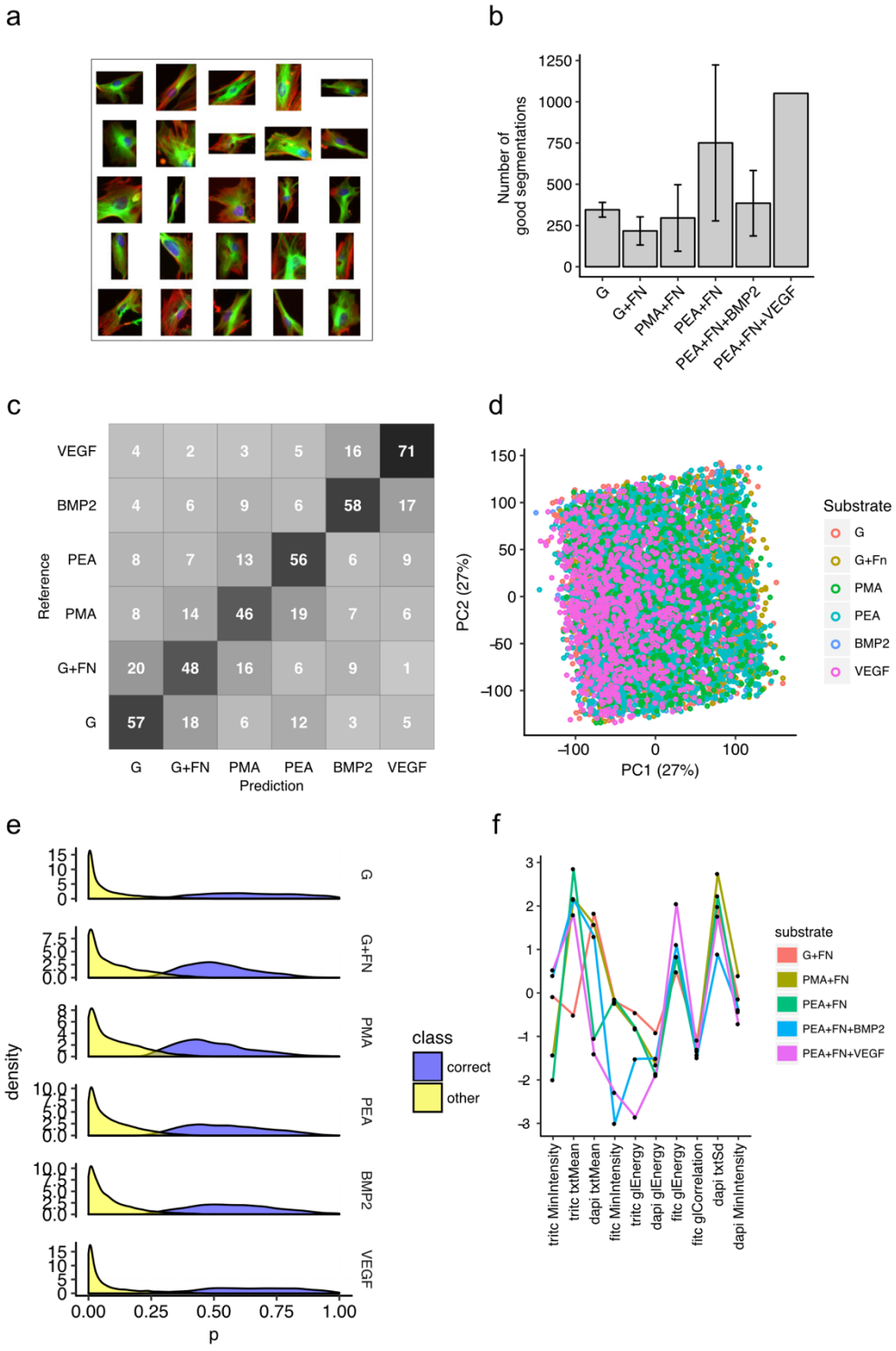

Figure 3.16 Results of a logistic regression used to predict which ECM a given MSC was cultured on after 3 d based on β-tubulin, actin, and nuclei image analysis. (a) Random sample of good segmentations. (b) Mean ± standard deviation of number of segmentations by artificial ECM type (8,346 good segmentations; n = 4, PMA n = 3, BMP2 n = 3, VEGF n = 1 sample replicate(s)). More variation was observed across PEA replicates. (c) Confusion matrix showing average percentage of classifications for each substrate, showing similar accuracy for all substrates. (d) PCA of all features. (e) Logistic regression probability of belonging to each class by correct or incorrect prediction, showing a spread of confidence for each class. (f) Top 10 most significant coefficients of the logistic regression fit for each ECM, containing features from each image channel. tritc = actin, dapi = nuclei, fitc = β-tubulin, "txt" prefix = intensity histogram feature, "gl" prefix = grey-level co-occurrence matrix feature.

Table 3.5 Per-class performance metrics for prediction of substrates at 3 d

	G	G+Fn	PMA+FN	PEA+FN	PEA+FN+BMP2	PEA+FN+VEGF	Mean
Precision	0.57	0.48	0.46	0.56	0.58	0.70	0.56
Recall	0.58	0.38	0.36	0.77	0.55	0.55	0.53
f-1 score	0.57	0.42	0.40	0.65	0.56	0.62	0.54

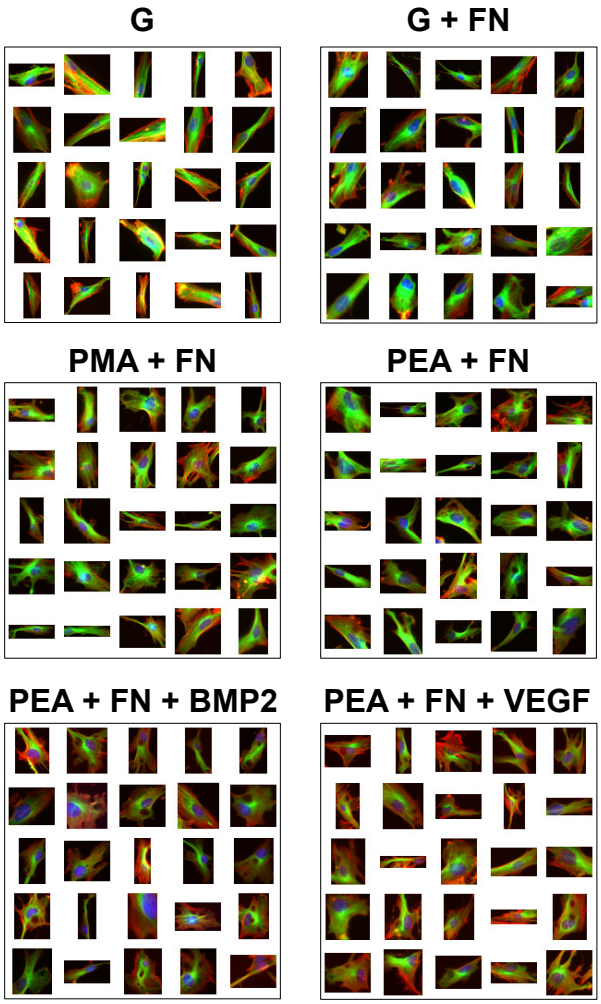

Figure 3.17 Random samples of correctly predicted MSCs after 3 d culture.

3.3.2.4 Analysis of Mid-Late Osteogenesis

ALP is an enzyme involved in osteogenesis, expressed at the cell surface (Sharma et al. 2014; Pittenger et al. 1999; Štefková et al. 2015), and necessary to initiate mineralization of bone nodules (Wennberg et al. 2000; Bellows et al. 1991). MSCs were imaged for ALP, actin, and nuclei at 14 d (Stein & Lian 1993) culture on the artificial ECMs. 24,797 segmentations were obtained from the images and 15,715 were predicted as good segmentations. A sample of segmentations kept for analysis is shown in Figure 3.18 (a), and the number of these per artificial ECM in Figure 3.18 (b) shows that cell number across the replicates was alike; note that

there was only one replicate for PMA with FN. The average prediction rates using a logistic regression on all the features was again accurate for cells cultured on control substrates, with those on PMA with FN and PEA with FN and VEGF being similarly well predicted. In contrast to previous predictions, cells cultured on PEA with FN were relatively accurate (Figure 3.18 (c)). Performance metrics were, however, particularly low (Table 3.6). No clustering was observed using PCA (Figure 3.18 (d)). The lower performance was reflected by the distribution of class probabilities in Figure 3.18 (e), where besides the control substrates, most probabilities were centred around lower values. Features for all three channels were contained in the top 10 most informative (Figure 3.18 (f)). In agreement with previous random samples there was no easily perceptible difference in images of correct segmentations besides a brighter actin stain on control substrates (Figure 3.19).

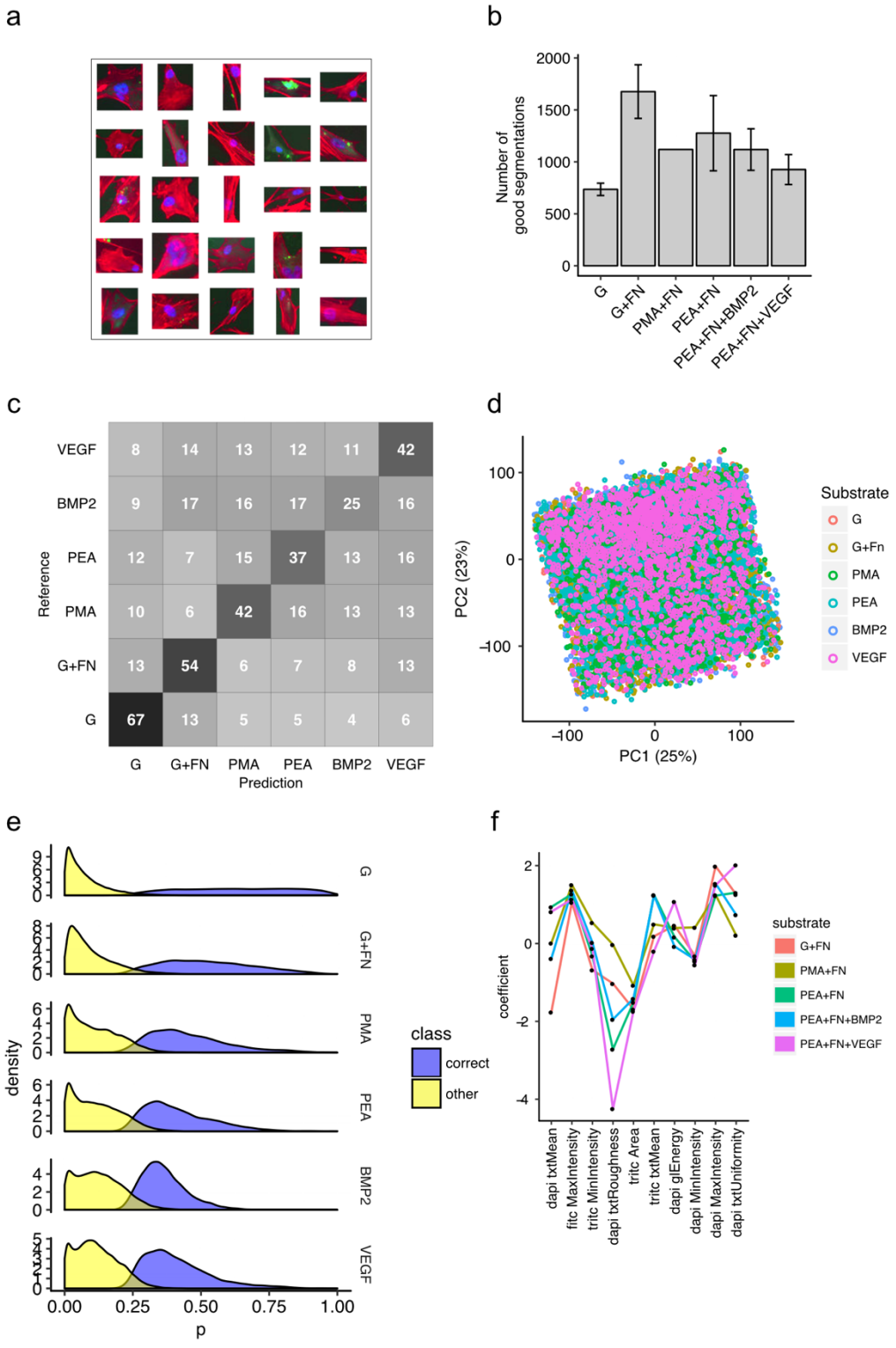

Figure 3.18 Results of a logistic regression used to predict which ECM a given MSC was cultured on after 3 d based on ALP, actin, and nuclei image analysis. (a) Random sample of good segmentations. (b) Mean ± standard deviation of number of segmentations by artificial ECM type (15,715 good segmentations; n = 3, G+FN n = 2, PMA n = 1 sample replicate(s)), showing small variation in cell number across replicates. (c) Confusion matrix showing average percentage of classifications for each substrate, showing higher accuracy from MSCs on control, PMA, PEA and VEGF substrates. (d) PCA of all features. (e) Logistic regression probability of belonging to each class by correct or incorrect prediction. Polymer-based MSC predictions were less confident in comparison to the control substrates. (f) Top 10 most significant coefficients of the logistic regression fit for each ECM, which contains features from all three image channels.

Table 3.6 Per-class performance metrics for prediction of substrates at 14 d

	G	G+Fn	PMA+FN	PEA+FN	PEA+FN+BMP2	PEA+FN+VEGF	Mean
Precision	0.67	0.54	0.42	0.37	0.25	0.42	0.45
Recall	0.51	0.56	0.22	0.52	0.41	0.30	0.42
f-1 score	0.58	0.55	0.29	0.43	0.31	0.35	0.42

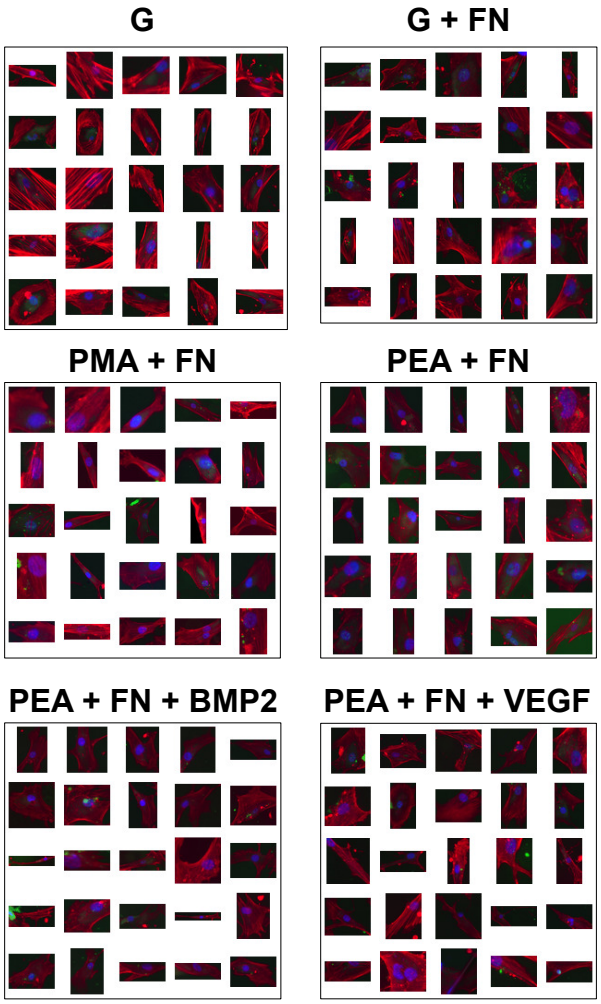

Figure 3.19 Random samples of correctly predicted MSCs after 14 d culture.

3.3.2.5 Analysis of Late Osteogenesis in BM

In the previous chapter, MSCs were cultured in BM for 21 d and their patterns of OCN, actin, and nuclei compared. Their high density meant the image processing pipeline under-segmented based on actin, so the following analysis focused on their nuclei alone, which were better segmented (Figure 3.20 (a)). 24,764 segmentations were detected in the images and 19,408 were predicted as good segmentations. In keeping with prior counts of segmentations kept for analysis, the number of segmentations was similar across replicates, apart from glass, which had fewer replicates (Figure 3.20 (b)). Figure 3.20 (c) shows the

prediction rates of the logistic regression. Cells on control substrates were accurately predicted in comparison to the other ECMs, and PMA was also relatively high. Notably, a large number of MSCs cultured on PEA with FN and VEGF were predicted to be cultured on glass with FN. Performance metrics were low, possibly attributed to the reduced number of features for analysis in comparison to those with more than one channel (Table 3.7). PCA revealed no clustering by condition (Figure 3.20 (d)). The distributions of prediction probabilities were similar except for glass with FN, which featured a bimodal distribution (Figure 3.20 (e)). While only nuclei features were included in this analysis, the mean intensity of DAPI staining was again considered informative for the logistic regression (Figure 3.20 (f)). No features were estimated to be significantly different as a function of substrate using the mixed effects model.

