CHAPTER II
LITERATURE REVIEW

2.1 Pluripotent stem cells (PSCs)

2.1.1 A defining characteristic of pluripotency

PSCs are a class of stem cells defined by their extraordinary ability to self–renew indefinitely in an undifferentiated state and to differentiate into all three embryonic germ layers: the ectoderm, mesoderm, and endoderm (Thomson et al., 1998a; Yamanaka and Thomson, 2010). This key characteristic of pluripotency makes them invaluable for regenerative medicine, drug discovery, and developmental biology research. The concept of pluripotency is distinct from totipotency, which describes a cell's ability to form all cell types of a complete organism, including extra–embryonic tissues like the placenta (Thomson et al., 1995). PSCs are the progenitors of all cell types within the body proper, but they cannot give rise to extra–embryonic tissues on their own.

PSCs can broadly be classified into two main types based on their origin: ESCs and induced pluripotent stem cells (iPSCs). While ESCs are derived directly from the inner cell mass of a blastocyst–stage embryo, iPSCs are generated artificially by reprogramming somatic cells back to a pluripotent state (Takahashi et al., 2007). Both types of cells share a common set of pluripotency markers, including the transcription factors *OCT4*, *SOX2*, and *NANOG*, which form a core regulatory network that maintains the undifferentiated state (Jaenisch and Young, 2008).

2.1.2 Embryonic stem cells (ESCs)

ESCs are arguably the most well characterized and foundational type of PSCs. They are isolated from the inner cell mass of a pre implantation embryo, typically at the blastocyst stage (Thomson et al., 1998b). The first successful isolation of mESCs was reported in 1981, and this was followed by the landmark isolation of hESCs by Thomson et al. (1998). This discovery unlocked unprecedented opportunities for studying human development and disease.

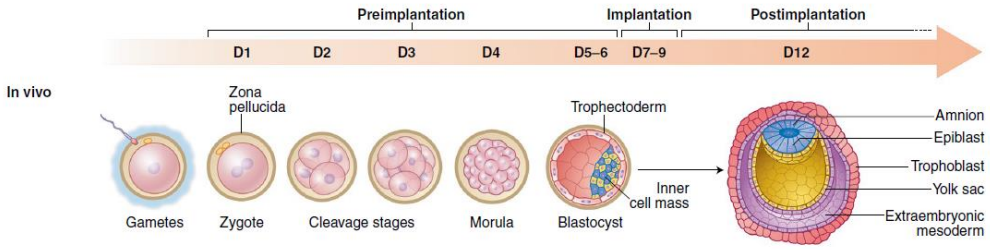

Figure 2.1 Early human development *in vivo*. (Mina et al., 2021).

A key advantage of ESCs is their robust and stable pluripotent state. They can be maintained in culture for long periods without losing their differentiation potential or acquiring significant genetic abnormalities, as long as appropriate culture conditions are maintained (Thomson et al., 1998b). The pluripotency of ESCs is functionally confirmed by their ability to differentiate into all three germ layers *in vitro* and, in the case of *in vivo* assays, by their ability to form teratomas benign tumors containing differentiated cells from all three germ layers when injected into immunocompromised animals (Adewumi et al., 2007).

Despite their scientific promise, the use of ESCs is associated with significant ethical considerations due to their embryonic origin (Thomson et al., 1998b). Additionally, their allogeneic nature means that derivatives of ESCs would subject to immune rejection when transplanted into a patient, necessitating immunosuppressive therapy or the development of immunologically matched cell banks (Depciuch et al., 2018). These challenges have driven the search for alternative sources of PSCs, leading to the discovery of iPSCs.

2.1.3 The relationship between PSCs and ESCs

The term "pluripotent stem cell" is a broader classification that encompasses ESCs. ESCs were the first type of PSCs to isolated and characterized, and they serve as the benchmark against which all other PSCs, particularly iPSCs, are measured (Jaenisch and Young, 2008). While iPSCs have largely overcome the ethical and immunological barriers associated with ESCs, a significant body of research is still dedicated to comparing the two cell types to ensure that iPSCs are functionally equivalent to ESCs (Szymanski et al., 2018). For example, studies often compare the gene expression profiles, epigenetic states, and differentiation efficiencies of iPSCs and ESCs to confirm that the reprogramming process has fully restored pluripotency (Reis et al., 2020).

mESCs were isolated first in 1981 (Evans and Kaufman 1981; Martin 1981). They were obtained by immunosurgically digesting the outer layer (trophectoderm) of mouse blastocysts to isolate the ICM that will adhere to the culture dish and

expand. The ICM is comprised of epiblast cells that give rise to the entire adult mouse. The mESCs harbor a normal karyotype even after long term culture. When initially isolated, mESCs exhibited enhanced growth and showed a higher propensity for differentiation when cultivated on mitotically inactivated mouse embryonic fibroblast (MEFs) feeder cells (Martin and Evans, 1974). The term "feeder" originally was adopted to define a cell providing critical nutrients or trophic factors support for the maintenance of an undifferentiated stem cell state. Furthermore, feeder cells may remove deleterious components from the media or may involve in cell–contact mediated mechanisms to prevent differentiation of the ESCs. The precise composition of the molecules secreted by feeder cells in any ESCs culture is currently unknown. mESCs, when injected into host blastocysts, produce viable chimaeras, by contributing to all fetal tissue lineages including the germline (Bradley et al., 1984).

ESCs from non–human primates (nhpESCs) and human (hESCs) were first derived in the mid–late–1990s by Thomson and colleagues (Thomson et al., 1995; Thomson et al., 1998c).

