QUANTIFYING BREAST CANCER ADAPTATION: PLATFORMS FOR STUDYING TRANSCRIPTIONAL CHANGES IN SINGLE CELLS

Debjani Saha

Department of Surgery and Cancer

Faculty of Medicine

Imperial College London

ABSTRACT

Hormone-dependent breast cancer (HDBC) is the most commonly diagnosed female malignancy. Patients are administered adjuvant endocrine therapies (ET) which induces temporary remission. However, up to 40% of patients relapse. Recent studies indicate that ET drives cells into a bottleneck where surviving micro-disseminated cells enter a reversible cell-cycle arrest – dormancy –from which, with time, they can asynchronously and spontaneously awaken *via* non-genetic cell state transitions. However, the exact mechanisms by which dormant cells awaken and re-enter the cell cycle remain poorly understood. Recent advances in single-cell technologies have enabled detailed cancer studies at unprecedented cellular scales. Unfortunately, traditional sequencing methods, are limited by their need to lyse cells and in doing so, provide only snapshots, preventing the tracking of dynamic transcriptional changes in the same cell over time. This limitation hinders our understanding of how cancer cells adapt and survive treatment, especially in HDBC, where dormancy and eventual recurrence are common.

This project bridges a key technological gap, presenting the development of multiple techniques that can ultimately be used to probe cells at a sub-cellular level; introducing two minimally invasive platforms using nanotweezers and nanopipettes, enabling the sequential single-cell sampling of cytoplasmic RNA without resultant cell death. These techniques rely on dielectrophoresis to trap highly expressed RNA and cytosolic volume, respectively. By facilitating continuous monitoring, these approaches capture dynamic transcriptional changes, crucial for understanding dormancy and adaptation in HDBC. Additionally, given the challenges experienced in dynamic technology development, alternative static approaches that yield insight into these states and variations in hormone-receptor expression with age, with spatial resolution, have also been optimised and presented. With further work, these techniques should enable the generation of a dynamic, longitudinal model of gene expression and the processes that underlie adaptation in HDBC, ultimately offering new perspectives on cancer progression and resistance.

STATEMENT OF ORIGINALITY

I hereby declare that the work presented in this thesis is my own, and is otherwise appropriately referenced and/or acknowledged where required. This work has not previously been submitted to satisfy any degree requirements in any form at this or any other institution.

COPYRIGHT DECLARATION

Debjani Saha,

1st November 2024

ACKNOWLEDGEMENTS

Firstly, a massive thank you to Professor Luca Magnani, Professor Joshua Edel and Dr Aleksandar Ivanov, for their support and guidance throughout this PhD, and the opportunity to embark on this challenging, but equally enjoyable journey.

To all the members of the Magnani Lab, past and present, thank you for creating an environment filled with curiosity and laughter and for always believing in me. A special thank you to Hannah for being there for me, through the struggles of failed experiments, to the many failures of our failed awakening experiments – this PhD journey would not have been the same without you. To Claudia, for the many coffee outings, the laughs, the rants and the beautiful friendship – thank you for everything.

To all the members of the Edel and Ivanov Labs, especially Binoy, Tony, Annie and Zoe, for all your help with the nanotweezer experiments. Alex D-L, for being such a good mentor and friend, and always having the wisest words to share.

Thank you to all the collaborators that have made this work possible – I really could not have done this without you.

To my friends, Cherry, Jo, Sona, Sara and Max – you have all been amazing, listening to my struggles, dealing with my stress and always being there should I need a friend!

Thank you, Edward, for being my number one fan and the best cheerleader one could ask for.

Bhaiya, thank you for always bringing your immaculate sense of humour to the party. Didu and Dada, thank you for always showering me with love!

Finally, my parents – I could not have done this without your endless love and support. Thank you for always encouraging me to become the best version of myself. I cannot express the immense gratitude I hold in my heart for you.