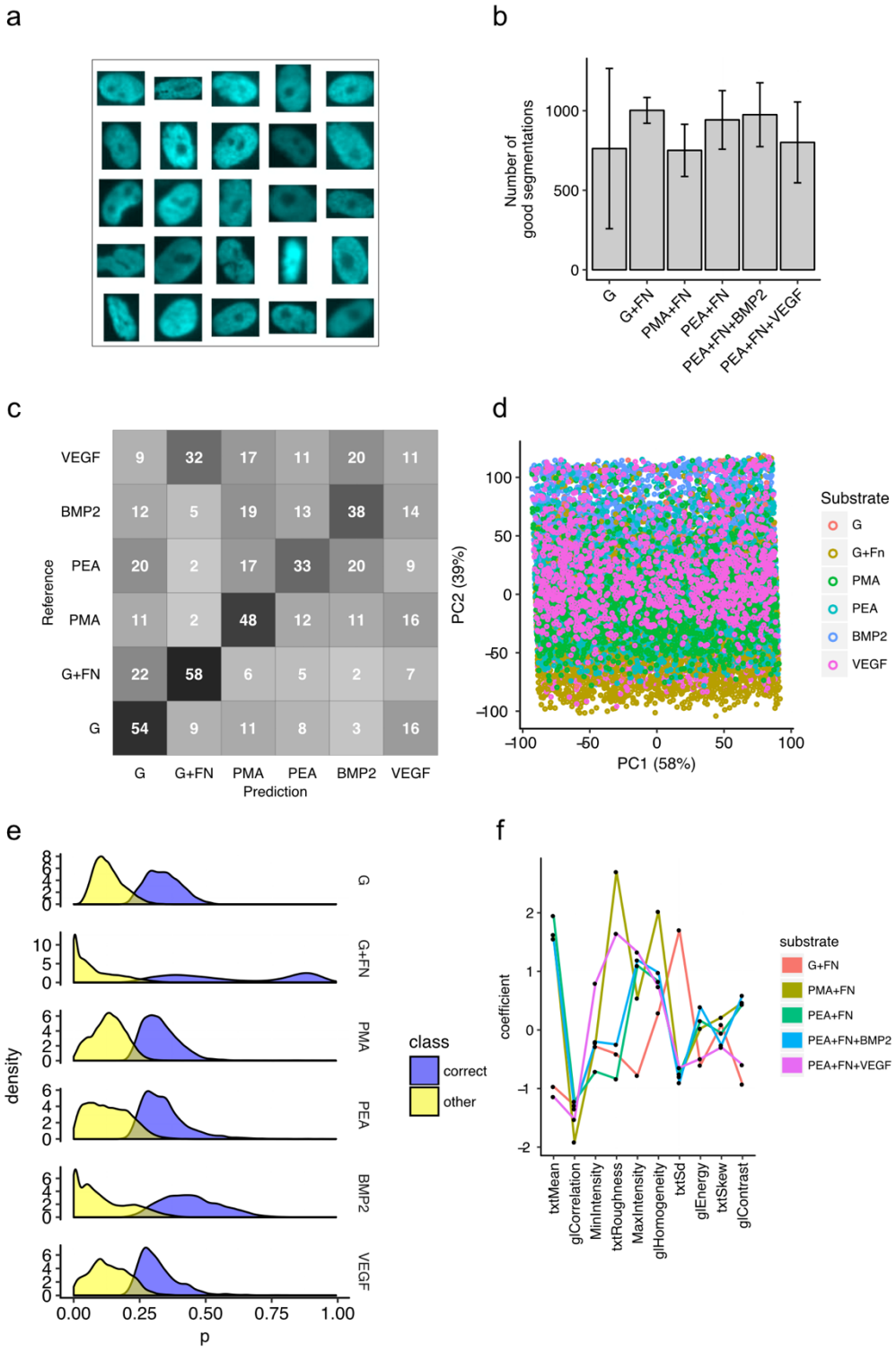

Figure 3.20 Results of a logistic regression used to predict which ECM a given MSC nucleus was cultured on after 21 d based on nucleus image analysis. (a) Random sample of good segmentations. (b) Mean ± standard deviation of number of segmentations by artificial ECM type (good segmentations 19,408; n = 4, glass n = 2 sample replicates). There was more variation across glass replicates than the other conditions. (c) Confusion matrix showing average percentage of classifications for each substrate. Predictions for MSCs on control and PMA substrates had the highest accuracies. (d) PCA of all features. (e) Logistic regression probability of belonging to each class by correct or incorrect prediction, showing a bimodal distribution for glass with FN, and similar distributions for the other substrates. (f) Top 10 most significant coefficients of the logistic regression fit for each ECM.

Table 3.7 Per-class performance metrics for prediction of substrates at 21 d

	G	G+Fn	PMA+FN	PEA+FN	PEA+FN+BMP2	PEA+FN+VEGF	Mean
Precision	0.54	0.58	0.48	0.33	0.37	0.11	0.40
Recall	0.23	0.61	0.39	0.45	0.44	0.16	0.38
f-1 score	0.32	0.60	0.43	0.38	0.42	0.13	0.38

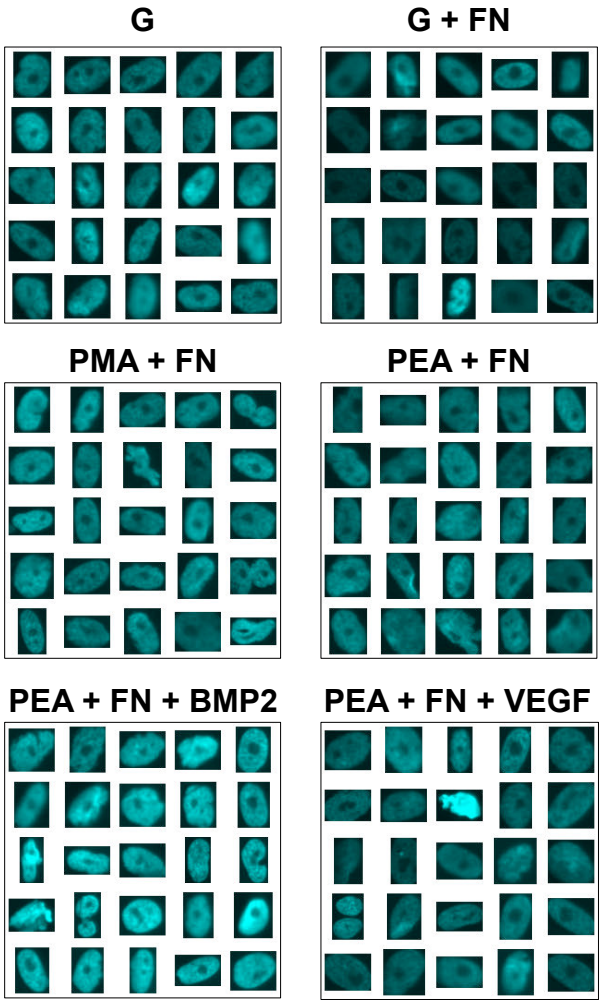

Figure 3.21 Random samples of correctly predicted MSC nuclei after 21 d culture.

3.3.2.6 Analysis of Late Osteogenesis in OM

In addition to the 21 d culture in BM analysed above, MSCs were also cultured in OM, the images of which were initially examined in the previous chapter. As above, images of the MSC nuclei segmented from these images were analysed. 37,072 segmentations were obtained from the images and 11,653 were predicted as good segmentations. Segmentations kept for analysis are shown in Figure 3.22 (a), but there was comparatively greater variation in the number of kept segmentations by replicate for PEA with FN and growth factors (Figure 3.22 (b)). MSC nuclei on control substrates were relatively accurately predicted, and those

on PMA and especially PEA with FN and VEGF were also well predicted (Figure 3.22 (c)), but overall prediction performance and confidence was low (Table 3.8, Figure 3.22 (e)). No clustering was observed using PCA (Figure 3.22 (d)). Perhaps importantly, most of the top 10 (out of 17) most informative features in this analysis (Figure 3.22 (f)) were also important for the previous BM analysis (Figure 3.20 (f)). MSCs cultured on PMA with FN or PEA with FN and VEGF had nuclei with greater perimeter and area than those cultured on glass with FN (Figure 3.24).

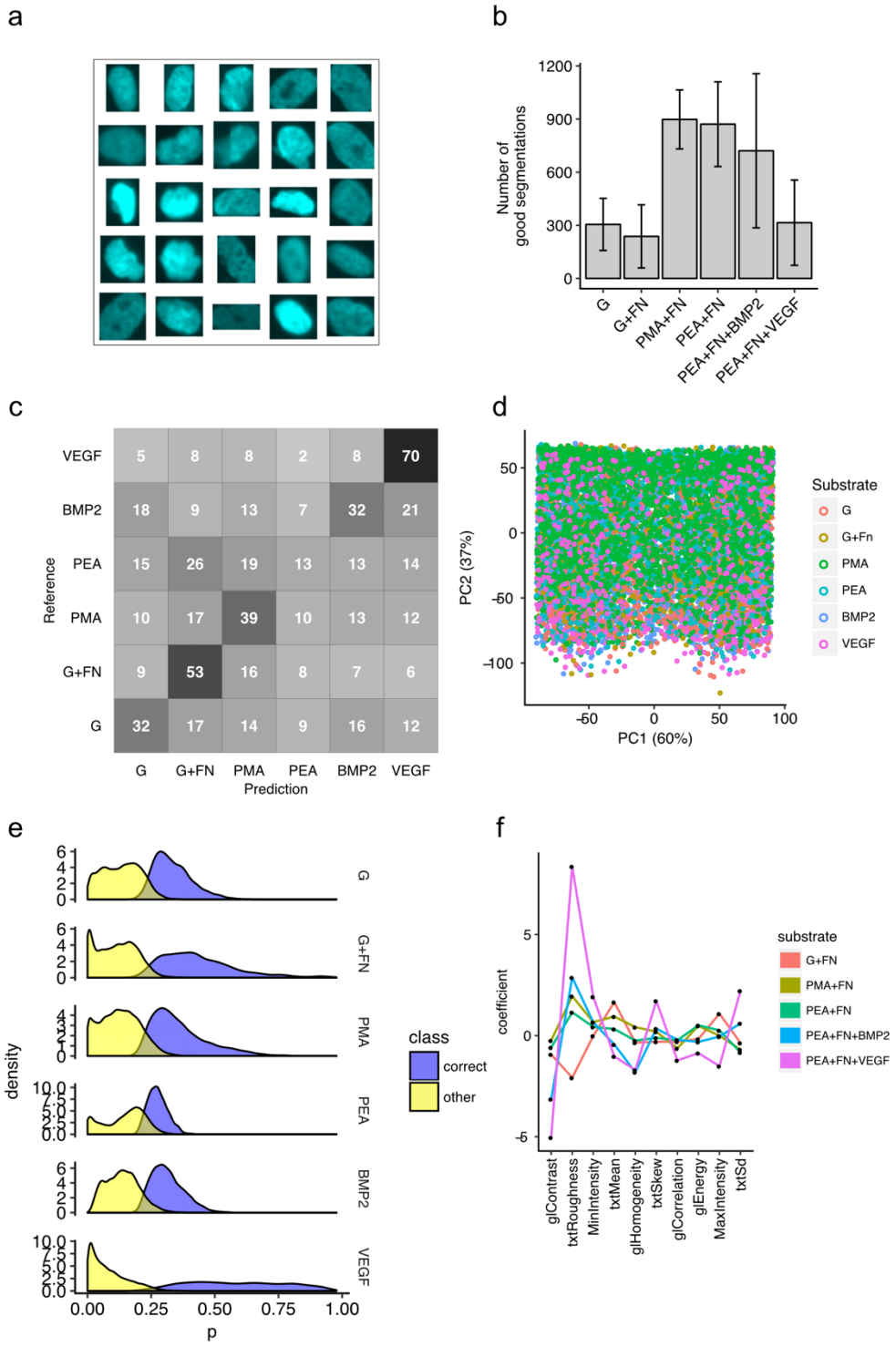

Figure 3.22 Results of a logistic regression used to predict which ECM a given MSC nucleus was cultured on after 21 d in OM based on nucleus image analysis. (a) Random sample of good segmentations. (b) Mean ± standard deviation of number of segmentations by artificial ECM type (11,653 good segmentations; n = 4, PEA n = 3 sample replicates). Variation across replicates was greater on the growth factor-based substrates. (c) Confusion matrix showing average percentage of classifications for each substrate. VEGF had a comparatively high accuracy. Performance for glass with FN and PMA was also relatively good. (d) PCA of all features. (e) Logistic regression probability of belonging to each class by correct or incorrect prediction. Probabilities of VEGF predictions were spread, whereas those for other classes were generally lower over a smaller range. (f) Top 10 most significant coefficients of the logistic regression fit for each ECM.

Table 3.8 Per-class performance metrics for prediction of substrates at 21 d in OM

	G	G+Fn	PMA+FN	PEA+FN	PEA+FN+BMP2	PEA+FN+VEGF	Mean
Precision	0.32	0.53	0.39	0.13	0.32	0.70	0.40
Recall	0.23	0.24	0.55	0.23	0.47	0.38	0.35
f-1 score	0.27	0.33	0.46	0.17	0.38	0.49	0.35

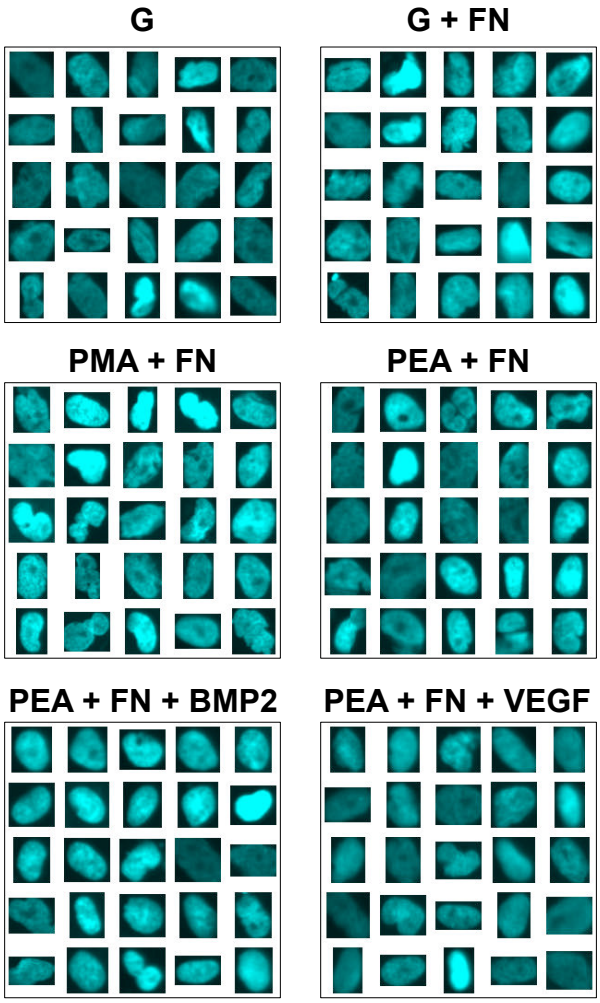

Figure 3.23 Random samples of correctly predicted MSC nuclei after 21 d culture in OM.

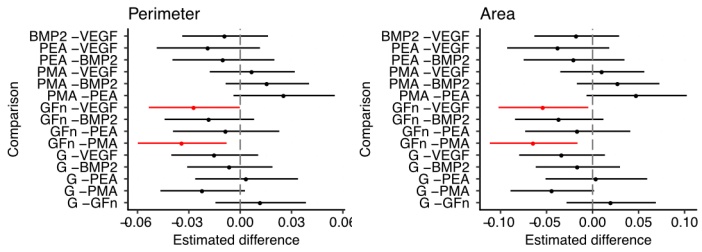

Figure 3.24 Features of MSC nuclei cultured for 21 d on the artificial ECMs in OM with significantly different estimated means ± 95 % confidence interval. Nuclei of MSCs cultured on PMA or VEGF tended to be larger than glass with FN. Red, significant (confidence interval does not include 0).

These results indicate that the addition of OM caused nuclei of MSCs cultured on PEA with FN and VEGF to become particularly distinct, and that cells cultured on control and PMA with FN substrates are also relatively distinguishable by their nuclei. Further, nuclei on PEA with FN and VEGF and on PMA tend to be larger in comparison to control conditions.

3.4 Discussion

In this chapter, an image processing and data analysis pipeline was developed to investigate MSCs using fluorescence images. It was used to analyse datasets of MSC morphology and osteogenic-related protein expression patterns studied in the previous and current chapters. This approach lent to a more robust conclusion compared to hand-picking individual cells, of which no difference could be discerned, and also yielded information about heterogeneity across and within the substrates (Caicedo et al. 2017).

The image processing pipeline was based around marker-based watershed segmentation of cells by their actin cytoskeleton and nuclei. This model-based approach is reported to work well for fluorescence images (Caicedo et al. 2017). Due to the size of the mosaic images it was infeasible to use commonly employed biological image analysis tools such as CellProfiler (Carpenter et al. 2006; Kamentsky et al. 2011), and more flexible to use MATLAB and its Image Processing Toolbox (Gonzalez & Woods 2010a; Gonzalez et al. 2009). While the pipeline was focused on whole-cell segmentation, it was built to allow nuclei alone to be analysed. Furthermore, it accommodated the addition of a third channel to allow investigation of nuclear or cytoplasmic proteins. A reported disadvantage of using custom programs is that they are not modular, unlike CellProfiler, or easy to use for routine high-throughput processing (Carpenter et al. 2006), but the pipeline here was built in an object-oriented style and it supported parallel processing using the MATLAB Parallel Computing Toolbox. Still, it was specifically written to process images obtained using the protocols described in the methods sections of this and the previous chapter, and so will not necessarily translate to other image sets (for example, those recorded using a different microscope) without modification, a common problem with scripted and CellProfiler-based pipelines (Carpenter et al. 2006; Caicedo et al. 2017). However, writing code for the flexible analysis of non-standardised images is an

accepted solution (Eliceiri et al. 2012). Future work could focus on enabling the analysis of different image magnifications and developing a graphical interface for the MATLAB pipeline to enable wider usage; the image labeller has already been written with a graphical interface.

The strength of the conclusions in this chapter should be considered in light of the effectiveness of the cell segmentation process. The segmentation pipeline worked well for nuclei as evidenced for example by Figure 3.21, but was not as successful in identifying the entire cell body using the actin cytoskeleton. In many cases the putative cell segmentations were fragments of the cell body surrounding the nucleus, for example those shown in Figure 3.5. While the filter process was developed to remove these bad segmentations, it was clear that some remained (Figure 3.5). As a result, the information about actin morphology and texture features was only partial in comparison to nuclei and therefore the contribution of nuclei features to conclusions would be more robust and informative than those of cell bodies. This may then underlie the fitted predictive models attributing higher importance to nuclei features in contrast to the actin-based features (for example Figure 3.10 (f), and in the high ranking of mean DAPI intensity in the other analyses). Further work is needed to improve the cell segmentation process either through the staining procedure or the segmentation algorithm.