The hESCs have been shown to harbor a normal karyotype even after long-term culture, to express high levels of telomerase activity, and to express specific cell surface markers including SSEA3, SSEA4, TRA–1–60 and TRA–1–81. Even after extended periods of culture, these cells still maintain their developmental potential in derivatives of all three embryonic germ layers, including 1) endoderm; gut epithelium 2) mesoderm; cartilage, bone, smooth muscle and striated muscle 3) ectoderm; neural epithelium, embryonic ganglia and stratified squamous epithelium. In addition, they can give rise to derivatives of the trophoblast. These cell lines hold great promises in human developmental biology, drug discovery, and regenerative medicine.

The nhpESCs lines from rhesus monkey (*Macaca mulatta*, rhESCs) were first isolated in 1995 (Thomson et al., 1995). These cells, harbor a normal karyotype, and expresses the same cell surface markers as hESCs. The rhESCs remain undifferentiated when grown on MEFs feeder layers but differentiate or die in the absence of fibroblasts. Similarly to mESCs, they require different culture conditions for their maintenance in the pluripotent state. The rhESCs cells allowed to differentiate *in vitro* secrete bioactive chorionic gonadotropin into the medium, express chorionic gonadotropin alpha– and beta–subunit mRNAs, and express alpha fetoprotein mRNA, indicating trophoblast and endoderm differentiation. When injected into severe combined

immunodeficient mice, rhESCs formed teratomas composed of derivatives of all three embryonic germ layers.

The nhpESCs lines from other species, including the common marmoset (*Callithrix jacchus*, Thomson et al., 1996), and the cynomolgus monkey ESCs (cyESCs, Suemori et al., 2001), were also derived from blastocyst stage embryos obtained by *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI).

In summary, ESCs are a specific and highly important type of PSCs, while the term PSCs describes any cell with the defining characteristics of pluripotency, regardless of its origin. The foundational work with ESCs has been instrumental in the entire field of stem cell biology, and their continued study, alongside iPSCs, is crucial for advancing our understanding of pluripotency and for realizing the full potential of regenerative medicine.

2.2 Characteristics of primed, naïve-like, and naïve pluripotent stem cell states

PSCs are defined by their ability to self–renew and differentiate into all three embryonic germ layers (Thomson et al., 1998a). However, this pluripotency is not a single, static state. Instead, it exists along a continuum, with different states representing varying levels of developmental maturity and functional properties (Nichols and Smith, 2009). The two most well–defined states in this spectrum are naïve and primed pluripotency. More recently, a transitional or intermediate state, often referred to as naïve-like, has also been characterized, primarily in human stem cells. Understanding these distinct states is crucial for their reliable use in research and therapeutic applications.

Table 2.1 The first ESC lines derived from nonhuman primates (NHPs) (Anwised et al., 2023)

Species	Name of cell lines	Culture medium	Cells differentiation analysis	Reference
Marmoset (*C. jacchus*)	Cj11.2, Cj25.1, Cj28, Cj33, Cj35, Cj36, Cj39, Cj62	DMEM culture medium/ feeder–MEF	embryoid bodies	Thomson et al., 1996
Rhesus monkey (*M. mulatta*)	R278, R366, R367, R394, R420, R456, R460	DMEM medium/ feeder–MEF	embryoid bodies teratomas	Thomson et al., 1995
Cynomolgus monkey (*M. fascicularis*)	CMK5, CMK6, CMK7, CMK9	DMEM medium and Ham's nutrient mixture F–12/ feeder–MEF	embryoid bodies teratomas	Suemori et al., 2001

2.2.1 Naïve pluripotency

Naïve pluripotency represents the ground state of pluripotency and is considered the developmental equivalent of cells in the pre implantation ICM of a blastocyst (Boroviak and Nichols, 2017). Naïve PSCs exhibit several defining characteristics:

Morphology: They typically form tightly packed, dome–shaped colonies and are cultured on an inactivated feeder layer or in specific media containing inhibitors of the MAPK/ERK and GSK3 pathways (2i medium) (Nichols and Smith, 2009).

Transcriptional profile: Naïve PSCs express a core set of pluripotency markers, including *OCT4*, *SOX2*, and *NANOG*, at high levels. Critically, they also express unique markers such as *KLF4* and *TFCP2L1*, which are downregulated in primed cells (Marks et al., 2012). The global gene expression profile is characterized by a low level of spontaneous differentiation, and a global hypomethylation of DNA (Theunissen and Jaenisch, 2014).

Functional characteristics: Naïve PSCs possess a more potent differentiation capacity and can contribute to both the embryo proper and the germline when injected into a host blastocyst (Thomson et al., 1995). They also exhibit X–chromosome reactivation in female cells, which is a hallmark of the early embryonic state (Nichols and Smith, 2009).

2.2.2 Primed pluripotency

Primed pluripotency represents a later developmental stage, analogous to the epiblast of a post implantation embryo (Boroviak and Nichols, 2017). hESCs were originally isolated and propagated in the primed state (Thomson et al., 1998b). The key characteristics of primed PSCs are:

Morphology: Primed PSCs grow as flat, compact colonies with more defined cell–cell junctions. They are typically cultured on a feeder layer or an extracellular matrix like Matrigel (Boroviak and Nichols, 2017).

Transcriptional profile: They also express core pluripotency factors like *OCT4* and *NANOG* but lack expression of naïve–specific markers. A key distinguishing feature is the high expression of lineage–priming genes, indicating a predisposition towards differentiation. The overall epigenetic landscape is characterized by global hypermethylation of DNA compared to the naïve state (Theunissen and Jaenisch, 2014).