THESIS OVERVIEW

Chapter 1 provides a brief overview of cancer, how cancers (specifically breast cancer), can evolve and adapt to therapy, and highlights the need for dynamic analytical techniques to fully probe the many states that cells can adopt and processes that underlie this shift from therapy to resistance. Existing dynamic single-cell analytical techniques, which are all capable of extracting material from cells, or measuring events across cell membranes, without impacting cell viability, are outlined. The work conducted in this thesis attempts to further explore dynamic single-cell analytical studies, develop understanding of the epigenetics and chromatin architecture of newly identified characteristics of dormancy and quantify variations in ERα expression in tumours with age. Each chapter contains a brief introduction containing relevant background information, materials and methods used for all pertaining experiments, and respective results, discussions and conclusions. Chapters 2, 3 and 4 additionally contain graphical summaries, with a brief comparison of alternative techniques and how the chosen technique described in the chapter compares.

Chapter 2 works on developing the minimally invasive, dielectrophoresis-based nanotweezer platform developed by Nadapurram *et al.*, expanding its use to the extraction of mRNA from MCF-7 cells and attempting to increase their trapping capacity using broad-spectrum surface functionalisation[1]. Difficulties in transcript detection were troubleshooted by optimising reagent combinations, pooling multiple samples together and validating gene localisation using RNAScope. Due to the complexity of transcript detection and gene quantification from mRNA extracts obtained using this technique, experiments using nanotweezers were discontinued, and alternative dynamic and static techniques were attempted.

Chapter 3 presents one such alternative dynamic single-cell analytical technique, combining patch-clamp measurements with angular, hopping scanning ion conductance microscopy and electroporation. Similar to the nanotweezer platform, this technique provides a minimally invasive method to extract and sample sub-cellular contents. Cytoplasmic access was validated in a number of configurations by injecting the cells with fluorescein and extracting material by application of various magnitudes of mechanical pressure or voltage pulses. The ability to extract material using this technique was characterised by specifically detecting low concentrations of λ-phage-DNA diluted within MCF-7 cell lysate.

Given the difficulty experienced in detecting material extracted from cells using the nanotweezer platform, outlined in chapter 2, a nucleic acid amplification technique was attempted. **Chapter 4** outlines various attempts at optimising a protocol for Nucleic Acid Sequence Based Amplification of RNA, with the hope that trapped mRNA could be amplified prior to downstream processing, thus enabling greater detection from the relatively small extracts. Despite the use of previously validated reagent combinations and attempts at optimising enzyme ratios, this method proved unsuccessful.

Chapter 5 outlines the development of a multi-omic pipeline for the probing of newly identified dynamic processes occurring during dormancy – failed awakening, wherein treatment-induced dormant cells transiently re-enter the cell cycle before rapidly regressing to form a new dormant persistor population, notably distinct to the final resistant cell population - as outlined by Rosano *et al.*[2] Using a reporter cell line (developed by Hannah Dewhurst), enabling the tracking of cells through the cell cycle, a novel protocol was developed, wherein both the transcriptome and chromatin architecture of cells at different cell cycle phases could be explored from the same single cell.

As dynamic cell experiments proved little success, another single-cell technique, utilising an RNAScope assay was attempted. **Chapter 6** presents a HiPlex RNAScope assay, exploring variations in *ESR1* and *FOXA1*

expression and influence of age on their expression in healthy and tumour tissue within archival samples obtained from patients with luminal breast-cancer. While only three targets were analysed for the purpose of this thesis, this chapter provides a basis for further exploration and expansion of the target panel with a hope to understand how expression of key oestrogen-responsive genes and variations in their expression between healthy and tumour tissue counterparts can impact disease aggression, ultimately providing insight into possible avenues for therapeutic intervention. The work in this chapter presents a ground-truth for larger spatial transcriptomic assays to be conducted in the future.