Accurate prediction by a classifier of which substrate a cell was cultured on indicates a relationship between the various measured features and properties of the substrate, albeit through a complex underlying model, here a logistic regression (Lever et al. 2016). The results presented here suggest that the FN conformations, particularly the isolated conformation on PMA, and the addition of growth factors, particularly PEA with FN and VEGF, had a discriminating effect on the measured protein and nuclei features, at least for a subpopulation of cells. These studies of using a classifier may be useful for future work focusing on variations of the artificial ECMs, for example, the addition of new or combinations of growth factors, and the use of blends of PEA and PMA (Mnatsakanyan et al. 2015), with an interest in establishing how similar the cell phenotypes are to those cultured on the artificial ECMs presented here — these

studies may benefit from using a supervised method previously demonstrated for fluorescence phenotyping (Jones et al. 2009; Sommer et al. 2011).

Using a mixed effects model allowed inferences to be made about differences in particular features while controlling for inherent variation across replicates (Krzywinski et al. 2014; Gelman & Hill 2007), but did limit these inferences to experiments where there was more than one replicate for each condition, and ignored relationships between features that would be captured by the classification analysis. Nevertheless, comparisons of cell morphometrics without adjusting for replicate variability has been used to contrast other cell types on the artificial ECMs. At 3 h in serum-free culture conditions, it was reported that mouse C3H10T1/2 MSCs on the isolated FN conformation were more circular compared to those on the network FN conformation (Rico et al. 2016), and L929 fibroblasts had a larger cell area on the network FN conformation (Vanterpool et al. 2014). After 24 h, cell shape, judged by measurements of the segmented actin cytoskeleton, tended to be longer or narrower — have a greater aspect ratio and eccentricity — on PMA with FN than PEA with FN. MSCs cultured on PMA with FN tended to have more homogeneous actin cytoskeletons and rounder nuclei at 5 d compared to PEA-based surfaces. While no specific feature of nuclei in MSCs after 21 d culture in BM was estimated to be different, those cultured in OM on PMA with FN or PEA with FN and VEGF had a larger perimeter and area compared to glass with FN.

Image processing studies of MSCs have suggested that actin and nuclei features at early time points (24 h – 3 d) were predictive of differentiated phenotypes (Liu et al. 2010; Treiser et al. 2010; Marklein et al. 2016; Matsuoka et al. 2013; Vega et al. 2012; Vega et al. 2015). Analysis of the actin cytoskeleton and nuclei of MSCs at 24 h and 5 d showed distinct differences in morphology for cells on PMA with FN and PEA with FN and VEGF to those on PEA with FN alone. In the previous chapter, MSCs on the isolated FN conformation had a higher expression of OPN and a distinct biochemical profile compared to those on the network conformation. The relationship between MSC morphology and osteogenesis has been reported using several biomaterials that constrict cell shape and cytoskeletal conformation (McBeath et al. 2004; Kilian et al. 2010).

Previous image analysis studies of MSC behaviour have suggested there are donor- and time-specific differences in image features (Liu et al. 2010; Treiser et al. 2010; Marklein et al. 2016; Matsuoka et al. 2013), but there are conflicting studies as to whether these dissimilarities extend to lineage-specific markers and proliferation (Stenderup et al. 2001; Surdo & Bauer 2012). The different datasets analysed in this chapter were from different donors but only one donor was analysed in each experiment; this may partly underlie the variation in segmentation counts across the artificial ECMs. However, there was a general theme in that cells on PMA with FN and PEA with FN and VEGF were most distinct. Repeats using more donors would strengthen these conclusions.

4 Metabolomics

4.1 Introduction

In the previous chapters, single protein measurements indicated varying degrees of osteogenesis in MSCs cultured on the artificial ECMs, and high-content image analysis of cytoskeletal, nuclear, and lineage-specific protein expression suggested that the ECM compositions differentially influenced MSC morphometrics. Further, SR-FTIR analyses revealed broad biochemical differences in lipid, protein, and nucleic acid contents in MSCs that were related to the underlying FN conformation.

In this chapter, the metabolomes of MSCs cultured on the artificial ECMs are compared using LC-MS to gain understanding of the different biochemical states alluded to by the previous results. While there were no individual biomarkers that distinguished the different artificial ECMs in this dataset, it is shown that metabolic pathway analyses revealed large differences between MSCs cultured on the different FN conformations.

PiMP was used to analyse the metabolomics data. To address the issues detailed in the introduction, of consolidating PiMP and improving how LC-MS metabolomics data is presented, the development of a tool to explore and visualise the relationships (Figure 4.1) that underlie the peak-metabolite-pathway annotations resulting from an LC-MS metabolomics experiment is also described in this chapter.

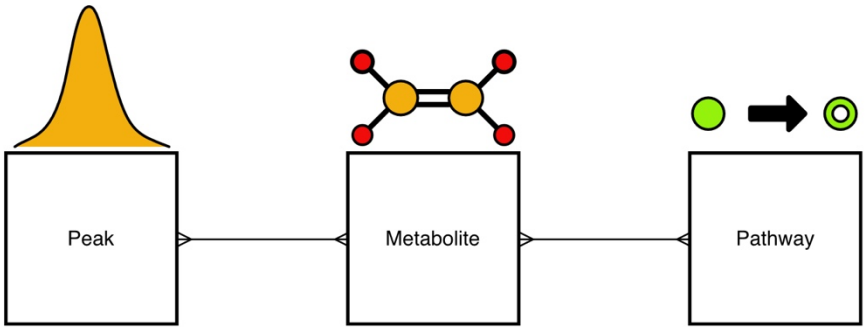

Figure 4.1 Peak-metabolite-pathway relationships. Several peaks can be annotated by a single metabolite. Several metabolites can annotate a single peak. Multiple metabolites can belong in the same pathway. A single metabolite can belong to several pathways.

The aims of this chapter were to: (1) develop a quick results (QR) application for visualising and exploring untargeted LC-MS metabolomics data to improve PiMP, and (2) use this and PiMP to compare the metabolomes of MSCs cultured on the artificial ECMs.

4.2 Materials and Methods

4.2.1 Software Development

Software was version controlled using a private GitLab server hosted by Glasgow Polyomics. Local instances of software were tested before running on a production server (http://www.polyomics.mvls.gla.ac.uk).

4.2.2 Filtering Related Data Interactively (FiRDI)

FiRDI was written in JavaScript using jQuery, AlaSQL, DataTables, and jQuery.

4.2.3 Quick Results (QR) Application

The QR application was written in Python using PiMP, Django, NumPy and Pandas packages; and in JavaScript using FiRDI, d3, and HighCharts. Pre-existing features in PiMP were used in the QR application, and Bootstrap was used for web page layout.

4.2.4 Metabolomics

MSCs were cultured as described in the previous chapters for 7 or 14 days on the artificial ECMs with n = 4 replicates. For extraction, media was discarded and samples were immersed in 1:3:1 methanol:chloroform:water on a shaker for 1 h at 4 °C. Extracts were then centrifuged at 4 °C for 3 min at 13,000 RPM before storage at -80 °C before use.

LC-MS was performed by Glasgow Polyomics (University of Glasgow, UK), using the following approach: Hydrophilic interaction liquid chromatography (HILIC) was carried out on a Dionex UltiMate 3000 RSLC system (Thermo Fisher Scientific, UK) using a iHILIC-Fusion(P) column (150 m, 4.6mm, 5 m column, Hilicon). Pooled samples were run for quality control.

The column was maintained at 30°C and samples were eluted with a linear gradient (20 mM ammonium carbonate in water (A) and acetonitrile (B) over 30 min at a flow rate of 0.3 ml min^{-1} as follows:

Time (min)	%A	%B
0	20	80
15	80	20
15	95	5
17	95	5
17	20	80
30	20	80

The injection volume was 10 µL and samples were maintained at 4 °C prior to injection. For the MS analysis, a Thermo Orbitrap Exactive (Thermo Fisher Scientific) was operated in polarity switching mode and the MS settings were as follows: resolution 50,000, AGC 106, m/z range 70–1400, sheath gas 40, auxiliary gas 5, sweep gas 1, probe temperature 150°C, capillary temperature 275°C.

For positive mode ionisation: source voltage +4.5 kV, capillary voltage +50 V, tube voltage +70 kV, skimmer voltage +20 V. For negative mode ionisation: source voltage -3.5 kV, capillary voltage -50 V, tube voltage -70 V, skimmer voltage -20 V.

Mass calibration was performed for each polarity immediately prior to each analysis batch. The calibration mass range was extended to cover small metabolites by inclusion of low-mass contaminants with the standard Thermo calmix masses (below *m/z* 1400), $C_2H_6NO_2$ for positive ion electrospray ionisation (PIESI) mode (*m/z* 76.0393) and $C_3H_5O_3$ for negative ion electrospray ionisation (NIESI) mode (*m/z* 89.0244). To enhance calibration stability, lock-mass correction was also applied to each analytical run using these ubiquitous low-mass contaminants.

4.2.5 Metabolomics Data Analysis

Analysis was performed using PiMP (Gloaguen et al. 2017) with the following data processing parameters: relative standard deviation (RSD) filter 0.5, noise filter 0.8, parts per million (ppm) window 3, minimum intensity 5000, minimum number of detections 3, RT window 0.05, RT alignment using CowCoda. Peaks

comparisons with p-value < 0.05 were rejected (Scheltema et al. 2011; Smith et al. 2006; Gloaguen et al. 2017). A modified pathway activity score based on the pathway-level analysis of gene expression (PLAGE) algorithm (Tomfohr et al. 2005) (Personal communication, Karen McCluskey, Glasgow Polyomics, University of Glasgow, UK) was used to compare metabolic pathway activity. The modified score of a pathway is its PLAGE score combined with a hypergeometric (overrepresentation) test using Stouffer's method (Whitlock 2005), which can be interpreted as adjusting the pathway activity to compensate for its size. Modified activity scores that were associated with pathways that had enough detected metabolites compared with the pathway size (overrepresentation test, false discovery rate < 0.1) and that had raw PLAGE score p-values < 0.05 in any comparison were considered significant.

4.3 Results

4.3.1 Developing the QR Application

The QR web application was designed to display the peaks, metabolites, and pathways found in a metabolomics experiment using PiMP (Section 1.4.1; Figure 1.7). In particular, so that this information was in a single page with interactivity to allow a user to browse the relationships between these biological entities. To do this, a JavaScript module named Filter related data interactively (FiRDI) was developed which can take any relational dataset and present it in a webpage in this manner. Displaying relational data in this way may be useful for other non-metabolomics and even non-biological datasets, so developing the QR application in this way — creating a distinct and reusable JavaScript module that was used to present the metabolomics data in the QR application — means FiRDI may be used to benefit other unrelated projects. In this section, first, FiRDI is described, and then the QR web application extension to PiMP, which uses FiRDI, is detailed.

4.3.1.1 FiRDI

FiRDI is designed to display related data tables in an interactive webpage. It presents each data type — for example, peak, metabolite, or pathway — in its own table, and each table is interactive in the sense that selecting any row will filter out the unrelated data in the other tables. As an example, consider that a

certain peak is of interest since it is changing in abundance between experimental conditions, and therefore picking out its associated metabolites is of biological interest. All the peaks and metabolites resulting from a metabolomics experiment are initially displayed in their respective tables, but selecting this peak restricts the list of displayed metabolites to only those related to that peak. This filter is removable by deselecting said peak.

FiRDI is run in a webpage created by the user that contains empty table elements for each data type to be displayed, meaning that the user controls the webpage layout. The relational data is provided to FiRDI in JavaScript object notation (JSON) format, and a JavaScript object in a prescribed format is also provided by the user which describes the relationship of the tables (Figure 4.2 (a)). From this JavaScript configuration object, FiRDI compiles a partial SQL query based on the described relationships. It is comprised of a SELECT clause for all table columns and a parameterised WHERE clause. The WHERE clause has a series of predicates that follow the relationships, and is parameterised such that when a user selects a table row, the WHERE clause will filter out any unrelated data when the query is run. FiRDI then repopulates the tables with the filtered dataset. Moreover, the filter is cumulative. For example, if visualising a peak-metabolite-pathway dataset, a pathway is selected to see all related metabolites and peaks, and a further filter can be added by selecting one of those metabolites to show only peaks related to that metabolite in that pathway. Further, these filters can be switched on and off in any order.

The relational data should be interpretable as a connected path graph and have only unique rows. In addition, the relationships should be embodied as junction/joining tables (Figure 4.2 (b)). As well as describing the table relationships, the configuration object also contains information about which tables are junction/joining tables, so that when the data is presented in the webpage they are hidden from view since they are not relevant for visualising the data (Figure 4.2 (c)).

FiRDI is not restricted to viewing metabolomics data nor limited to the number of tables of data that can be displayed and filtered (within the limits of computational resources). Selecting related data to a datum of interest is a common and general task, and FiRDI allows the display and querying of datasets

that can be arranged in the same way as the metabolomics data example (Figure 4.2 (b)). Finally, FiRDI also provides functionality to replace the initial data with a new dataset that has identical relationships and structure without reloading the entire webpage. This is useful, for example, when a dataset is perhaps time-sensitive and needs to be replaced with up-to-date data.

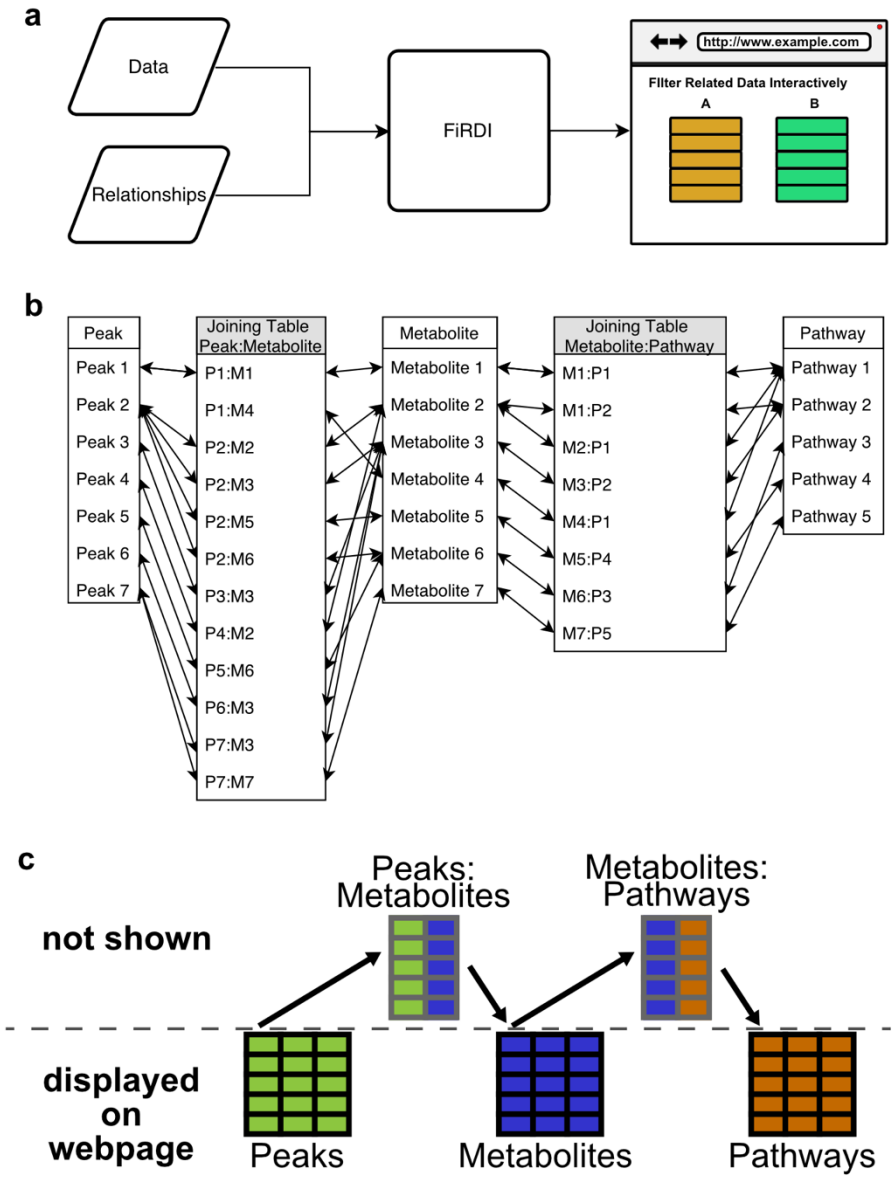

Figure 4.2 How FiRDI works. (a) FiRDI is used by providing a set of relational data and a description of their relationships. It then populates tables in a web page created by the user and provides interactive filtering functionality to visualise and explore the relationships in a dataset. (b) An example metabolomics dataset that would be provided to FiRDI. Each Peak (peak 1, peak 2, …, peak 7) is annotated (related) by a metabolite via a joining table Peak:Metabolite. The same goes for metabolites that are part of a pathway. (c) Using the example metabolomics dataset (b), only the Peak, Metabolites, and Pathway tables are

displayed by FiRDI, and the Peak:Metabolite and Metabolite:Pathway joining tables are hidden from view.

4.3.1.2 PiMP QR Application

FiRDI was described in the previous section, which is a JavaScript module that enables the visualisation and interactive filtering of related data in a webpage. The use of FiRDI to extend PiMP with the QR application is detailed in this section.

PiMP displays information about the peaks, metabolites, and pathways in a metabolomics dataset in different places. Further, the relationships underlying the peak-metabolite-pathway mapping could be made more explicit to assist the understanding of metabolomics data. These two problems are addressed using the QR application, which displays the data in a single interface and makes clear the aforementioned complexity in LC-MS metabolomics features.

The QR application is a single webpage accessed through a link in PiMP (Figure 4.3). It presents the results of an "analysis", which is the result of pre-processing of a metabolomics dataset and comparison between experimental conditions (Gloaguen et al. 2017). The title of the analysis and a settings menu heads the QR webpage (Figure 4.3 (1)), and below it is the data in tables produced by FiRDI. The columns of the peaks table (Figure 4.3 (2)) display peak identifiers (ID), RTs, masses, and a symbol representing the direction of significant log fold changes across comparisons (titled c0, c1, c2) — the "?" symbol above the peaks table is a key for the comparisons. The metabolites table (Figure 4.3 (3)) lists the metabolites, and the pathways table lists the pathways (Figure 4.3 (4)). Each table is orderable, paged, and selectable. Below each table is an information pane (Figure 4.3 (5, 6, 7)) where information related to the table row that was selected is displayed.