Functional characteristics: Primed PSCs have a more restricted differentiation potential and cannot contribute efficiently to chimeras when injected into blastocysts (Guo et al., 2017). In female primed cells, one of the two X–chromosomes is inactivated, mirroring the *in vivo* post implantation epiblast state (Nichols and Smith, 2009).

2.2.3 Naïve-like pluripotent stem cells

The concept of a naïve-like state has emerged primarily in the context of human stem cell research, where researchers have successfully reprogrammed or converted human primed ESCs to a naïve state using specific culture conditions (Theunissen et al., 2014). These "naïve-like" cells often share many of the morphological, transcriptional, and functional characteristics of true naïve cells (i.e., mouse ESCs), but their full equivalence to the pre implantation human ICM remains an active area of investigation (Boroviak and Nichols, 2017).

The significance of the naïve-like state is immense, as it offers a more robust and efficient platform for genetic manipulation and disease modeling (Baker et al., 2014). For example, it has been shown that converting primed hESCs to the naïve-like state increases their efficiency for genome editing via CRISPR/Cas9 (Theunissen and Jaenisch, 2014). However, the long–term stability and true biological fidelity of these converted cells are still under scrutiny.

The understanding that pluripotency is not a single state but rather a dynamic spectrum has revolutionized stem cell research. The distinct characteristics of naïve, primed, and naïve-like PSCs provide a crucial framework for guiding research and therapeutic development. While primed cells are valuable for their ability to model post implantation development, naïve cells offer a more flexible and potent tool for fundamental research and advanced genetic engineering. The ability to switch between these states provides an unprecedented level of control over stem cell behavior, paving the way for more effective and targeted applications in regenerative medicine.

Differences between naïve and primed pluripotency have been extensively studied, including culture conditions, transcriptomic, epigenetic and metabolic profiles as outlined in the recent review by Weinberger et al. (2016) (Table 2.2). The naïve state represents the cellular state of the pre–implantation mouse blastocyst epiblast, while the primed state is representative of the post implantation epiblast cells. These two cell types exhibit clearly distinct developmental potential, as evidenced by the fact that naïve cells are able to contribute to blastocyst chimeras, while primed cells cannot. The epigenetic, transcriptomic, metabolic and proteomic differences should lead to a better understanding of the fundamental properties of these two states of pluripotency (Takahashi et al., 2018). Nowadays, many techniques allow deep characterization of pluripotent stem cell states such as immunocytochemistry, RNA–seq, qPCR as well as FTIR microspectroscopy.

2.3 Rhesus macaques (*Macaca mulatta*)

Rhesus macaques are a critical model in biomedical research, and their embryonic stem cells serve as a vital tool for understanding human pluripotency and for developing cell–based therapies. The close genetic and physiological similarities

between rhesus macaques and humans make them an ideal translational model for studying human diseases that cannot accurately replicated in rodent models, such as HIV (Human immunodeficiency virus) , Alzheimer's, and Parkinson's disease. The establishment and characterization of rhESCs from this species have paved the way for the derivation of hESCs and have provided a platform for preclinical studies in regenerative medicine.

Table 2.2 Characteristic features of naïve and primed pluripotent stem cell states (Kumari, 2016)

Property	Naïve state	Primed state
1. Colony morphology	Compact dome–shaped	Flattened
2. Expressed genes	High expression of *OCT4, NANOG, SOX2, DPPA3, TFCP2L1, ZFP42, KLF2, KLF4, KLF5, ESRRB, FGF4, CDH1*	*OCT4, SOX2, OTX2, DNMT3B, FGF5, POU3F1, MEIS1, SOX11, GDF3*
3. Growth factor dependence	LIF	Activin, FGF2
4. Single–cell mortality	Low	High
5. XCI status in female cells	XaXa	XaXi
6. H3K27Me3 over developmental regulators	Low	High

2.3.1 Rhesus macaques (*Macaca mulatta*) as a biomedical model

The rhesus macaque (*Macaca mulatta*) is one of the most widely used non–human primates in biomedical research due to its close phylogenetic relationship to humans (Thomson et al., 1995). With a genome that is approximately 93% homologous to the human genome, the rhesus macaque serves as a highly relevant translational model for studying human physiology and disease progression (Szymanski et al., 2018). The species' use has been instrumental in numerous medical breakthroughs, including the development of the polio vaccine and the discovery of the Rh factor in blood typing (Reis et al., 2020).

The value of the rhesus macaque model extends across a wide range of research areas, including infectious diseases, such as HIV/AIDS (acquired immunodeficiency syndrome), where studies in rhesus macaques infected with simian immunodeficiency virus (SIV) have provided invaluable insights into viral pathogenesis

and have guided the development of antiretroviral therapies and vaccines (Wong et al., 2021). The longevity and natural aging processes of rhesus macaques also make them an excellent model for studying age–related diseases like Alzheimer's, as they naturally accumulate the misfolded amyloid–beta proteins observed in human patients (Baker et al., 2014).

2.3.2 Rhesus macaque embryonic stem cells (rhESCs)

rhESCs are pluripotent stem cells isolated from the ICM of rhesus monkey blastocysts. Their derivation by Thomson et al. 1995) was a landmark achievement that preceded the isolation of hESCs in 1998 (Thomson et al., 1995; Thomson et al., 1998a). The establishment of rhESCs provided the foundational knowledge and culture protocols that were essential for the subsequent work with human cells.

rhESCs share key characteristics with their human counterparts, including the ability for self–renewal and pluripotency (Thomson and Marshall, 1998a). They can proliferate indefinitely in an undifferentiated state *in vitro* and maintain the potential to differentiate into cells of all three embryonic germ layers: ectoderm, mesoderm, and endoderm (Thomson and Marshall, 1998a). This pluripotency is confirmed by the expression of a common set of pluripotency markers, such as the transcription factors *OCT4*, *SOX2*, and *NANOG*, and cell surface markers like SSEA–4 and TRA–1–60 (Adewumi et al., 2007).