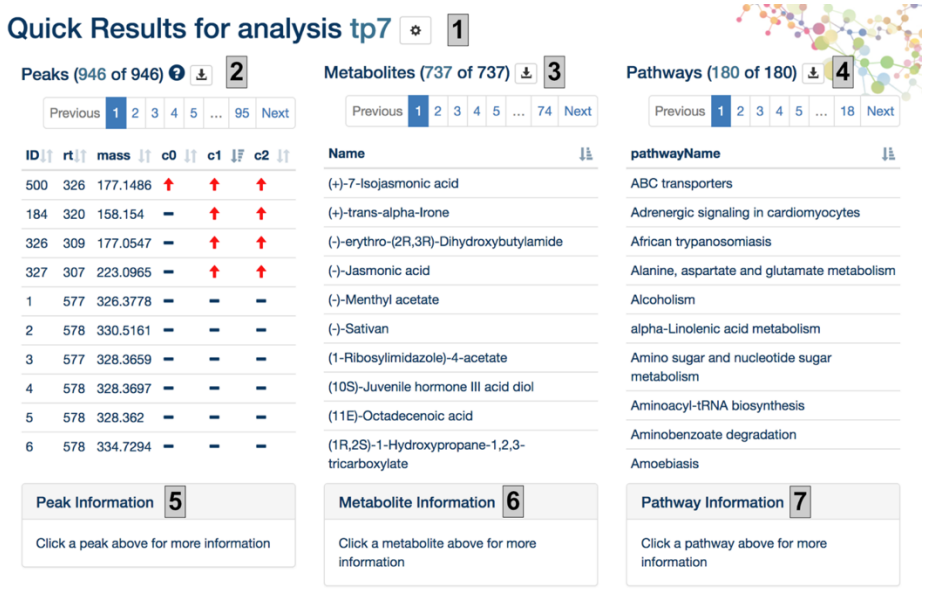

Figure 4.3 Screenshot of the QR webpage with an example analysis titled "tp7". (1) Title and settings button. (2, 3, 4) Interactive tables produced by FiRDI, with ordering, pagination, and row selection functions. (5, 6, 7) Information panes that display information related to the selected table row above them. Note that nothing is selected in this example.

When a peak is selected (Figure 4.3 (2)) the peak information pane (Figure 4.3 (5)) is populated with relevant information (Figure 4.4): the ID, type of peak, and MS polarity mode that the peak was recorded in is shown. Further, two charts are offered to the user through loading buttons that call asynchronous functions, since the charts could take a long time to load (Figure 4.4 (1,2)). The first (Figure 4.4 (1)) is an extracted ion chromatogram of the peak across all of the experimental samples, which is taken directly from PiMP. The second is a box-and-whisker plot of the intensity of that peak across experimental groups (Figure 4.4 (2)). The metabolite information pane (Figure 4.5 (left)) provides the international chemical identifier key (InChiKey) of the selected metabolite, its parts per million, and what adduct was detected. It also shows the Kyoto Encyclopaedia of Genes and Genomes (KEGG) structure for the compound and a link to the related KEGG compound database entry (Kanehisa et al. 2011). The pathway information pane displays a link to the related KEGG pathway database (Figure 4.5 (right)) (Kanehisa et al. 2011).

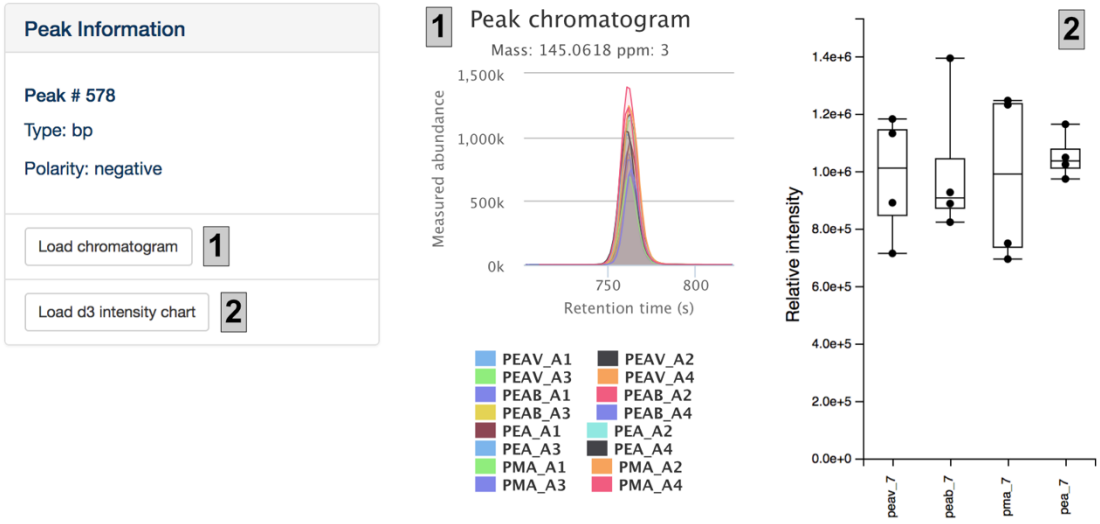

Figure 4.4 Peak information pane. (left) The peak information pane displays the peak ID, type, and polarity of the selected peak. It also provides two charts related to the peak, (1) an extracted ion chromatogram for all samples in the analysis, and (2) a box-and-whisker plot of peak intensities across experimental groups.

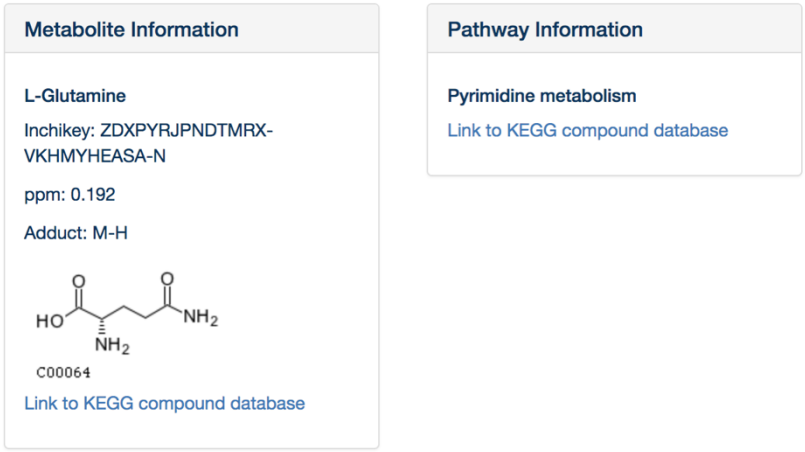

Figure 4.5 Metabolite and pathway information panes. Information related to the selected metabolite, its InChiKey, ppm, detected adduct, structure, and link to the KEGG compound database entry (left); and pathway, a link to its entry in the KEGG maps database (right) are shown in these panes.

The settings menu (Figure 4.6), accessed through a settings button (Figure 4.3 (1)), has controls to show only peaks that are significantly different across experimental groups, and to only show peaks that are annotated with authentic chemical standards (identifications).

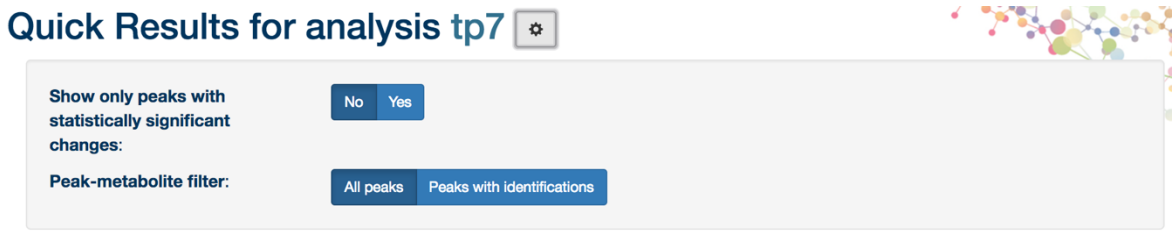

Figure 4.6 QR settings. The settings button (Figure 4.3 (1)) opens a settings menu containing two options to control what analysis data is shown: significantly changing peaks only, and peaks annotated with an authentic chemical standard only.

An example of exploring metabolomics data and its relationships using the QR application is shown in Figure 4.7. The pyrimidine metabolism pathway has been selected in the pathways table (Figure 4.7 (1)) to restrict the visualised data to peaks (Figure 4.7 (3)) and metabolites (Figure 4.7 (2)) in that pathway. Next, the metabolite L-glutamine has been selected to filter out unrelated peaks and pathways (Figure 4.7 (4)). It is seen that two peaks are related to L-glutamine, and selecting each peak in turn and viewing the peak information pane reveals that these are the MS positive and negative mode peaks for this metabolite.

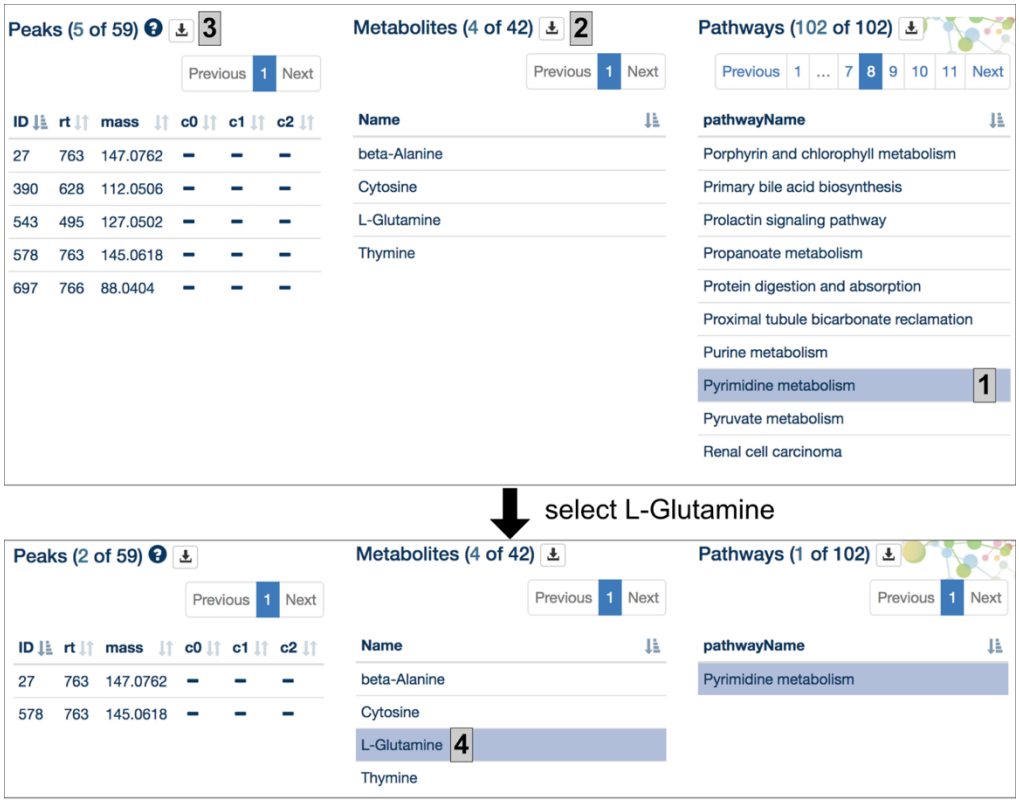

Figure 4.7 Example of using FiRDI to interactively filter metabolomics data in the QR application. Selecting the pyrimidine metabolism pathway in the pathways table (1) removes all metabolites (2) and their cognate peaks (3) that are not in the pathway. Selecting L-glutamine in the metabolites table after having filtered on the pyrimidine metabolism pathway shows only the peaks annotated with L-glutamine (4).

Figure 4.8 shows an outline of how PiMP, FiRDI, and the QR application work together. A user navigates to the QR page for their analysis of interest using PiMP, and the request is sent to a Django QR view (Figure 4.8 (1)). The view populates a Django template, the QR page, which contains the layout shown in Figure 4.3 (Figure 4.8 (2)). This runs the QR JavaScript code, which gets the analysis data from another Django view, the QR Data function (Figure 4.8 (3)), and together with a JavaScript configuration object, it is used to initialise FiRDI (Figure 4.8 (4)). FiRDI then fills the tables in the template (Figure 4.8 (5)) and the QR page is ready for use. Changing the settings (Figure 4.6) uses the functionality in FiRDI to load new data without having to reload the webpage (Figure 4.9): for example, a user changes the peak identification setting (Figure 4.9 (1)) and the QR JavaScript code gets the new data from the Django QR data view (Figure 4.9 (2)) and passes it to FiRDI (Figure 4.9 (3)), which empties and repopulates the peak, metabolites, and pathways tables in the QR page with the new data (Figure 4.9 (4)).

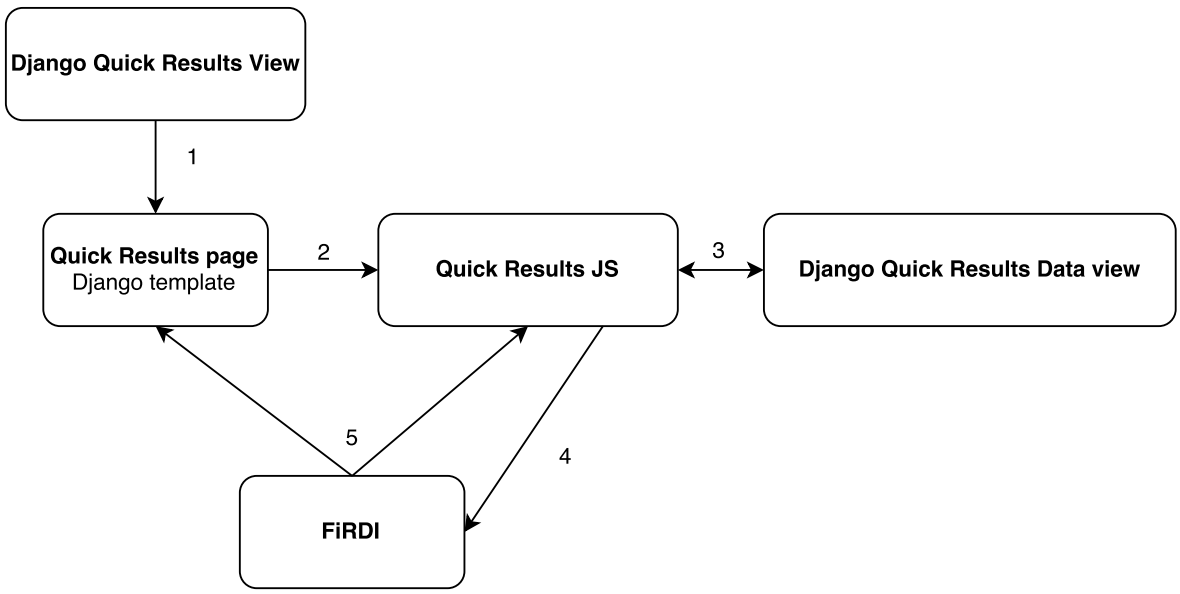

Figure 4.8 Starting the QR application. (1) Request to quick_results_view(), which fills the quick_result.html template. (2) quick_results.html runs pimp_quick_results_firdi.js. (3) pimp_quick_results_firdi.js gets the data from another view, get_quick_results_data(). (4) pimp_quick_results_firdi.js runs FiRDI.js with the data and configuration. (5) FiRDI.js populates the quick_results.html page and provides control to pimp_quick_results_firdi.js. The webpage is now ready to be used. Names suffixed with "()" indicate Python functions/Django views. Names suffixed with ".js" indicate JavaScript functions.

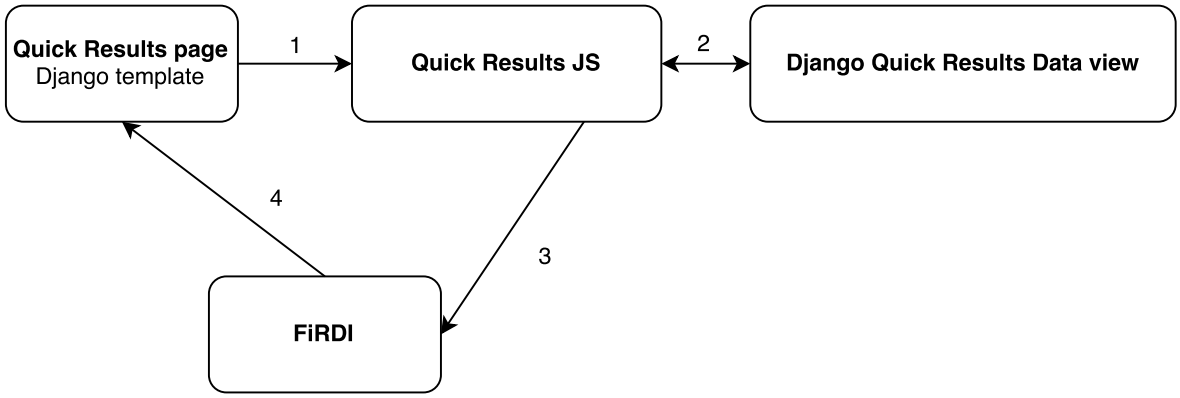

Figure 4.9 Loading new metabolomics data from the settings menu. (1) User redefines data (for example, show only significant peaks or identified metabolites) by changing a setting in the QR settings menu. (2) The new dataset is requested and the view (get_quick_results_data(), a Python function/Django view) returns the data. (3) The new data is sent to FiRDI, (4) which empties the existing tables and populates them with the new data.

4.3.2 Metabolomics Analysis of MSCs Cultured on the Artificial ECMs

Metabolomes of MSCs were extracted at 7 and 14 d culture on the artificial ECMs and measured using LC-MS. The TIC chromatogram shown in Figure 4.10 shows high instrument reproducibility. 946 peaks were detected in the 7 d dataset and 1032 were detected at 14 d. A plot of the first two principal components of a principal components analysis (PCA) of peak intensities revealed little separation between ECMs at 7 d, but at 14 d there was slight clustering between PMA and PEA-based samples but not those with added growth factors. Using the QR application to analyse the data revealed that no peaks annotated by authentic chemical standards were significantly changing between ECMs at 7 or 14 d.