The similarities between rhESCs and hESCs make the rhesus macaque a crucial model for preclinical studies of cell–based therapies. It allows researchers to evaluate the safety, efficacy, and potential for immune rejection of stem cell derivatives in a physiological context that closely mimics the human body (Miller et al., 2014). This is particularly important for therapies aimed at treating neurodegenerative diseases, as rhesus macaques naturally develop many of the same neurological conditions as humans (Chan and Kazarian, 2013).

2.3.3 Significance and future directions

The study of rhesus macaques and their embryonic stem cells continue to be of paramount importance in advancing our understanding of human biology and disease. The use of this model has not only facilitated the development of stem cell research but has also provided a bridge for translating laboratory discoveries into clinical applications. The ongoing research into rhESCs, including their genetic stability, differentiation potential, and immune compatibility, is essential for ensuring the safety and success of future regenerative medicine therapies.

2.4 Characterization of rhESCs by standard methods

rhESCs are pluripotent stem cells derived from the inner cell mass of blastocyst stage rhesus monkey embryos (Thomson et al., 1995). As a non–human primate model, the rhesus macaque is phylogenetically close to humans, and its stem cells share many biological similarities with hESCs. This makes rhESCs an invaluable model for studying fundamental pluripotency mechanisms, and for preclinical safety and efficacy studies of cell–based therapies before their application in humans (Thomson et al., 1995; Thomson and Marshall, 1998a). To ensure the proper identity, quality, and functionality of these cell lines, a rigorous and standardized set of characterization methods is essential.

The routine characterization of rhESCs involves a combination of morphological, molecular, and functional assays to confirm their key properties: pluripotency, self–renewal, and genetic stability.

2.4.1 Morphological and self–renewal assessment

The most basic step in rhESC characterization is the assessment of their morphology and ability to self–renew in an undifferentiated state (Thomson et al., 1995). Undifferentiated rhESCs typically grow as compact, three–dimensional colonies with sharp borders. The cells within these colonies are small, have a high nuclear–to–cytoplasmic ratio, and possess large, prominent nucleoli (Figure 4.1). This morphology is a key indicator of their undifferentiated state (Suemori et al., 2006). The ability of the cells to maintain this morphology and proliferate for an extended period in culture, without spontaneous differentiation, is evidence of their self–renewal capacity (Thomson and Marshall, 1998a).

2.4.2 Pluripotency marker expression

Pluripotency is the defining characteristic of ESCs, and it is confirmed by the expression of specific molecular markers. These markers can be detected at both the protein and gene levels.

Protein expression: Key pluripotency associated proteins include OCT4, SOX2, and NANOG (Thomson et al., 1995; Thomson and Marshall, 1998c). These transcription factors form a core regulatory network that maintains the pluripotent state. Their presence is commonly verified using immunocytochemistry or flow cytometry (Thomson et al., 1995). Other surface markers, such as SSEA–4 and TRA–1–60, which are found on the cell surface of pluripotent primate cells, are also routinely used to confirm pluripotency (Thomson et al., 1995; Adewumi et al., 2007).

Gene expression: The expression of pluripotency related genes, including *POU5F1* (encoding *OCT4*), *SOX2*, and *NANOG*, is typically measured using reverse

transcription polymerase chain reaction (RT–PCR) or quantitative real time PCR (qRT–PCR). The high expression levels of these genes in undifferentiated rhESCs, followed by their downregulation upon differentiation, is a standard metric for characterization (Suemori et al., 2006).

2.4.3 Pluripotency functional assays

Molecular marker expression is necessary but not sufficient to prove pluripotency. Functional assays are required to demonstrate the cells' ability to differentiate into all three germ layers: ectoderm, mesoderm, and endoderm.

In vitro differentiation: This involves inducing the cells to differentiate into derivatives of all three germ layers in a controlled cell culture environment (Thomson and Marshall, 1998). The formation of different cell types, such as neurons (ectoderm), cardiomyocytes (mesoderm), and gut–like cells (endoderm), can confirmed by analyzing the expression of specific lineage markers using immunocytochemistry or RT–PCR (Thomson et al., 1995).

In vivo teratoma formation: The gold standard for confirming pluripotency is the teratoma formation assay (Adewumi et al., 2007). This involves injecting undifferentiated rhESCs into an immunocompromised mouse. The formation of a benign tumor called a teratoma, which contains differentiated cell types from all three embryonic germ layers, is definitive proof of pluripotency. Histological analysis of the resulting tumor is used to identify the presence of these diverse cell types (Thomson et al., 1995; Thomson and Marshall, 1998).

2.4.4 Genetic stability assessment

Maintaining a stable and normal karyotype is critical for the safe and reliable use of rhESCs (Thomson et al., 1995). Prolonged culture can lead to the acquisition of chromosomal abnormalities, which can compromise the cells' properties or lead to oncogenic potential (Adewumi et al., 2007). Karyotyping is performed to visualize the chromosomes and confirm that the cells have a normal diploid chromosomal complement (2n=42 for rhesus macaque) (Thomson and Marshall, 1998c). This analysis is an essential part of quality control for rhESC lines.

The comprehensive characterization of rhESCs using a combination of morphological, molecular, and functional assays is a critical step in their establishment and application. The standard methods, including the assessment of morphology, expression of pluripotency markers (e.g., OCT4, SSEA–4), functional pluripotency assays (*in vitro* differentiation and teratoma formation), and genetic stability analysis, provide a robust framework for validating the identity and quality of these valuable cell lines.

These methods ensure that rhESCs serve as a reliable and reproducible model for biomedical research and as a crucial stepping stone for human clinical trials.

2.5 Immunocytochemistry: Principles and limitations

2.5.1 Principles of immunocytochemistry

Immunocytochemistry (ICC) is a widely used laboratory technique that harnesses the specific binding of antibodies to visualize the location and distribution of a target protein or antigen within a cell (Coons, 1956). The fundamental principle of ICC is the highly specific interaction between an antibody and its corresponding antigen. The term "immunocytochemistry" is generally used for the analysis of isolated cells or cell cultures, while a similar technique applied to tissue sections is called immunohistochemistry (Haines et al., 2007).

The process of ICC involves several key steps. First, the cells are fixed to a solid support (e.g., a glass slide) to preserve their morphological integrity and immobilize the proteins (Buchwalow et al., 2011). Next, the cell membrane may permeabilized to allow antibodies to access intracellular antigens. The cells are then incubated with a primary antibody that is raised against the target antigen of interest. This antibody binds directly to the protein being studied (Miller et al., 2014).

To visualize this primary antibody antigen complex, a detection system is employed. There are two main approaches:

2.5.1.1 Direct ICC

A primary antibody is directly conjugated to a reporter molecule, such as a fluorophore (e.g., FITC, Cy3) or an enzyme (e.g., horseradish peroxidase, HRP) (Depciuch et al., 2018). While this method is simple and fast, it often suffers from lower signal amplification and is less sensitive than indirect methods.

2.5.1.2 Indirect ICC

This more common and sensitive method uses an unconjugated primary antibody and a secondary antibody that is conjugated to a reporter molecule. The secondary antibody is raised against the species of the primary antibody (e.g., a goat anti rabbit secondary antibody for a primary antibody raised in a rabbit). Multiple secondary antibodies can bind to a single primary antibody, leading to a significant amplification of the signal (Buchwalow et al., 2011). This method is highly sensitive and allows for the use of a wide range of commercially available labeled secondary antibodies.

Following the labeling steps, the cells are typically counterstained to visualize cellular structures, such as the nucleus, and then imaged using a microscope (Ramos-Vara and Miller, 2014).

2.5.2 Limitations of immunocytochemistry

Despite being a powerful and widely used technique, ICC has several inherent limitations that must considered for accurate and reliable results.

2.5.2.1 Antibody specificity and cross reactivity

The reliability of ICC is critically dependent on the specificity of the antibodies used. An antibody might bind to a protein other than the intended target (off target binding), leading to false positive results or a misinterpretation of the localization (Haines et al., 2007). Cross reactivity with related proteins is a common issue, particularly with polyclonal antibodies. Therefore, it is essential to validate antibodies rigorously using controls, such as Western blotting, to confirm that they recognize the correct protein (Buchwalow et al., 2011).

2.5.2.2 Fixation and permeabilization artifacts

The fixation and permeabilization steps, while necessary to preserve cellular morphology and allow antibody access, can introduce artifacts. Fixatives like paraformaldehyde can cross link proteins, potentially masking the epitope (the specific site on the antigen that the antibody recognizes), thereby leading to a weak or absent signal (Ramos-Vara and Miller, 2014). Similarly, detergents used for permeabilization can alter the native structure of the cell, potentially affect protein localization or cause them to washed away (Coons, 1956). The choice of fixative and permeabilization agent must carefully optimized for each specific antigen to minimize these issues (Baker et al., 2014).

2.5.2.3 Quantification challenges

While ICC provides excellent qualitative information on protein localization, its use for precise quantitative analysis is challenging. The intensity of the fluorescent or chromogenic signal is influenced by numerous factors, including antibody concentration, incubation times, and the efficiency of the reporter molecule (Coons, 1956). Direct comparison of signal intensity between different experiments is often unreliable. Therefore, ICC is generally not suitable for determining the absolute quantity of a protein but rather for assessing its relative abundance or cellular distribution (Miller et al., 2014; Depciuch et al., 2018). More sophisticated techniques, such as flow cytometry or western blotting, are often required for accurate quantification.

2.6 RNA sequencing: Principles and limitations

2.6.1 Principles of RNA sequencing

RNA sequencing (RNA–seq) is a high throughput technology that uses next generation sequencing (NGS) to analyze the expression, structure, and function of an organism's transcriptome (Wang et al., 2009). The fundamental principle of RNA–seq is to convert an RNA population, which is highly unstable, into a more stable complementary DNA (cDNA) library. This library is then sequenced to generate millions of short reads that represent the original RNA molecules (Stark et al., 2019). The digital nature of this data provides a more precise and quantitative measure of gene expression compared to traditional methods like microarrays.

The process of RNA–seq generally follows a standardized workflow:

RNA isolation: Total RNA, including mRNA, ribosomal RNA (rRNA), and non–coding RNA, is extracted from the sample. The quality and integrity of the RNA are critical for accurate results (Reis et al., 2020).

Library preparation: This is a crucial step that prepares the RNA for sequencing. It typically involves mRNA enrichment to remove the highly abundant rRNA, which would otherwise dominate the sequencing data. The enriched mRNA is then fragmented, and a reverse transcriptase enzyme is used to synthesize a stable cDNA copy (Mortazavi et al., 2008). Sequencing adapters are then ligated to the ends of the cDNA fragments.