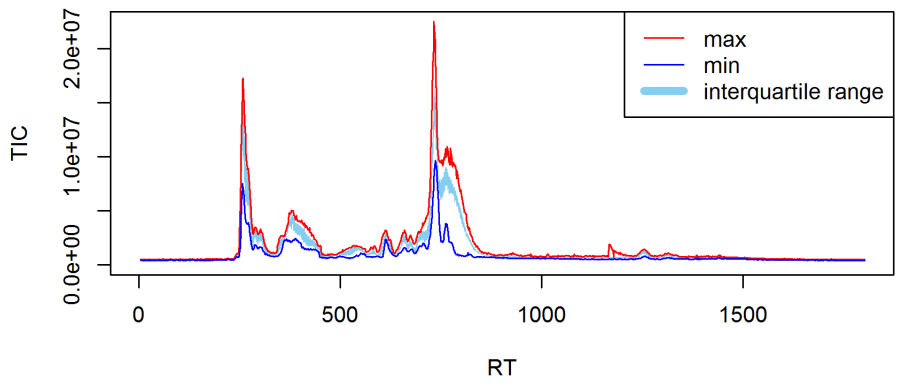

Figure 4.10 TIC statistics for positive mode pooled samples.

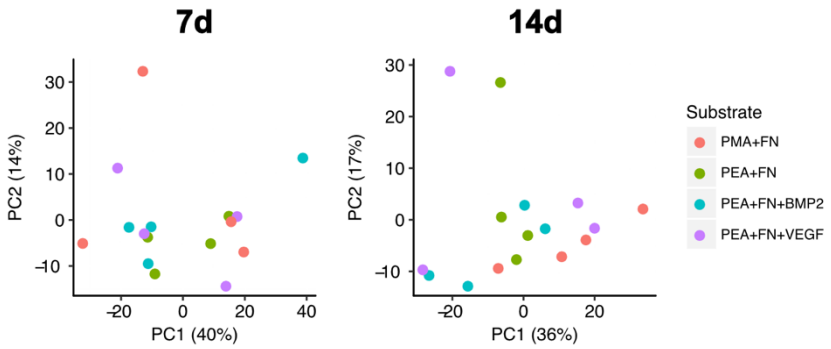

Figure 4.11 PCA of metabolomics data. At 7 d, there was little evidence of clustering by substrate, but at 14 d there is slightly more separation between samples cultured on PEA with FN and PMA with FN. The variance accounted for by the corresponding principal component is shown in brackets.

Instead of comparing individual metabolites, the pathways they belong to were compared. Small changes in individual compounds may be insignificant and therefore missed, but these small changes may amount to a large effect if they are consistent and grouped by a metabolic pathway (Tomfohr et al. 2005). A modified pathway-level analysis of gene expression (PLAGE) score that is part of the PiMP ecosystem (Karen McCluskey, Polyomics, University of Glasgow, unpublished) was used to compare the metabolomes of MSCs cultured on the artificial ECMs. This score combines the pathway activity score calculated by PLAGE of changes in compound formulae abundance with an overrepresentation test to compensate for the size of a pathway relative to how many compounds were detected in it. In general, the lower the modified pathway score, the greater the difference in activity between the experimental groups. At 7 d there did not appear to be much difference between the artificial ECMs (Figure 4.12), whereas at 14 d the activity of many pathways of MSCs cultured on PMA with FN compared to PEA with FN were different (Figure 4.13).

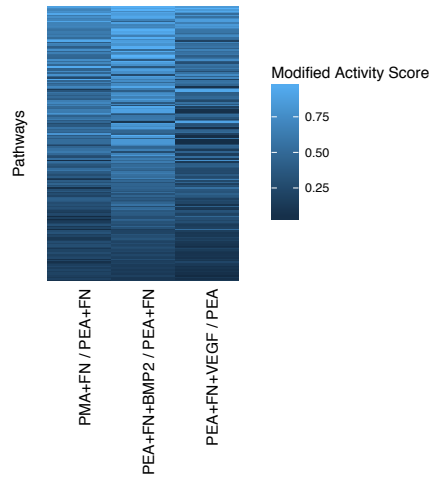

Figure 4.12 Comparison of pathway activity between MSCs cultured on the artificial ECMs at 7 d. Note that there was little difference between conditions. Darker/lower modified activity score indicates a larger difference in activity.

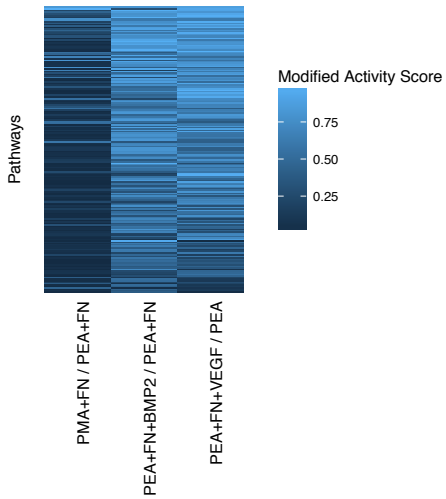

Figure 4.13 Comparison of pathway activity between MSCs cultured on the artificial ECMs at 14 d. Many pathways had different activity levels across the FN conformations (PMA + FN / PEA + FN), but few were different when comparing the growth factor based ECMs with PEA with FN. Darker/lower modified activity score indicates a larger difference in activity.

While no pathway activity met the threshold of significance at 7 d, several pathways had significant differences in activity at 14 d (Figure 4.14). These differences were largely observed between MSCs cultured on PEA with FN and those on PMA with FN, the two different FN conformations. Some of these pathways should be considered carefully as they are ostensibly unrelated to MSC or mammalian behaviour, such as bacterial chemotaxis, but, on the other hand, metabolites can be in several pathways. Nevertheless, the general trend was that there were several metabolic pathway differences based on FN

conformation, and the HMG-CoA reductase inhibitors pathway is different between PEA + FN and PEA + FN + VEGF.

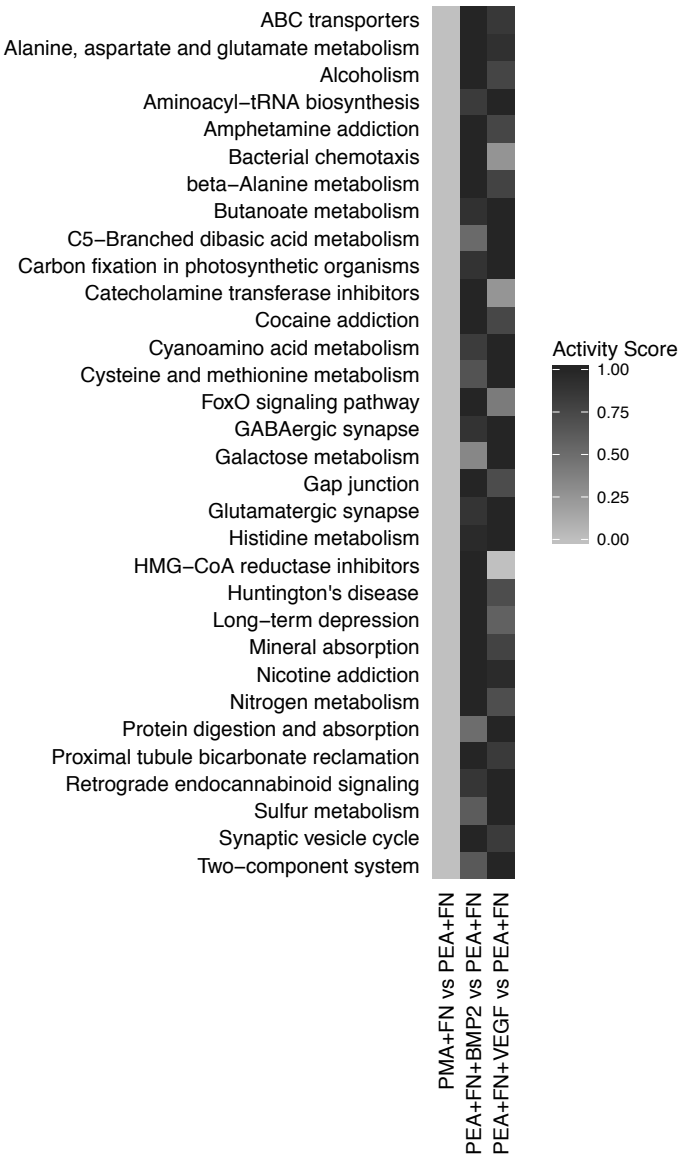

Figure 4.14 Pathways with different activity levels at 14 d as indicated by a modified activity score. Several pathways have a different activity depending on whether MSCs were cultured on PMA and PEA, whereas only the HMG-CoA reductase inhibitors pathway was different in cells cultured on PEA with FN and VEGF compared to PEA with FN. No pathway activity met the criteria for significance comparing PEA with FN and BMP2 to PEA with FN. A lower activity score (representing the modified activity score) indicates a greater difference in activity between the compared conditions.

4.4 Discussion

This chapter described the QR application extension to the PiMP ecosystem, which streamlines the visualisation and analysis of metabolomics data. Further, it was shown that metabolic pathway analysis of the LC-MS metabolomics data

suggested several pathways with different activities in MSCs that depended on the FN conformation of the underlying substrate.

Other approaches to profiling the global state of a cell include transcriptomics and proteomics. While metabolomics is comparatively newer, it is equally viewed as providing a global view into cell phenotype (Fiehn 2002). As well as its newness, metabolomics has shown to be a promising approach to improving the understanding of biomaterial-guided cell behaviour (Alakpa et al. 2016; McNamara et al. 2012). Together, these points justify the use of metabolomics in this study. Further, other omics techniques may complement metabolomics data and vice versa (Groen et al. 2016), providing grounds for future work.

The three main LC-MS metabolomics features — peaks, metabolites, and pathways — can be explored using PiMP, but doing so requires navigation to different parts of the application; following the flow of information from peak to pathway could be made easier. This is necessary in order to highlight the ambiguity in the list of metabolites and pathways output by such an analysis. The presence of a metabolite or pathway in the list does not guarantee that it is part of the true biological context of the system under investigation because of possible peak misannotation (Lu et al. 2017). By unifying these features of a PiMP analysis into a single QR webpage, the relationships among them are explicit and the underlying complexity clear, so that the user can make more informed decisions about what aspects of the dataset to focus on. Further, this consolidation can expedite data exploration.

In the process of developing the QR application several choices were made with respect to what features to include, and some features are still missing. Future work could focus on incorporating the modified pathway activity scores used above into the pathways table of the QR page. Further, other databases exist that contain biological information about compounds and metabolic pathways. Only KEGG pathways (Kanehisa et al. 2011) are currently used in PiMP, but for example, the metabolites and pathways in the human metabolome database (Wishart et al. 2017) could be used. With the field of metabolomics advancing at pace, the modular design of PiMP and the QR application enables the addition of new features.

The activity changes in metabolic pathways described here can be used to generate targets or hypotheses for further experiments that study how the artificial ECMs influence MSC behaviour. However, some of the highlighted KEGG metabolic pathways contain large numbers of metabolites and some are very broad, for example the aminoacyl tRNA biosynthesis and ATP-binding cassette (ABC) transporter pathways contain amino acids, and so will appear to have changes in activity (Booth et al. 2013), but could be false positive findings. With consideration of these points, several of these pathways can be related to previous studies of stem cells. Variation in the expression of ABC transporter genes distinguished MSCs from other ESCs (Barbet et al. 2012) and their expression appeared to identify adult stem cells in mice (Ramalho-Santos et al. 2002). Studies of the metabolomes of MSCs undergoing osteogenesis on nanotopographies made from bacterial inclusion bodies suggested that activity of the aminoacyl tRNA biosynthesis pathway did not change with differentiation (Seras-Franzoso et al. 2014). Beta-alanine appeared not to change after hypoxic pre-treatment of human MSCs implanted into mice, despite the treatment improving MSC retention and survival (Beegle et al. 2015). MSCs treated with sodium butyrate (a butanoate salt) underwent increased osteogenesis in OM compared to OM alone (Chen et al. 2007). Sodium butyrate is a histone deacetylase inhibitor, and given that nuclei measurements based on DAPI DNA staining contributed to the discrimination of which artificial ECM an MSC was cultured on (described in the previous chapter), a possible set of experiments is to explore the epigenome, or at least specific histone modifications (Jenuwein & Allis 2001; Strahl & Allis 2000), across the artificial ECMs; biomaterials are reported to be able to regulate stem cell epigenetics (Crowder et al. 2016). In all, this information could be used to inform the rational design of novel biomaterials for guiding MSC and stem cell fate for the purpose of tissue engineering and stem cell biology (McNamara et al. 2012).

5 General Discussion and Conclusions

5.1 Discussion

5.1.1 Artificial ECMs for MSC Control

The control of stem cells *in vitro* is important for understanding their biology and to exploit their potential for TERM. Biomaterials provide this control in the form of a tailored extracellular environment. For example, biomaterials have been customised to present variations in stiffness (Engler et al. 2006), topography (Dalby et al. 2007), and chemistry (Curran et al. 2005). The work presented here is the first characterisation of primary STRO-1$^+$ MSC behaviour on material-driven FN and growth factor matrices using high-content automated fluorescence microscopy and untargeted metabolomics.

In vivo, protein sequestration of growth factors and ECM constituents has a significant role in controlling stem cell behaviour, and so biomaterials that mimic or present ECM-inspired cues are particularly promising for *in vitro* stem cell guidance (Hettiaratchi et al. 2016). Synthetic polymers are a popular approach to control the assembly and patterning of ECM proteins *in vitro* because they can be biocompatible, customisable, and compliant with various manufacturing techniques to build complex substrates (Donnelly et al. 2017), and differences in their chemistry lead to modulation of adsorbed proteins (Krutty et al. 2016). For example, substrates covered with long polymer chains terminated with different chemical groups have been used to vary the pattern of FN adsorption and therefore cell behaviour (Araújo et al. 2016; Lin et al. 2015).

The most notable findings related to the artificial ECMs in this thesis were that the difference in FN conformation had a notable effect on the cytoskeletal structure, nuclei organisation, and metabolic activity of the cultured MSCs. Prior work demonstrating that, for either the same MSCs or a model mouse MSC cell line, PMA was generally associated with differentiation and PEA with maintenance of multipotency (González-García et al. 2012; Rico et al. 2016), is in line with these data. Several studies have also investigated the variation in substrate chemistry to control adsorbed FN conformation (Keselowsky et al. 2003; García et al. 1999; Mitsi et al. 2006; Phillips et al. 2010; Araújo et al.

2016; Klotzsch et al. 2014), but these have not produced such physiologically-similar FN matrices as those seen on PEA (Salmerón-Sánchez et al. 2011). Other ECM proteins are used in studies of controlling protein adsorption for controlling cell behaviour, such as vitronectin and collagen (Salasznyk et al. 2004), but FN is particularly well studied and has been used as a part of many biomaterials (Wilson et al. 2005; Zollinger & Smith 2017).

The addition of growth factors to the FN network conformation has been shown to enhance the activity of BMP2 and VEGF in a synergistic effect (Llopis-Hernández et al. 2016; Moulisová et al. 2017). BMP2 is associated with inducing osteogenesis in MSCs (Carreira et al. 2014), and VEGF is reported to enhance mineralization, proliferation, and survival in MSCs (Mayer et al. 2005; Pons et al. 2008). Other approaches to present these growth factors to cells that have had reported effects include, for example, artificially crosslinking BMP2 to FN (Hauff et al. 2015), or directly covalently immobilising VEGF to the substrate (Guex et al. 2014), but these approaches required chemical modification of the growth factors for linking to occur, whereas the ECMs used here rely on the natural binding properties of FN and its growth factor binding site for adsorption. With these studies in mind, especially those on the PEA-based FN network matrices with growth factors, it was hypothesised that the artificial ECMs with growth factors would also affect STRO-1$^+$ MSC differentiation. However, there was no consistent difference between the FN network with and without growth factors. The only indication of an effect was from the results of the image classification task, where MSCs cultured on the FN network with VEGF could be accurately predicted. There was no such discrimination when using BMP2 instead. However, these prior studies of the synergistic effect between the FN network and growth factors used different cell types (commercially-sourced MSCs and HUVECs) and measurement techniques, which may explain this difference. At least for topographies, it has been reported that there are cell type-specific responses to identical biomaterials (Biela et al. 2009).

With respect to using these artificial ECMs for TERM, using primary cells originating from the target tissue, as opposed to transformed cell lines or those that are processed using proprietary methods before procurement, is important so that results can be as similar as possible to the *in vivo* environment. The

human STRO-1$^+$ bone marrow cell population is believed to contain actual MSCs, but there are also other markers like STRO-1$^+$, such as CD146 (Sacchetti et al. 2007), NESTIN (Méndez-Ferrer et al. 2010; Isern et al. 2013), CD271 (Tormin et al. 2011), or VCAM1 (Gronthos et al. 2003), associated with the same claim (Bianco 2014), highlighting the lack of consensus on MSC identity (Lv et al. 2014). A downside of working with these specific primary cells is that they are obtained from donors undergoing surgery, meaning supply is limited and unpredictable, which forms a bottleneck in throughput. Small molecules discovered to accelerate the rate of MSC self-renewal while maintaining their multipotency may be a useful solution to this problem (Schugar et al. 2008).

Each substrate used in this work was produced one at a time, meaning that throughput was limited not only by the supply of cells but also by the rate of substrate production. Several studies have demonstrated the use of high-throughput approaches to quickly produce combinatorial arrays of biomaterial components, referred to as polymer microarrays. This has been used to produce arrays of different ECM proteins, synthetic polymers, artificial peptides, and small molecules in order to discover which combinations have a desirable effect on stem cell behaviour (Coyle et al. 2016). For example, arrays comprised of patches of varying concentrations of BMP2 and collagen to discover that which elicited the greatest degree of MSC osteogenesis (Ghaemi et al. 2016). A microarray of a variety of synthetic polymers was tested with the aim of determining which best maintained MSCs in long term culture (Duffy et al. 2014). The approach has even been used to make arrays of hydrogels with varying stiffness in order to examine the effect on MSC osteogenesis (Gobaa et al. 2011). An obvious improvement to the experimental methods used in this thesis is therefore to use polymer microarrays to rapidly increase the number of replicates and add the capacity for novel variants of the artificial ECMs, such as changes in protein concentrations or entirely new combinations of growth factors and proteins.