Sequencing: The prepared cDNA library is sequenced using a high throughput platform, such as an Illumina sequencer (Miller et al., 2014). This generates millions of short sequencings reads from the ends of the cDNA fragments.

Data analysis: The short reads are computationally mapped to a reference genome or transcriptome. The number of reads that map to a specific gene or transcript is then counted to quantify its expression level (Wang et al., 2009). This provides a comprehensive gene expression profile of the sample, which can used to identify differentially expressed genes, novel transcripts, gene fusions, and single nucleotide variants (SNVs).

2.6.2 Limitations of RNA sequencing

Despite its widespread use and powerful capabilities, RNA–seq has several limitations that must addressed for reliable data interpretation.

2.6.2.1. Cost and computational demands

RNA–seq can expensive, especially for projects with a large number of samples or a requirement for deep sequencing to detect low abundance transcripts (Krimm and Bandekar, 1986). In addition, the vast amount of data generated by a single

sequencing run requires significant computational resources for storage, processing, and analysis. This necessitates access to high performance computing clusters and specialized bioinformatics expertise, which can a barrier for some research groups (Szymanski et al., 2018).

2.6.2.2 Challenges in quantification and normalization

Quantifying gene expression from RNA–seq data is not straightforward. The number of reads obtained for a gene is influenced not only by its true expression level but also by factors such as gene length and sequencing depth (Wong et al., 2021). To enable accurate comparisons between samples, the data must normalize to account for these technical variations (Depciuch et al., 2018). The choice of normalization method (e.g., TPM, FPKM, or DESeq2) can significantly impact the final results and conclusions (Baker et al., 2014).

2.6.2.3 Sample and technical variability

The quality of the starting RNA material is critical. Degraded RNA can lead to biases in the data, particularly towards the 3' end of transcripts (Miller et al., 2014). Additionally, technical variations introduced during library preparation, such as biases in reverse transcription or PCR amplification, can affect the final gene expression counts. While these biases can partially mitigate through rigorous quality control and careful experimental design, they remain a source of variability (Stark et al., 2019).

2.6.2.4 Difficulty in detecting low abundance transcripts

Sequencing depth directly impacts the ability to detect and quantify transcripts. For genes expressed at very low levels, a shallow sequencing depth may not generate enough reads to accurately represent their expression, leading to false negatives (Depciuch et al., 2018). Conversely, very high sequencing depth can cost prohibitive. Optimizing the balance between sequencing depth and cost is a common challenge in experimental design.

2.7 Characterization of ESCs by FTIR microspectroscopy

2.7.1 Principles, mechanisms and limitations of FTIR microspectroscopy

FTIR microspectroscopy is a powerful analytical technique that combines the principles of FTIR spectroscopy with optical microscopy, allowing for chemical analysis with high spatial resolution (Baker et al., 2014). The fundamental principle relies on the interaction between infrared radiation and matter. Specifically, molecules possess unique vibrational and rotational modes, such as stretching (changes in bond length) and bending (changes in bond angle) (Stuart, 2004). When the frequency of the incident IR radiation matches the natural vibrational frequency of a specific chemical

bond, the bond absorbs the energy, transitioning to a higher vibrational state. This absorption is unique to each type of bond and functional group, creating a distinctive molecular "fingerprint" that can used to identify and characterize the sample's chemical composition (Krimm and Bandekar, 1986).

The mechanism of generating the spectrum involves a Michelson interferometer. An incoming broadband IR beam is split into two paths by a beamsplitter, one path directed to a fixed mirror and the other to a movable mirror (Griffiths, 1983). The two beams are recombined, creating an interference pattern called an interferogram, which is a signal of the total IR intensity as a function of the optical path difference (OPD) between the two beams. The raw interferogram is not directly interpretable; therefore, a mathematical operation known as the Fourier transform is applied to convert the time domain interferogram into a frequency domain spectrum (absorbance versus wavenumber) (Stuart, 2004). The integration of this system with a microscope allows for the IR beam to focused onto a small area of the sample, enabling the collection of spectral data point by point, thereby generating a chemical map or image that links chemical information to spatial location (Wong et al., 2021).

2.7.2 Limitations of FTIR microspectroscopy

Despite its significant advantages, FTIR microspectroscopy has several inherent limitations that must carefully considered during experimental design and data interpretation.

2.7.2.1 Spatial resolution

One of the primary limitations is the diffraction limit of light, which restricts the achievable spatial resolution. For mid IR radiation, which is typically used, the wavelength is in the range of 2.5 to 25 μm, meaning the theoretical spatial resolution is limited to several micrometers (Reis et al., 2020). This can a major constraint when analyzing subcellular components or nanoparticles, which are often smaller than the diffraction limit. While techniques like attenuated total reflection (ATR) can improve resolution by using a crystal with a high refractive index, they require direct contact with the sample and are limited by the physical size of the ATR crystal (Chan and Kazarian, 2013).

2.7.2.2 Sample preparation and handling

Proper sample preparation is crucial but can challenging, particularly for biological samples. Samples must very thin (<10 μm) and uniform to avoid strong IR absorption, which can saturate the signal and obscure important spectral features (Miller et al., 2014). The presence of water is a particularly significant issue, as it has a strong and broad IR absorption band that can mask the signals from other key

biomolecules, such as proteins and lipids (Baker et al., 2014). Consequently, biological samples often need to dried, which can potentially alter their native biochemical and structural state.

2.7.2.3. Data processing and analysis

FTIR microspectroscopy generates large, complex datasets that necessitate sophisticated data processing and analysis. The raw spectra often contain noise, baseline artifacts, and overlapping peaks, requiring extensive pre-processing steps, including baseline correction and spectral normalization (Depciuch et al., 2018). Furthermore, the interpretation of the spectral data often requires the use of advanced chemometric methods such as PCA, hierarchical cluster analysis (HCA), and partial least squares (PLS) to extract meaningful information and identify subtle differences between samples (Depciuch et al., 2018). These complex computational demands require specialized software and expertise, which can a barrier for some researchers.

2.8 Diagnostic applications and performance of FTIR microspectroscopy

2.8.1 Diagnostic applications

FTIR microspectroscopy has emerged as a powerful label free diagnostic tool with diverse applications in biotechnology and medicine. Its ability to provide a "molecular fingerprint" of a sample by identifying the presence and relative abundance of biomolecules (e.g., proteins, lipids, carbohydrates, and nucleic acids) makes it suitable for various diagnostic purposes (Baker et al., 2014; Miller et al., 2014).

2.8.1.1 Cancer diagnosis and histopathology

One of the most significant applications is in the diagnosis of cancer. FTIR microspectroscopy can differentiate between healthy and cancerous tissues based on subtle biochemical changes that occur during carcinogenesis, often before morphological changes are evident (Depciuch et al., 2018). For instance, studies have shown that cancerous cells exhibit altered lipid and protein content, changes in nucleic acid conformation, and increased glycolysis, all of which are detectable through their unique IR spectral signatures (Lassiter et al., 2019; Reis et al., 2020). This technology is being explored for the rapid analysis of tissue biopsies, offering an alternative or complement to traditional histopathology, which relies on subjective visual interpretation by pathologists. It has been successfully applied to the diagnosis of various cancers, including breast (Lassiter et al., 2019), colon (Szymanski et al., 2018), and brain tumors (Miller et al., 2014).

2.8.1.2 Microbiology and infectious disease

FTIR microspectroscopy is also a valuable tool in microbiology for the rapid identification and classification of microorganisms, including bacteria, fungi, and viruses. Each microbial species possesses a unique macromolecular composition that yields a distinct IR spectrum (Mendelsohn and Moore, 2012). This allows for the differentiation of closely related species and strains, which is crucial for epidemiological studies and the rapid diagnosis of infectious diseases (Wong et al., 2021). The technique is particularly useful for identifying antibiotic resistant strains, offering a quick method to guide clinical treatment (Depciuch et al., 2018).

2.8.1.3 Stem cell and regenerative medicine

In the field of regenerative medicine, FTIR microspectroscopy provides a non–destructive method to monitor the differentiation and maturation of stem cells. The biochemical changes that accompany stem cell differentiation—such as changes in lipid profiles and protein secondary structures—can tracked in real time. This provides insight into the differentiation pathways and quality control for cell–based therapies (Kazarian and Chan, 2017).

2.8.2 Performance

The performance of FTIR microspectroscopy in a diagnostic context is defined by its sensitivity, specificity, and speed.

2.8.2.1. Sensitivity and specificity

The sensitivity and specificity of FTIR microspectroscopy in distinguishing between different sample types, such as healthy and diseased tissues, are generally very high, often exceeding 90% in well controlled studies (Baker et al., 2014). The high sensitivity is attributed to the fact that the technique measures the collective signal from thousands of molecules within a single measurement, making it highly sensitive to even subtle biochemical changes (Reis et al., 2020). The specificity is derived from the unique spectral fingerprint of different biomolecules and the use of sophisticated chemometric algorithms, such as PCA and LDA, to classify spectra accurately (Depciuch et al., 2018).

2.8.2.2 Speed and throughput

The speed of FTIR microspectroscopy is a significant performance advantage, especially when using modern instrumentation. Unlike many traditional diagnostic methods that require extensive staining and processing time, FTIR provides results in minutes. Advances in technology, such as the use of FPA detectors, have dramatically increased data acquisition speed by allowing for the simultaneous collection of spectra from a large area, making it possible to image entire tissue

sections in a single pass (Miller et al., 2014). This high throughput makes FTIR microspectroscopy a promising tool for high volume clinical and research applications.

2.9 Detectors in fourier transform infrared (FTIR) microspectroscopy

2.9.1 Principles and types of IR detectors

The performance of FTIR microspectroscopy is critically dependent on the type and sensitivity of the detector used to measure the IR radiation transmitted through or reflected from a sample (Griffiths, 1983). IR detectors operate by converting the incident IR radiation into an electrical signal. They are broadly classified into two main categories: thermal detectors and photon detectors (Stuart, 2004).

2.9.1.1 Thermal detectors

These detectors, such as the deuterated triglycine sulfate (DTGS) detector, measure the temperature change caused by the absorption of IR radiation. They are relatively inexpensive and do not require cryogenic cooling, which makes them easy to use (Depciuch et al., 2018). However, their response time is slower and their sensitivity is generally lower than that of photon detectors. For this reason, DTGS detectors are often used in routine or low cost FTIR systems where high speed and sensitivity are not the primary requirements (Miller et al., 2014).

2.9.1.2 Photon detectors

These detectors, such as the MCT detector, measure the change in the electrical properties (e.g., resistance) of a semiconductor material when it absorbs IR photons. MCT detectors are highly sensitive and have a very fast response time, making them ideal for high–speed data acquisition and for analyzing samples with low IR signal intensity (Baker et al., 2014). To achieve maximum performance, MCT detectors typically require cryogenic cooling with liquid nitrogen to minimize thermal noise (Wong et al., 2021). The high sensitivity of MCT detectors is particularly valuable for FTIR microspectroscopy, where the amount of IR radiation reaching the detector from a small sample area is often limited (Reis et al., 2020).