5.1.2 High-Content Biomaterial Analyses

Fluorescence image analysis is a keystone in many of the artificial ECM studies. It was also a predominant technique in this project, but it was used for the high-content analysis of automated microscopy images as opposed to manually

selected pictures used in these previous works. This was critical to ensure robustness of conclusions and for abrogating the bias associated with selecting regions of a sample to image. The availability of general purpose fluorescence image analysis packages like CellProfiler (Kamentsky et al. 2011) and Fiji (Schindelin et al. 2012) have been met with broad adoption (Caicedo et al. 2017) but they do not necessarily meet the requirements of every imaging task, entailing the use of a bespoke MATLAB and R pipeline for the analysis (Eliceiri et al. 2012). Nevertheless, the pipeline followed the general outline of a standard fluorescence image analysis procedure (Eliceiri et al. 2012).

The microscopy hardware used in this thesis, while automated, was a fast, single sample mosaic image solution. Though the software controlling the microscope allowed for the mosaic imaging of multiple samples at the same time, the set up did not support a predefined arrangement of positions to image. Samples had to be arbitrarily placed on the microscope stage for imaging, and for each use these positions had to be mapped using the software. This process was slow in comparison to using high-throughput microscopes, which are typically used to image each well of a tissue culture plate or each section of a regularly organised array of cell cultures (Pepperkok & Ellenberg 2006). Indeed, many polymer array studies make use of high-throughput fluorescence microscopes (Pepperkok & Ellenberg 2006), and the throughput of future work would benefit from this.

A limitation of using fluorescence microscopy was the false positive staining inherent in using certain antibodies and dyes. Actin was also not a perfect candidate for cell segmentation. These impediments became particularly clear when examining the segmentations produced by the MATLAB pipeline, for example in the form of DAPI-stained matter that was not relevant cellular material, or the high number of accurate rejections made during the filtering step, such as those that were incomplete segmentations. There are alternatives to using antibody-based labelling including label-free imaging such as by phase contrast microscopy (Meijering 2012) or whole homogenous cell staining such as a cytoplasmic dye like CellTracker (Reynolds et al. 2013). However, the advantage of using a cellular protein like actin is that, as well as being useful for segmentation, its provides other features that are relevant to cell behaviour on biomaterials, such as its distribution. The alternative suggestions only provide

morphological information. There is a balance between the extra information afforded by a channel containing an informative marker and one dedicated solely to improving segmentation performance.

High-content fluorescence image analysis is not the only high-information option for directly studying cellular interactions with biomaterials. For example, focal-plane array FTIR imaging enables rapid IR measurement over a large area (Heraud & Tobin 2009), and this has been used to profile sizeable regions (approximately 250 µm x 250 µm at 4 µm resolution) of MSCs undergoing differentiation in various culture conditions (Krafft et al. 2007). FTIR is label-free and produces data that can be processed using techniques similar to the multivariate analyses used in this thesis. Equally, there are several high-throughput measurement techniques to analyse the biomaterials themselves (Urquhart et al. 2007). Time-of-flight secondary ionisation mass spectrometry is a tested approach for profiling the protein composition of biomaterials and polymer arrays (Mei et al. 2010; Ghaemi et al. 2016). Together, the relationship between material properties and cell behaviour can be rapidly mapped using high information content approaches and multivariate statistics.

It should be noted that the performance metrics for the image classification task were generally low, with precision and recall rarely ascending above 50 %. The classifier performed similarly for all classes overall, with no one class particularly more successful or worse than the others. Classification is not necessarily an easy task, and will be particularly hampered by variability within the classes of the training data. Many reasons could be suggested to explain this variability. Perhaps only a subset of the MSCs have the capacity to respond, or respond with a similar phenotypic change, to the artificial ECMs. This is supported by the previously mentioned uncertainty about the true *in vivo* identity of MSCs and over the efficiency of markers for enriching for MSCs — not all STRO-1$^+$ cells have differentiation potential (Stewart et al. 1999; Gronthos et al. 1994). Even with populations of cells derived from a single ancestor there can be substantial variability within gene and protein expression (Cote et al. 2016; Elowitz et al. 2002). There is also likely flexibility in the range of expression values that underlie a given phenotype (Huang 2010), that may make classification difficult. Further, the changes that occur needed to be measured

with the chosen fluorescence markers. It could be that there were ECM-wide phenotypic changes occurring, but those phenotypic markers were not labelled.

For some of the image processing results, the classification results did not always tally with those of the inference task. Classification works by computing a relationship between image features but inference compares the features in isolation. The relationships may be more informative for explaining differences in MSC phenotype than any given feature alone. The limitation of using classification is that the model chosen to represent the relationship is difficult to comprehend with respect to the real-world biology. A logistic regression was chosen as the model since it is one of the simpler choices, and the coefficients of the model can be interpreted in the context of the features measured. Random forests, support vector machines, and neural networks were also tested, which are more complex but flexible models that may have better performance, yet there was no substantial difference in their capabilities for these datasets compared to the logistic regression (neural networks tested by Vihang Godbole, Glasgow Polyomics, University of Glasgow).

5.1.3 Metabolomics and Metabolomics Data Analysis

Metabolomics has been used to identify small molecules related to the effects of biomaterials upon stem cells that could be used to control their behaviour without biomaterials or improve understanding of the mechanisms behind differentiation (McNamara et al. 2012). Doing so follows the approach of identifying individually significantly changing metabolites related to a desirable biomaterial effect, such as maintaining multipotency. For example, untargeted LC-MS metabolomics was used to identify changes in lipids that accompanied MSC bone or chondrogenic differentiation on hydrogels of varying stiffness (Alakpa et al. 2016). Unfortunately, no single metabolite that was matched to an authentic standard was changing significantly in this work. Perhaps at earlier time points significant changes would be observed, as sharp changes in individual metabolites could be evened out by flux with time. This could conceivably manifest as the observed changes in pathway activity, which absorb the changes as part of homeostasis. Admittedly, the cited identification of lipids associated with MSC differentiation was performed at 1 and 4 weeks, so this theory may not hold true in all cases. The low numbers of metabolomics replicates used in this

thesis may also underlie the inability to make statistically backed inferences about metabolite changes.

The interpretation of untargeted LC-MS metabolomics data is complicated by the complexity of the underlying metabolite identification process. The QR web application enables exploration of the metabolites in an LC-MS metabolomics dataset while visualising this complexity in a straightforward manner. Coupled with the information and infrastructure already provided by PiMP, such as the pathways and comparisons between conditions, the QR page is expected to replace the current user interface for PiMP, after some of the changes discussed in the relevant section have been made. Further, FiRDI provides a means of visualising related data in a web page while providing functionality to interactively query the relations of a data point of interest. In bioinformatics, this may be useful for such areas as proteomics by visualising measured peptides, the proteins they are believed to represent, and perhaps the functions they are associated with (Mallick & Kuster 2010). More generally, many data sets can be interpreted under a relational model, for which FiRDI may be of use.

Alternative web-based metabolomics pipelines to PiMP exist, such as XCMS online (Tautenhahn et al. 2012; Gowda et al. 2014), workflow4metabolomics (Giacomoni et al. 2015), and MetaboAnalyst (Xia et al. 2015), but they either are not streamlined or do not offer the same level of functionality and biological context (Gloaguen et al. 2017; Spicer et al. 2017). Further, given that PiMP forms part of the provision of metabolomics data by Glasgow Polyomics, it was the best pipeline to extend with the QR application. Nevertheless, since the relevant functionality has been abstracted as FiRDI, it could be used in other web-based applications. While stand-alone applications for metabolomics (for a review, see Booth et al. 2013) may be less user-friendly because of set up, installation, and hardware requirements, the concepts that are embodied by the QR application could be readily adopted.

There were certain compromises between complexity, flexibility, and usability in the QR application. For example, it provides an option to filter out non-statistically significantly changing metabolites/peaks, but this was implemented at the cost of flexibility: users are restricted to using the standard multiple correction p-value threshold of 0.05 (Altman & Krzywinski 2016). This may not

be a problem, but some users may wish to have more control. For those focussed on pathway analysis, this may be of no consequence anyway. It is anticipated that the pathway-level analysis tools used in this thesis will be integrated into the QR application, and that changes will be made to allow custom statistical thresholding of both pathways and metabolites/peaks. With this in mind, the QR application has received positive feedback from users.

5.1.4 General Biological Discussion

The results of this thesis did not persuasively demonstrate that the artificial ECMs induced osteogenesis in the cultured MSCs, although control experiments did indicate that these cells were capable of differentiation. In particular, the lack of changes in osteogenic markers, as measured by ICW and image analysis, suggested this. Furthermore, no individual identified metabolites were significantly changing between conditions, but changes were anticipated because of the relationship between biomaterials and changes in the metabolome in stem cells (McNamara et al. 2015; Alakpa et al. 2016; Agathocleous & Harris 2013), and the coverage of many common pathways by the list of potential identified metabolites (Glasgow Polyomics, University of Glasgow). Other phenotypic differences were indicated in the image analysis classification, image analysis inference, and metabolomics sections, such as the relatively successful classification, observed differences in nuclei features, and changes in pathway activity of MSCs cultured on PMA with FN. These results support the argument that the different FN conformations alter MSC phenotype (González-García et al. 2012) or other cell types (Rico et al. 2016; Moulisová et al. 2017; Salmerón-Sánchez et al. 2011), but the evidence presented herein was limited by the poor image classification performance, lack of individually changing metabolites, and little discrimination of the MSCs cultured on the corresponding PEA with FN substrate. The addition of growth factors to the FN network conformation did not convincingly lead to phenotypic changes such as osteogenesis, in contrast to previous studies (Llopis-Hernández et al 2016; Moulisová et al. 2017). It was expected that the techniques used here would have shown changes in MSCs cultured on these surfaces because of these works. It is possible that differences in cell source and cell culture conditions to the published works underpin these conclusions.

5.2 Future Work

FN has binding activity for a range of growth factors beyond BMP2 and VEGF that have roles in the regulation of a variety of cells (Martino & Hubbell 2010; Rodrigues et al. 2010), and it would be interesting whether the effect of these or combinations thereof would be enhanced by being presented using the PEA-based FN network conformations. This future work would particularly benefit from high-throughput biomaterial manufacturing techniques and therefore high-throughput analysis techniques.

The metabolomes of MSCs on the artificial ECMs were investigated in this thesis, but there are other omics technologies that provide a similarly detailed insight into cell behaviour. Further, it is believed that by integrating omics data sets, particularly transcriptomics with metabolomics data, more information about the biology of the system can be obtained (Groen et al. 2016), and recent developments in PiMP have enabled the analysis of metabolomics datasets in the context of genome-scale metabolic networks (the prediction of metabolic networks using transcriptome data) using MetExplore (Cottret et al. 2010). A greater understanding of the biology of MSCs cultured on these artificial ECMs may come from studies using multiple omics technologies.

Many metabolic pathways appeared to be changing in MSCs between the different FN conformations, and with more time these could be studied in detail. It is important to work out which pathways are already associated with stem cell metabolic activity, which are irrelevant, and which of these are novel findings. When viable pathways are identified, it is important to modulate their activity, perhaps using stimulating or inhibiting small molecules targeted against certain members, to test the findings and establish whether they would be useful candidates as composites of new biomaterials, or as alternatives to regulate stem cell self-renewal and differentiation for TERM (Schugar et al. 2008).

The process of mapping from measured molecules to actual biological knowledge is common in omics technologies, for example, from DNA fragment to the gene it was derived from, or from a peptide to the protein it comprised. Together this knowledge is more informative than the list of putative proteins or genes alone, and sometimes is critical for interpretation. Other web-enabled pipelines for

analysing or visualising omics data may benefit from the functionality offered by FiRDI in the future.

5.3 Conclusion

In this thesis, human bone marrow STRO-1$^+$ MSCs were characterised on artificial ECMs comprised of different FN conformations with added growth factors. This revealed substrate-dependent changes in MSC behaviour, highlighting the potential for modulating protein conformation in order to control cell behaviour. During this process, new techniques were developed or demonstrated to improve existing approaches to analyse cells on these biomaterials, namely using automated microscopy and high-content image analyses, and using metabolomics. Further, as a result of needing to improve the analysis of untargeted LC-MS metabolomics data, a metabolomics data visualisation tool was developed, which also has further reaching implications for the general visualisation and interaction of related data.

List of References

Aarts, E. et al., 2014. A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience*, 17(4), pp.491–496.

Agathocleous, M. & Harris, W.A., 2013. Metabolism in physiological cell proliferation and differentiation. *Trends in Cell Biology*, 23(10), pp.484–492.

Aguilar, H.N. et al., 2010. Quantification of rapid Myosin regulatory light chain phosphorylation using high-throughput in-cell Western assays: comparison to Western immunoblots. A. Lucia, ed. *PLoS ONE*, 5(4), p.e9965.

Alakpa, E.V. et al., 2016. Tunable Supramolecular Hydrogels for Selection of Lineage-Guiding Metabolites in Stem Cell Cultures. *Chem*, 1(2), pp.298–319.

Altman, N. & Krzywinski, M., 2016. Points of significance: P values and the search for significance. *Nature Methods*, 14(1), pp.3–4.

Araújo, A.R. et al., 2016. Surfaces Mimicking Glycosaminoglycans Trigger Different Response of Stem Cells via Distinct Fibronectin Adsorption and Reorganization. *ACS Applied Materials & Interfaces*, 8(42), pp.28428–28436.

Aretz, I. & Meierhofer, D., 2016. Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology. *International Journal of Molecular Sciences*, 17(5), p.632.

Atherton, P. et al., 2016. Mechanosensitive components of integrin adhesions: Role of vinculin. *Experimental Cell Research*, 343(1), pp.21–27.

Baker, M.J. et al., 2014. Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols*, 9(8), pp.1771–1791.

Baksh, D., Song, L. & Tuan, R.S., 2004. Adult mesenchymal stem cells: characterization, differentiation, and application in cell and gene therapy. *Journal of Cellular and Molecular Medicine*, 8(3), pp.301–316.

Baneyx, G., Baugh, L. & Vogel, V., 2002. Fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), pp.5139–5143.

Barbet, R. et al., 2012. Expression of the 49 human ATP binding cassette (ABC) genes in pluripotent embryonic stem cells and in early- and late-stage multipotent mesenchymal stem cells: possible role of ABC plasma membrane transporters in maintaining human stem cell pluripotency. *Cell Cycle*, 11(8), pp.1611–1620.

Barczyk, M., Carracedo, S. & Gullberg, D., 2010. Integrins. *Cell and Tissue Research*, 339(1), pp.269–280.

Barth, A., 2007. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta*, 1767(9), pp.1073–1101.

Bassan, P. et al., 2010. RMieS-EMSC correction for infrared spectra of biological

cells: Extension using full Mie theory and GPU computing. *Journal of Biophotonics*, 3(8-9), pp.609–620.

Bates, D. et al., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), pp.1–48.

Baujard-Lamotte, L. et al., 2008. Kinetics of conformational changes of fibronectin adsorbed onto model surfaces. *Colloids and Surfaces B: Biointerfaces*, 63(1), pp.129–137.

Bayat, A., 2002. Science, medicine, and the future: Bioinformatics. *BMJ (Clinical research ed.)*, 324(7344), pp.1018–1022.

Beegle, J. et al., 2015. Hypoxic preconditioning of mesenchymal stromal cells induces metabolic changes, enhances survival, and promotes cell retention in vivo. *Stem Cells*, 33(6), pp.1818–1828.

Bellows, C.G., Aubin, J.E. & Heersche, J.N., 1991. Initiation and progression of mineralization of bone nodules formed in vitro: the role of alkaline phosphatase and organic phosphate. *Bone and Mineral*, 14(1), pp.27–40.

Bianco, P., 2014. "Mesenchymal" stem cells. *Current Opinion in Structural Biology*, 30(1), pp.677–704.

Bianco, P. & Robey, P.G., 2001. Stem cells in tissue engineering. *Nature*, 414(6859), pp.118–121.

Biela, S.A. et al., 2009. Different sensitivity of human endothelial cells, smooth muscle cells and fibroblasts to topography in the nano-micro range. *Acta Biomaterialia*, 5(7), pp.2460–2466.

Bilem, I., Chevallier, P., et al., 2017a. Interplay of Geometric Cues and RGD/BMP-2 Crosstalk in Directing Stem Cell Fate. *ACS Biomaterials Science & Engineering*, 3(10), pp.2514–2523.

Bilem, I., Plawinski, L., et al., 2017. The spatial patterning of RGD and BMP-2 mimetic peptides at the subcellular scale modulates human mesenchymal stem cells osteogenesis. *Journal of Biomedical Materials Research - Part A*.

Bissels, U., Eckardt, D. & Bosio, A., 2013. Characterization and Classification of Stem Cells. In G. Steinhoff, ed. *Regenerative Medicine: From Protocol to Patient*. Regenerative Medicine: From Protocol to Patient. Dordrecht: Springer Netherlands, pp. 155–176.

Booth, S.C., Weljie, A.M. & Turner, R.J., 2013. Computational tools for the secondary analysis of metabolomics experiments. *Computational and Structural Biotechnology Journal*, 4(5), p.e201301003.

Burridge, K. & Chrzanowska-Wodnicka, M., 1996. Focal adhesions, contractility, and signaling. *Annual Review of Cell and Developmental Biology*, 12(1), pp.463–518.

Burridge, K. et al., 1988. Focal adhesions: transmembrane junctions between the extracellular matrix and the cytoskeleton. *Annual Review of Cell*

Biology, 4(1), pp.487–525.

Caicedo, J.C. et al., 2017. Data-analysis strategies for image-based cell profiling. *Nature Methods*, 14(9), pp.849–863.

Carisey, A. & Ballestrem, C., 2011. Vinculin, an adapter protein in control of cell adhesion signalling. *European Journal of Cell Biology*, 90(2-3), pp.157–163.

Carpenter, A.E. et al., 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biololgy*, 7(10), p.R100.

Carpenter, A.E., Kamentsky, L. & Eliceiri, K.W., 2012. A call for bioimaging software usability. *Nature Methods*, 9(7), pp.666–670.

Carreira, A.C. et al., 2014. Bone Morphogenetic Proteins: structure, biological function and therapeutic applications. *Archives of Biochemistry and Biophysics*, 561, pp.64–73.