2.9.2 Focal plane array (FPA) detectors

A significant advancement in FTIR microspectroscopy has been the development of FPA detectors. Unlike single element detectors that acquire spectral data sequentially, FPA detectors consist of a two–dimensional array of individual detector elements (pixels) (Miller et al., 2014). This allows for the simultaneous acquisition of an entire IR image, where each pixel in the array collects a full IR spectrum from a corresponding point on the sample (Chan and Kazarian, 2013). This parallel data acquisition dramatically reduces the time required to generate a chemical

map or image of a sample, making it possible to analyze large areas of tissue or cell cultures in minutes, rather than hours (Kazarian and Chan, 2017). The high throughput and high sensitivity of FPA based systems have revolutionized the field, enabling new applications in diagnostic histopathology and high content screening (Lassiter et al., 2019). The use of FPA detectors allows for the rapid visualization of biochemical heterogeneities, which is crucial for distinguishing between different tissue types or identifying disease states.

2.9.3 Performance and limitations

The choice of detector has a direct impact on the performance of the FTIR microspectroscopy system. The signal to noise ratio (SNR), which is a critical measure of spectral quality, is significantly higher for MCT and FPA detectors compared to DTGS detectors, enabling the detection of subtle spectral features (Depciuch et al., 2018). While FPA detectors offer unparalleled speed and spatial information, they can more expensive and require specialized software for data processing due to the large file sizes generated (Wong et al., 2021). The cryogenic cooling required for MCT and FPA detectors also adds complexity and cost to the instrumentation. Despite these limitations, the superior sensitivity and speed of these detectors have made them the standard for advanced research and clinical applications of FTIR microspectroscopy (Reis et al., 2020).

2.10 Multivariate data analysis of FTIR microspectroscopy

2.10.1. The challenge of complex FTIR microspectroscopy data

FTIR microspectroscopy generates large and complex datasets, especially when using modern FPA detectors (Baker et al., 2014). A single chemical map can contain thousands of individual spectra, with each spectrum comprising hundreds or even thousands of data points (wavenumbers). This results in a highly dimensional dataset, where the spectral variations are often subtle and overlapping, making them difficult to interpret by visual inspection alone (Miller et al., 2014). To extract meaningful information from these complex datasets and identify subtle biochemical differences, multivariate data analysis (chemometrics) is essential (Reis et al., 2020). These statistical techniques are designed to handle and simplify multi–dimensional data, revealing underlying patterns and relationships.

2.10.2 Common multivariate data analysis techniques

Several chemometric techniques are routinely applied to FTIR microspectroscopy data. These can broadly be classified into unsupervised and supervised methods.

2.10.2.1. Unsupervised methods: exploring data without prior knowledge

Unsupervised methods are used to explore the data and identify natural groupings or patterns without any prior knowledge or labels.

Principal component analysis (PCA): PCA is the most widely used unsupervised method for analyzing FTIR data (Depciuch et al., 2018). It works by reducing the dimensionality of the dataset by identifying the principal components (PCs), which are new variables that capture the maximum variance in the original data. The first few PCs often represent the most significant sources of variation in the sample, such as differences between healthy and diseased tissues or different cell types (Wong et al., 2021). The results are often visualized in a scores plot to show how samples are clustered and loadings plot to identify which wavenumbers (and thus which chemical bonds) are responsible for the observed clustering (Kazarian and Chan, 2017).

Hierarchical cluster analysis (HCA): HCA is a clustering method that groups similar spectra together based on a measure of spectral distance (Miller et al., 2014). The output is a dendrogram that visually represents the relationships between spectra, with similar spectra located close to each other. HCA is often used in combination with PCA to provide an objective way of classifying different regions within a sample, such as identifying a tumor margin in a tissue section (Chan and Kazarian, 2013).

2.10.2.2. Supervised methods: classification and prediction

Supervised methods are used when there is prior knowledge about the samples (e.g., control vs. treatment, healthy vs. diseased). These methods use a training set of labeled data to build a model that can predict the class of a new, unknown sample.

Linear discriminant analysis (LDA): LDA is a classification technique often used after PCA to enhance the separation between predefined groups (Szymanski et al., 2018). It finds a linear combination of features that best separates two or more classes of objects. When combined with PCA (known as PCA–LDA), it becomes a powerful tool for building robust diagnostic models for applications such as cancer detection (Lassiter et al., 2019).

Partial least squares (PLS): PLS is a regression technique used to model the relationship between the spectral data (X–variables) and an external variable of interest (Y–variable), such as a concentration or a disease score (Depciuch et al., 2018). This method is particularly useful for quantitative analysis, for example, predicting the concentration of a specific metabolite or the degree of differentiation in a cell culture.

2.10.3 Performance and interpretation

The successful application of multivariate data analysis is crucial for the diagnostic and quantitative performance of FTIR microspectroscopy. These methods enable the differentiation of samples with high sensitivity and specificity, often exceeding 90% (Baker et al., 2014). However, proper application requires careful data pre–processing (e.g., baseline correction, normalization) to remove artifacts and a deep understanding of the biological and chemical context to correctly interpret the statistical output (Reis et al., 2020). The results from these analyses can reveal the subtle biochemical changes that are the hallmarks of disease or cellular processes, making chemometrics an indispensable component of modern FTIR microspectroscopy.