Cassimeris, L. et al., 2012. Fueled by microtubules: does tubulin dimer/polymer partitioning regulate intracellular metabolism? *Cytoskeleton*, 69(3), pp.133–143.

Cavalcanti-Adam, E.A. et al., 2008. Cell adhesion and response to synthetic nanopatterned environments by steering receptor clustering and spatial location. *HFSP journal*, 2(5), pp.276–285.

Cavalcanti-Adam, E.A. et al., 2007. Cell spreading and focal adhesion dynamics are regulated by spacing of integrin ligands. *Biophysical Journal*, 92(8), pp.2964–2974.

Chan, C.K.F. et al., 2015. Identification and specification of the mouse skeletal stem cell. *Cell*, 160(1-2), pp.285–298.

Chazotte, B., 2011. Labeling nuclear DNA using DAPI. *Cold Spring Harbor protocols*, 2011(1), p.pdb.prot5556.

Chen, T.-H. et al., 2007. Sodium butyrate activates ERK to regulate differentiation of mesenchymal stem cells. *Biochemical and Biophysical Research Communications*, 355(4), pp.913–918.

Chonanant, C. et al., 2011. Characterisation of chondrogenic differentiation of human mesenchymal stem cells using synchrotron FTIR microspectroscopy. *Analyst*, 136(12), pp.2542–2551.

Cinque, G. et al., 2016. World First for Diamond in Synchrotron-Based IR Photothermal Nanospectroscopy. *Synchrotron Radiation News*, 29(4), pp.37–39.

Cock, P.J.A. et al., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), pp.1422–1423.

Cote, A.J. et al., 2016. Single-cell differences in matrix gene expression do not predict matrix deposition. *Nature Communications*, 7, p.10865.

Cottret, L. et al., 2010. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Research*, 38(Web Server issue), pp.W132–7.

Coyer, S.R. et al., 2012. Nanopatterning reveals an ECM area threshold for focal adhesion assembly and force transmission that is regulated by integrin activation and cytoskeleton tension. *Journal of Cell Science*, 125(Pt 21), pp.5110–5123.

Coyle, R., Jia, J. & Mei, Y., 2016. Polymer microarray technology for stem cell engineering. *Acta Biomaterialia*, 34, pp.60–72.

Crowder, S.W. et al., 2016. Material Cues as Potent Regulators of Epigenetics and Stem Cell Function. *Cell Stem Cell*, 18(1), pp.39–52.

Curran, J.M. et al., 2010. Introducing dip pen nanolithography as a tool for controlling stem cell behaviour: unlocking the potential of the next generation of smart materials in regenerative medicine. *Lab on a Chip*, 10(13), pp.1662–1670.

Curran, J.M., Chen, R. & Hunt, J.A., 2005. Controlling the phenotype and function of mesenchymal stem cells in vitro by adhesion to silane-modified clean glass surfaces. *Biomaterials*, 26(34), pp.7057–7067.

Curran, J.M., Chen, R. & Hunt, J.A., 2006. The guidance of human mesenchymal stem cell differentiation in vitro by controlled modifications to the cell substrate. *Biomaterials*, 27(27), pp.4783–4793.

Dalby, M.J. et al., 2007. The control of human mesenchymal cell differentiation using nanoscale symmetry and disorder. *Nature Materials*, 6(12), pp.997–1003.

Ding, L. et al., 2012. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature*, 481(7382), pp.457–462.

Dingal, P.C.D.P. & Discher, D.E., 2014. Combining insoluble and soluble factors to steer stem cell fate. *Nature Materials*, 13(6), pp.532–537.

Discher, D.E., Mooney, D.J. & Zandstra, P.W., 2009. Growth factors, matrices, and forces combine and control stem cells. *Science*, 324(5935), pp.1673–1677.

Dominguez, R. & Holmes, K.C., 2011. Actin structure and function. *Annual Review of Biophysics*, 40(1), pp.169–186.

Dominici, M. et al., 2006. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy*, 8(4), pp.315–317.

Donaldson, P.M. et al., 2016. Broadband near-field infrared spectromicroscopy using photothermal probes and synchrotron radiation. *Optics Express*, 24(3), pp.1852–1864.

Donnelly, H. et al., 2017. Current approaches for modulation of the nanoscale

interface in the regulation of cell behavior. *Nanomedicine: nanotechnology, biology, and medicine*. In press. 10.1016/j.nano.2017.03.020

Ducy, P., 2000. Cbfa1: a molecular switch in osteoblast biology. *Developmental Dynamics*, 219(4), pp.461–471.

Duffy, C.R.E. et al., 2014. A high-throughput polymer microarray approach for identifying defined substrates for mesenchymal stem cells. *Biomaterials Science*, 2(11), pp.1683–1692.

Dunn, W.B. et al., 2013. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, 9(1), pp.44–66.

Dutcher, S.K., 2001. The tubulin fraternity: alpha to eta. *Current Opinion in Cell Biology*, 13(1), pp.49–54.

Eliceiri, K.W. et al., 2012. Biological imaging software tools. *Nature Methods*, 9(7), pp.697–710.

Elowitz, M.B. et al., 2002. Stochastic gene expression in a single cell. *Science*, 297(5584), pp.1183–1186.

Engler, A.J. et al., 2006. Matrix elasticity directs stem cell lineage specification. *Cell*, 126(4), pp.677–689.

Fiehn, O., 2002. Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, 48(1-2), pp.155–171.

Fisher, M.B. & Mauck, R.L., 2013. Tissue engineering and regenerative medicine: recent innovations and the transition to translation. *Tissue Engineering. Part B, Reviews*, 19(1), pp.1–13.

Fraioli, R. et al., 2016. Surface guidance of stem cell behavior: Chemically tailored co-presentation of integrin-binding peptides stimulates osteogenic differentiation in vitro and bone formation in vivo. *Acta Biomaterialia*, 43, pp.269–281.

García, A.J., Vega, M.D. & Boettiger, D., 1999. Modulation of cell proliferation and differentiation through substrate-dependent changes in fibronectin conformation. *Molecular Biology of the Cell*, 10(3), pp.785–798.

Geiger, B. et al., 2001. Transmembrane crosstalk between the extracellular matrix--cytoskeleton crosstalk. *Nature Reviews Molecular Cell Biology*, 2(11), pp.793–805.

Geiger, B., Spatz, J.P. & Bershadsky, A.D., 2009. Environmental sensing through focal adhesions. *Nature Reviews Molecular Cell Biology*, 10(1), pp.21–33.

Gelman, A. & Hill, J., 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Ghaemi, S. et al., 2016. Synergistic influence of collagen I and BMP 2 drives

osteogenic differentiation of mesenchymal stem cells: A cell microarray analysis. *Acta Biomaterialia*, 34, pp.41–52.

Giacomoni, F. et al., 2015. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 31(9), pp.1493–1495.

Glassford, S.E., Byrne, B. & Kazarian, S.G., 2013. Recent applications of ATR FTIR spectroscopy and imaging to proteins. *Biochimica et Biophysica Acta*, 1834(12), pp.2849–2858.

Gloaguen, Y. et al., 2017. PiMP my metabolome: An integrated, web-based tool for LC-MS metabolomics data. *Bioinformatics*.

Gobaa, S. et al., 2011. Artificial niche microarrays for probing single stem cell fate in high throughput. *Nature Methods*, 8(11), pp.949–955.

Goecks, J., Nekrutenko, A. & Taylor, J., 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8), p.R86.

Gonzalez, R., Woods, R. & Eddins, S., 2009. *Digital Image Processing Using MATLAB*, Tata McGraw-Hill Education.

Gonzalez, R.C. & Woods, R.E., 2010. *Digital Image Processing*, Addison-Wesley.

González-García, C. et al., 2012. Surface mobility regulates skeletal stem cell differentiation. *Integrative Biololgy*, 4(5), pp.531–539.

Goodwin, S., McPherson, J.D. & McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), pp.333–351.

Gowda, H. et al., 2014. Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Analytical Chemistry*, 86(14), pp.6931–6939.

Groen, N. et al., 2016. Stepping into the omics era: Opportunities and challenges for biomaterials science and engineering. *Acta Biomaterialia*, 34, pp.133–142.

Gronthos, S. et al., 1999. Differential cell surface expression of the STRO-1 and alkaline phosphatase antigens on discrete developmental stages in primary cultures of human bone cells. *Journal of Bone and Mineral Research*, 14(1), pp.47–56.

Gronthos, S. et al., 2003. Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow. *Journal of Cell Science*, 116(Pt 9), pp.1827–1835.

Gronthos, S. et al., 1994. The STRO-1+ fraction of adult human bone marrow contains the osteogenic precursors. *Blood*, 84(12), pp.4164–4173.

Guerra, N.B. et al., 2010. Subtle variations in polymer chemistry modulate substrate stiffness and fibronectin activity. *Soft Matter*, 6(19), pp.4748–8.

Guex, A.G. et al., 2014. Covalent immobilisation of VEGF on plasma-coated electrospun scaffolds for tissue engineering applications. *Colloids and Surfaces B: Biointerfaces*, 123, pp.724–733.

Hagen, J.B., 2000. The origins of bioinformatics. *Nature Reviews Genetics*, 1(3), pp.231–236.

Haralick, R.M., Shanmugam, K. & Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), pp.610–621.

Hauff, K. et al., 2015. Matrix-Immobilized BMP-2 on Microcontact Printed Fibronectin as an in vitro Tool to Study BMP-Mediated Signaling and Cell Migration. *Frontiers in Bioengineering and Biotechnology*, 3, p.62.

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

Hettiaratchi, M.H., Guldberg, R.E. & McDevitt, T.C., 2016. Biomaterial strategies for controlling stem cell fate via morphogen sequestration. *Journal of Materials Chemistry B*, 4(20), pp.3464–3481.

Horgan, R.P. & Kenny, L.C., 2011. "Omic" technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 13(3), pp.189–195.

Horton, E.R. et al., 2015. Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. *Nature Cell Biology*, 17(12), pp.1577–1587.

Hothorn, T., Bretz, F. & Westfall, P., 2008. Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), pp.346–363.

Huang, S., 2010. Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules. *PLoS Biology*, 8(5), p.e1000380.

Hubbard, B. et al., 2016. Fibronectin Fiber Extension Decreases Cell Spreading and Migration. *Journal of Cellular Physiology*, 231(8), pp.1728–1736.

Hulshof, F.F.B. et al., 2017. Mining for osteogenic surface topographies: In silico design to in vivo osseo-integration. *Biomaterials*, 137, pp.49–60.

Humphrey, J.D., Dufresne, E.R. & Schwartz, M.A., 2014. Mechanotransduction and extracellular matrix homeostasis. *Nature Reviews Molecular Cell Biology*, 15(12), pp.802–812.

Humphries, J.D., Byron, A. & Humphries, M.J., 2006. Integrin ligands at a glance. *Journal of Cell Science*, 119(Pt 19), pp.3901–3903.

Hynes, R.O., 2002. Integrins: bidirectional, allosteric signaling machines. *Cell*, 110(6), pp.673–687.

Isern, J. et al., 2013. Self-renewing human bone marrow mesenspheres promote hematopoietic stem cell expansion. *Cell Reports*, 3(5), pp.1714–1724.

Javazon, E.H., Beggs, K.J. & Flake, A.W., 2004. Mesenchymal stem cells: paradoxes of passaging. *Experimental Hematology*, 32(5), pp.414–425.

Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science*, 293(5532), pp.1074–1080.

Jones, T.R. et al., 2009. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp.1826–1831.

Kamentsky, L., Jones, T.R., Fraser, A., Bray, M.-A., Logan, D.J., Madden, K.L., Ljosa, V., Rueden, C., Eliceiri, K.W. & Carpenter, A.E., 2011. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics*, 27(8), pp.1179–1180.

Kanchanawong, P. et al., 2010. Nanoscale architecture of integrin-based cell adhesions. *Nature*, 468(7323), pp.580–584.

Kanehisa, M. et al., 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research, 40, pp.D109-D114* .

Kapuscinski, J., 1995. DAPI: a DNA-specific fluorescent probe. *Biotechnic & Histochemistry*, 70(5), pp.220–233.

Karageorgiou, V. et al., 2004. Bone morphogenetic protein-2 decorated silk fibroin films induce osteogenic differentiation of human bone marrow stromal cells. *Journal of Biomedical Materials Research - Part A*, 71(3), pp.528–537.

Keselowsky, B.G., Collard, D.M. & García, A.J., 2003. Surface chemistry modulates fibronectin conformation and directs integrin binding and specificity to control cell adhesion. *Journal of Biomedical Materials Research - Part A*, 66(2), pp.247–259.

Kilian, K.A. et al., 2010. Geometric cues for directing the differentiation of mesenchymal stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), pp.4872–4877.

Kim, S.-H., Turnbull, J. & Guimond, S., 2011. Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. *The Journal of Endocrinology*, 209(2), pp.139–151.

Kinney, M.A. & McDevitt, T.C., 2013. Emerging strategies for spatiotemporal control of stem cell fate and morphogenesis. *Trends in Biotechnology*, 31(2), pp.78–84.

Klotzsch, E. et al., 2014. Conformational distribution of surface-adsorbed fibronectin molecules explored by single molecule localization microscopy. *Biomaterials Science*, 2(6), pp.883–892.

Knobloch, M. et al., 2013. Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis. *Nature*, 493(7431), pp.226–230.

Komori, T., 2010. Regulation of bone development and extracellular matrix protein genes by RUNX2. *Cell and Tissue Research*, 339(1), pp.189–195.

Krafft, C. et al., 2007. Differentiation of individual human mesenchymal stem cells probed by FTIR microscopic imaging. *Analyst*, 132(7), pp.647–653.

Krutty, J.D. et al., 2016. Surface functionalization and dynamics of polymeric cell culture substrates. *Current Opinion in Biotechnology*, 40, pp.164–169.

Krzywinski, M., Altman, N. & Blainey, P., 2014. Points of significance: nested designs. *Nature Methods*, 11(10), pp.977–978.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), pp.1–26.

Kulterer, B. et al., 2007. Gene expression profiling of human mesenchymal stem cells derived from bone marrow during expansion and osteoblast differentiation. *BMC Genomics*, 8(1), p.70.

Lanza, R., Langer, R. & Vacanti, J.P., 2013. *Principles of Tissue Engineering*, Academic Press.

Lever, J., Krzywinski, M. & Altman, N., 2016. Points of Significance: Logistic regression. *Nature Methods*, 13(7), pp.541–542.

Li, S., Guan, J.-L. & Chien, S., 2005. Biochemistry and biomechanics of cell motility. *Annual Review of Biomedical Engineering*, 7(1), pp.105–150.

Lian, J.B. & Stein, G.S., 1992. Concepts of osteoblast growth and differentiation: basis for modulation of bone cell development and tissue formation. *Critical Reviews in Oral Biology and Medicine*, 3(3), pp.269–305.

Lin, F. et al., 2011. Fibronectin growth factor-binding domains are required for fibroblast survival. *The Journal of Investigative Dermatology*, 131(1), pp.84–98.

Lin, M. et al., 2015. Adsorption force of fibronectin on various surface chemistries and its vital role in osteoblast adhesion. *Biomacromolecules*, 16(3), pp.973–984.

Liu et al., 2016. Synchrotron FTIR microspectroscopy reveals early adipogenic differentiation of human mesenchymal stem cells at single-cell level. *Biochemical and Biophysical Research Communications*, 478(3), pp.1286–1291.

Liu, E. et al., 2009. High-content profiling of cell responsiveness to graded substrates based on combinyatorially variant polymers. *Combinatorial Chemistry & High Throughput Screening*, 12(7), pp.646–655.

Liu, E. et al., 2010. Parsing the early cytoskeletal and nuclear organizational cues that demarcate stem cell lineages. *Cell Cycle*, 9(11), pp.2108–2117.

Ljosa, V. & Carpenter, A.E., 2009. Introduction to the quantitative analysis of two-dimensional fluorescence microscopy images for cell-based screening. *PLoS Computational Biology*, 5(12), p.e1000603.

Llopis-Hernández, V. et al., 2013. Role of material-driven fibronectin fibrillogenesis in protein remodeling. *BioResearch Open Access*, 2(5), pp.364–373.

Llopis-Hernández, V. et al., 2016. Material-driven fibronectin assembly for high-efficiency presentation of growth factors. *Science Advances*, 2(8), p.e1600188.

Lu, W. et al., 2017. Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annual Review of Biochemistry*, 86, pp.277–304.

Lutolf, M.P. & Hubbell, J.A., 2005. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nature Biotechnology*, 23(1), pp.47–55.

Lutolf, M.P., Gilbert, P.M. & Blau, H.M., 2009. Designing materials to direct stem-cell fate. *Nature*, 462(7272), pp.433–441.

Lv, F.-J. et al., 2014. Concise review: the surface markers and identity of human mesenchymal stem cells. *Stem Cells*, 32(6), pp.1408–1419.

Mallick, P. & Kuster, B., 2010. Proteomics: a pragmatic perspective. *Nature Biotechnology*, 28(7), pp.695–709.

Marklein, R.A. et al., 2016. High Content Imaging of Early Morphological Signatures Predicts Long Term Mineralization Capacity of Human Mesenchymal Stem Cells upon Osteogenic Induction. *Stem Cells*, 34(4), pp.935–947.

Martino, M.M. & Hubbell, J.A., 2010. The 12th-14th type III repeats of fibronectin function as a highly promiscuous growth factor-binding domain. *FASEB Journal*, 24(12), pp.4711–4721.

Martino, M.M. et al., 2011. Engineering the growth factor microenvironment with fibronectin domains to promote wound and bone tissue healing. *Science Translational Medicine*, 3(100), pp.100ra89–100ra89.

Martino, M.M. et al., 2015. Extracellular matrix-inspired growth factor delivery systems for bone regeneration. *Advanced Drug Delivery Reviews*, 94, pp.41–52.

Martino, M.M. et al., 2014. Growth factors engineered for super-affinity to the extracellular matrix enhance tissue healing. *Science*, 343(6173), pp.885–888.

Matsuoka, F. et al., 2013. Morphology-based prediction of osteogenic differentiation potential of human mesenchymal stem cells. *PLoS ONE*, 8(2), p.e55082.

Mayer, H. et al., 2005. Vascular endothelial growth factor (VEGF-A) expression in human mesenchymal stem cells: autocrine and paracrine role on osteoblastic and endothelial differentiation. *Journal of Cellular Biochemistry*, 95(4), pp.827–839.

McBeath, R. et al., 2004. Cell Shape, Cytoskeletal Tension, and RhoA Regulate

Stem Cell Lineage Commitment. *Developmental Cell*, 6(4), pp.483–495.

McMurray, R.J. et al., 2011. Nanoscale surfaces for the long-term maintenance of mesenchymal stem cell phenotype and multipotency. *Nature Materials*, 10(8), pp.637–644.

McNamara, L.E. et al., 2012. Metabolomics: a valuable tool for stem cell monitoring in regenerative medicine. *Journal of the Royal Society, Interface*, 9(73), pp.1713–1724.

McNamara, L.E. et al., 2011. Skeletal stem cell physiology on functionally distinct titania nanotopographies. *Biomaterials*, 32(30), pp.7403–7410.

McQuillan, D.J., Richardson, M.D. & Bateman, J.F., 1995. Matrix deposition by a calcifying human osteogenic sarcoma cell line (SAOS-2). *Bone*, 16(4), pp.415–426.

Mei, Y. et al., 2010. Combinatorial development of biomaterials for clonal growth of human pluripotent stem cells. *Nature Materials*, 9(9), pp.768–778.

Meijering, E., 2012. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Processing Magazine*, 29(5), pp.140–145.

Meyer, F. & Beucher, S., 1990. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1), pp.21–46.

Méndez-Ferrer, S. et al., 2010. Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature*, 466(7308), pp.829–834.

Mitsi, M. et al., 2006. Heparin-mediated conformational changes in fibronectin expose vascular endothelial growth factor binding sites. *Biochemistry*, 45(34), pp.10319–10328.

Mnatsakanyan, H. et al., 2015. Controlled Assembly of Fibronectin Nanofibrils Triggered by Random Copolymer Chemistry. *ACS Applied Materials & Interfaces*, 7(32), pp.18125–18135.

Moulisová, V. et al., 2017. Engineered microenvironments for synergistic VEGF – Integrin signalling during vascularization. *Biomaterials*, 126, pp.61–74.

Mouw, J.K., Ou, G. & Weaver, V.M., 2014. Extracellular matrix assembly: a multiscale deconstruction. *Nature Reviews Molecular Cell Biology*, 15(12), pp.771–785.

Movasaghi, Z., Rehman, S. & ur Rehman, D.I., 2008. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews*, 43(2), pp.134–179.

Murphy, W.L., McDevitt, T.C. & Engler, A.J., 2014. Materials as stem cell regulators. *Journal of Dental Research*, 13(6), pp.547–557.

Ng, D.H.J. et al., 2014. Microtubule-dependent modulation of adhesion complex composition. M. Parsons, ed. *PLoS ONE*, 9(12), p.e115213.

NIST Mass Spec Data Center, 2017. “Infrared Spectra.” In P. J. Linstrom & W. G.

Mallard, eds. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*. Gaithersburg MD.

Oakley, B.R., 2000. An abundance of tubulins. *Trends in Cell Biology*, 10(12), pp.537–542.

Oburoglu, L. et al., 2014. Glucose and glutamine metabolism regulate human hematopoietic stem cell lineage specification. *Cell Stem Cell*, 15(2), pp.169–184.

Odorico, J.S., Kaufman, D.S. & Thomson, J.A., 2001. Multilineage differentiation from human embryonic stem cell lines. *Stem Cells*, 19(3), pp.193–204.

Okumura, A. et al., 2001. Substrate affects the initial attachment and subsequent behavior of human osteoblastic cells (Saos-2). *Biomaterials*, 22(16), pp.2263–2271.

Orkin, S.H., 2000. Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews Genetics*, 1(1), pp.57–64.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp.62–66.

Paine, M.R.L. et al., 2017. Visualizing molecular distributions for biomaterials applications with mass spectrometry imaging: a review. *Journal of Materials Chemistry B*, 5(36), pp.7444–7460.

Panopoulos, A.D. et al., 2012. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Research*, 22(1), pp.168–177.

Parsons, J.T., Horwitz, A.R. & Schwartz, M.A., 2010. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Reviews Molecular Cell Biology*, 11(9), pp.633–643.

Patti, G.J., Yanes, O. & Siuzdak, G., 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13(4), pp.263–269.

Pautke, C. et al., 2004. Characterization of osteosarcoma cell lines MG-63, Saos-2 and U-2 OS in comparison to human osteoblasts. *Journal of Orthopaedic Research*, 24(6), pp.3743–3748.

Peng, H., 2008. Bioimage informatics: a new area of engineering biology. *Bioinformatics*, 24(17), pp.1827–1836.

Pepperkok, R. & Ellenberg, J., 2006. High-throughput fluorescence microscopy for systems biology. *Nature Reviews Molecular cell biology*, 7(9), pp.690–696.

Phillips, J.E. et al., 2010. Human mesenchymal stem cell differentiation on self-assembled monolayers presenting different surface chemistries. *Acta Biomaterialia*, 6(1), pp.12–20.

Pinho, S. et al., 2013. PDGFRα and CD51 mark human nestin+ sphere-forming

mesenchymal stem cells capable of hematopoietic progenitor cell expansion. *Journal of Experimental Medicine*, 210(7), pp.1351–1367.

Pittenger, M.F. et al., 1999. Multilineage potential of adult human mesenchymal stem cells. *Science*, 284(5411), pp.143–147.

Pons, J. et al., 2008. VEGF improves survival of mesenchymal stem cells in infarcted hearts. *Biochemical and Biophysical Research Communications*, 376(2), pp.419–422.

Ramalho-Santos, M. et al., 2002. "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science*, 298(5593), pp.597–600.

Reynolds, P.M. et al., 2013. Label-free segmentation of Co-cultured cells on a nanotopographical gradient. *Nano Letters*, 13(2), pp.570–576.

Rico, P. et al., 2016. Material-Driven Fibronectin Assembly Promotes Maintenance of Mesenchymal Stem Cell Phenotypes. *Advanced Functional Materials*, 26(36), pp.6563–6573.

Rico, P. et al., 2009. Substrate-induced assembly of fibronectin into networks: influence of surface chemistry and effect on osteoblast adhesion. *Tissue Engineering Part A*, 15(11), pp.3271–3281.

Rodrigues, M., Griffith, L.G. & Wells, A., 2010. Growth factor regulation of proliferation and survival of multipotential stromal cells. *Stem Cell Research & Therapy*, 1(4), p.32.

Rodríguez, J.P. et al., 2004. Cytoskeletal organization of human mesenchymal stem cells (MSC) changes during their osteogenic differentiation. *Journal of Cellular Biochemistry*, 93(4), pp.721–731.

Rutledge, D.N. & Jouan-Rimbaud Bouveresse, D., 2013. Independent Components Analysis with the JADE algorithm. *TrAC Trends in Analytical Chemistry*, 50(Supplement C), pp.22–32.

Sacchetti, B. et al., 2007. Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. *Cell*, 131(2), pp.324–336.

Salasznyk, R.M. et al., 2004. Adhesion to Vitronectin and Collagen I Promotes Osteogenic Differentiation of Human Mesenchymal Stem Cells. *BioMed Research International*, 2004(1), pp.24–34.

Salmerón-Sánchez, M. et al., 2011. Role of material-driven fibrillogenesis in cell differentiation. *Biomaterials*, 32(8), pp.2099–2105.

Salmerón-Sánchez, M. & Dalby, M.J., 2016. Synergistic growth factor microenvironments. *Chemical Communications*, 52(91), pp.13327–13336.

Salmerón-Sánchez, M. et al., 2011. Role of material-driven fibronectin fibrillogenesis in cell differentiation. *Biomaterials*, 32(8), pp.2099–2105.

Scadden, D.T., 2006. The stem-cell niche as an entity of action. *Nature*, 441(7097), pp.1075–1079.

Scheltema, R.A. et al., 2011. PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Analytical Chemistry*, 83(7), pp.2786–2793.

Schindelin, J. et al., 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), pp.676–682.

Schindelin, J. et al., 2015. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, 82(7-8), pp.518–529.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), pp.671–675.

Schröder, H.C. et al., 2005. Mineralization of SaOS-2 cells on enzymatically (silicatein) modified bioactive osteoblast-stimulating surfaces. *Journal of Biomedical Materials Research - Part B Applied Biomaterials*, 75(2), pp.387–392.

Schugar, R.C., Robbins, P.D. & Deasy, B.M., 2008. Small molecules in stem cell self-renewal and differentiation. *Gene Therapy*, 15(2), pp.126–135.

Seras-Franzoso, J. et al., 2014. Topographically targeted osteogenesis of mesenchymal stem cells stimulated by inclusion bodies attached to polycaprolactone surfaces. *Nanomedicine*, 9(2), pp.207–220.

Shamir, L. et al., 2010. Pattern recognition software and techniques for biological image analysis. *PLoS Computational Biology*, 6(11), p.e1000974.

Sharma, U., Pal, D. & Prasad, R., 2014. Alkaline phosphatase: an overview. *Indian Journal of Clinical Biochemistry*, 29(3), pp.269–278.

Shmueli, G., 2010. To Explain or to Predict? *Statistical Science*, 25(3), pp.289–310.

Siddappa, R. et al., 2007. Donor variation and loss of multipotency during in vitro expansion of human mesenchymal stem cells for bone tissue engineering. *Journal of Orthopaedic Research*, 25(8), pp.1029–1041.

Simmons, P.J. & Torok-Storb, B., 1991. Identification of stromal cell precursors in human bone marrow by a novel monoclonal antibody, STRO-1. *Blood*, 78(1), pp.55–62.

Singh, P., Carraher, C. & Schwarzbauer, J.E., 2010. Assembly of fibronectin extracellular matrix. *Current Opinion in Structural Biology*, 26(1), pp.397–419.

Smith, C.A. et al., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), pp.779–787.

Smith, M.L. et al., 2007. Force-induced unfolding of fibronectin in the extracellular matrix of living cells. *PLoS Biology*, 5(10), p.e268.

Soille, P., 2013. *Morphological Image Analysis*, Springer Science & Business

Media.

Sommer, C. & Gerlich, D.W., 2013. Machine learning in cell biology – teaching computers to recognize phenotypes. *Journal of Cell Science*, 126(24), pp.5529–5539.

Sommer, C. et al., 2011. Ilastik: Interactive learning and segmentation toolkit. *Proceedings - International Symposium on Biomedical Imaging*, pp.230–233.

Sonowal, H. et al., 2013. Inhibition of actin polymerization decreases osteogeneic differentiation of mesenchymal stem cells through p38 MAPK pathway. *Journal of Biomedical Science*, 20(1), p.71.

Spicer, R. et al., 2017. Navigating freely-available software tools for metabolomics analysis. *Metabolomics*, 13(9), p.106.

Stehbens, S. & Wittmann, T., 2012. Targeting and transport: how microtubules control focal adhesion dynamics. *The Journal of Cell Biology*, 198(4), pp.481–489.

Stein, G.S. & Lian, J.B., 1993. Molecular mechanisms mediating proliferation/differentiation interrelationships during progressive development of the osteoblast phenotype. *Endocrine Reviews*, 14(4), pp.424–442.

Stein, G.S. et al., 2004. Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. *Oncogene*, 23(24), pp.4315–4329.

Steiner, G. et al., 2007. Conformational changes during protein adsorption. FT-IR spectroscopic imaging of adsorbed fibrinogen layers. *Analytical Chemistry*, 79(4), pp.1311–1316.

Stenderup, K. et al., 2001. Number and proliferative capacity of osteogenic stem cells are maintained during aging and in patients with osteoporosis. *Journal of Bone and Mineral Research*, 16(6), pp.1120–1129.

Stewart, K. et al., 1999. Further characterization of cells expressing STRO-1 in cultures of adult human bone marrow stromal cells. *Journal of Bone and Mineral Research*, 14(8), pp.1345–1356.

Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), pp.41–45.

Sumner, L.W. et al., 2007. Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3(3), pp.211–221.

Surdo, Lo, J. & Bauer, S.R., 2012. Quantitative approaches to detect donor and passage differences in adipogenic potential and clonogenicity in human bone marrow-derived mesenchymal stem cells. *Tissue Engineering. Part C, Methods*, 18(11), pp.877–889.

Surrati, A. et al., 2016. Non-destructive characterisation of mesenchymal stem cell differentiation using LC-MS-based metabolite footprinting. *The Analyst*,

141(12), pp.3776–3787.

Swift, J. et al., 2013. Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation. *Science*, 341(6149), pp.1240104–1240104.

Štefková, K., Procházková, J. & Pacherník, J., 2015. Alkaline phosphatase in stem cells. *Stem Cells International*, 2015(10), pp.628368–11.

Tautenhahn, R. et al., 2012. XCMS Online: a web-based platform to process untargeted metabolomic data. *Analytical Chemistry*, 84(11), pp.5035–5039.

To, W.S. & Midwood, K.S., 2011. Plasma and cellular fibronectin: distinct and independent functions during tissue repair. *Fibrogenesis & Tissue Repair*, 4(1), p.21.

Tomfohr, J., Lu, J. & Kepler, T.B., 2005. Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics*, 6(1), p.225.

Tormin, A. et al., 2011. CD146 expression on primary nonhematopoietic bone marrow stem cells is correlated with in situ localization. *Blood*, 117(19), pp.5067–5077.

Treiser, M.D. et al., 2010. Cytoskeleton-based forecasting of stem cell lineage fates. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), pp.610–615.

Trevisan, J. et al., 2012. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *The Analyst*, 137(14), pp.3202–3215.

Tsimbouri, P. et al., 2014. Nanotopographical effects on mesenchymal stem cell morphology and phenotype. *Journal of Cellular Biochemistry*, 115(2), pp.380–390.

Unadkat, H.V. et al., 2011. An algorithm-based topographical biomaterials library to instruct cell fate. *Proceedings of the National Academy of Sciences of the United States of America*, 108(40), pp.16565–16570.

Unadkat, H.V. et al., 2013. High content imaging in the screening of biomaterial-induced MSC behavior. *Biomaterials*, 34(5), pp.1498–1505.

Urquhart, A.J. et al., 2007. High Throughput Surface Characterisation of a Combinatorial Material Library. *Advanced Materials*, 19(18), pp.2486–2491.

van der Walt, S. et al., 2014. scikit-image: image processing in Python. *PeerJ*, 2(2), p.e453.

Vanterpool, F.A. et al., 2014. A material-based platform to modulate fibronectin activity and focal adhesion assembly. *BioResearch Open Access*, 3(6), pp.286–296.

Vega, S.L. et al., 2012. High-Content Imaging-Based Screening of Microenvironment-Induced Changes to Stem Cells. *Journal of Biomolecular*

Screening, 17(9), pp.1151–1162.

Vega, S.L. et al., 2015. Organizational metrics of interchromatin speckle factor domains: integrative classifier for stem cell adhesion & lineage signaling. *Integrative Biology*, 7(4), pp.435–446.

Venables, W.N. & Ripley, B.D., 2002. *Modern Applied Statistics with S*, Springer.

Vimalraj, S. et al., 2015. Runx2: Structure, function, and phosphorylation in osteoblast differentiation. *International Journal of Biological Macromolecules*, 78, pp.202–208.

Wang, N., Tytell, J.D. & Ingber, D.E., 2009. Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus. *Nature Reviews Molecular Cell Biology*, 10(1), pp.75–82.

Wennberg, C. et al., 2000. Functional characterization of osteoblasts and osteoclasts from alkaline phosphatase knockout mice. *Journal of Bone and Mineral Research*, 15(10), pp.1879–1888.

Whitelock, 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology*, 18(5), pp.1368–1373.

Wickham, H., 2016. *ggplot2*, Springer.

Wickham, H., 2011. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software, Articles*, 40(1), pp.1–29.

Wijelath, E.S. et al., 2006. Heparin-II domain of fibronectin is a vascular endothelial growth factor-binding domain: enhancement of VEGF biological activity by a singular growth factor/matrix protein synergism. *Circulation Research*, 99(8), pp.853–860.

Wijelath, E.S. et al., 2002. Novel vascular endothelial growth factor binding domains of fibronectin enhance vascular endothelial growth factor biological activity. *Circulation Research*, 91(1), pp.25–31.

Williams, E.L., White, K. & Oreffo, R.O.C., 2013. Isolation and enrichment of Stro-1 immunoselected mesenchymal stem cells from adult human bone marrow. *Methods in Molecular Biology*, 1035(Chapter 7), pp.67–73.

Wilson, C.J. et al., 2005. Mediation of biomaterial-cell interactions by adsorbed proteins: a review. *Tissue Engineering*, 11(1-2), pp.1–18.

Wishart, D.S. et al., 2017. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research*, pp.gkx1089–gkx1089 VL – IS –.

Wozniak, M.A. et al., 2004. Focal adhesion regulation of cell behavior. *Biochimica et Biophysica Acta*, 1692(2-3), pp.103–119.

Xia, J. et al., 2015. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Research*, 43(W1), pp.W251–7.

Xie, L. et al., 2015. Characterization of Nestin, a Selective Marker for Bone

Marrow Derived Mesenchymal Stem Cells. *Stem Cells International*, 2015(7-8), pp.762098–9.

Yang, C. et al., 2016. Spatially patterned matrix elasticity directs stem cell fate. *Proceedings of the National Academy of Sciences of the United States of America*, 113(31), pp.E4439–45.

Yourek, G., Hussain, M.A. & Mao, J.J., 2007. Cytoskeletal changes of mesenchymal stem cells during differentiation. *ASAIO Journal*, 53(2), pp.219–228.

Zamboni, N., Saghatelian, A. & Patti, G.J., 2015. Defining the metabolome: size, flux, and regulation. *Molecular Cell*, 58(4), pp.699–706.

Zhou, S. et al., 2008. Age-related intrinsic changes in human bone-marrow-derived mesenchymal stem cells and their differentiation to osteoblasts. *Aging Cell*, 7(3), pp.335–343.

Zollinger, A.J. & Smith, M.L., 2017. Fibronectin, the extracellular glue. *Matrix Biology*, 60-61, pp.27–37.