Finally **Chapter 7** summarises overall conclusions from this work and future outlooks to enable optimisation of techniques outlined in chapters 2 and 3 for expansion to a wider pool of samples with increased detection, possible routes to amplify RNA to increase starting detection amounts, key steps that need to be adopted to enable a cohesive multi-omic pipeline for probing cells' transcriptomes and chromatin architecture and future prospects and implications of understanding variations in oestrogen expression with age. Overall, the experiments conducted in this thesis provide a basis for the development of both dynamic and static sub-cellular exploration, upon further expansion and optimisation.

TABLE OF CONTENTS

LIST OF FIGURES

hybrid technique that combines atomic force microscopy with microfluidics to enable the mechanical puncture and extraction of defined volumes of subcellular material using defined forces; **D** Nanotube endoscopy – gold nanoparticle-coated carbon nanotubes inserted into cells that act as channels for molecular transfer and **E** Nanostraws – vertically aligned hollow alumina nanostructures, integrated into the membrane of cells to facilitate passive and electroporation-driven molecular transport from live, adherent cells.[1,89–92] Notably, by ensuring cells are kept alive throughout sampling, these techniques have revolutionised longitudinal sampling of individual cells within heterogenous populations. While sub-cellular sampling techniques have only recently been coupled with downstream omics applications such as single-cell transcriptomics, the underlying engineering of these methods has been evolving since the early 2000s. Each technique offers unique advantages in terms of spatial precision, volume control, throughput and compatibility with downstream molecular analysis. However, successful application of these techniques in cellular contexts is heavily dependent on the biological context and cell type, requiring careful optimization of sampling parameters. Key features of the aforementioned techniques are outlined in **table 2.1**. This thesis focuses on techniques A and B: with nanotweezers explored in chapter 2 and nanopipettes, coupled with SICM and electroporation, explored in chapter 3. Fluid force microscopy (C) is referenced for comparison, and techniques D and E are not referenced further in this work, though their recent advancements are acknowledged herein in the broader context of nanobiopsy development. Despite promising early results, reproducibility remains a significant challenge due to the heterogeneity of live-cell systems and the ultra-low input levels involved. Throughout this thesis, difficulties encountered during sampling and amplification are critically examined, and potential solutions are proposed to improve both the robustness and analytical utility of nanobiopsy-derived samples. Panel **F** illustrates advanced applications of nanotweezers in live hippocampal neurons, as demonstrated by Sahota et al. **i)** EGFP-filled hippocampal neurons were sequentially samples using DEP at three subcellular regions: (1) axons, (2) dendrites and (3) somas (nanobiopsy sites marked in orange). **(ii)** 18S rRNA was successfully amplified from biopsied material via RT-qPCR and copy numbers of 8 replicates were quantified by region against no template controls. **(iii)** Longitudinal tracking capabilities were demonstrated by measuring MAP1A expression six neurons over 11 days, with repeated sampling at both soma and dendrites. This was supplemented by the analysis of CAMK2A expression at dendritic spines before and after NMDA/glycine stimulation **(iv)**. While this study highlights the versatility of this platform and the ability to target different spatial compartments and track changes with time, as discussed further in chapters 2 and 3, this application is limited by low RNA yield, sampling bias due to transcript localisation and reliance on RT-qPCR for signal detection. Top figure panel created with Biorender. Figures adapted with permission from S. G. Higgins et al., 2017 and Sahota et al., 2025.[93,94] Copyright (2017) Science & (2025) ACS Nano. 56

Figure 2.2 **Electrophoretic and dielectrophoretic forces exerted on charged and neutral species in uniform and non-uniform electric fields. A** A charged particle in a uniform electric field experiences EP forces, while the neutral particle will not experience a net force. **B** However, when in a non-uniform electric field, the neutral particle experiences an overall net force, DEP, as the magnitude of forces acting on the induced dipoles is unequal. Charges are shown in white on the particles. Figure adapted with permission from J. Voldman, 2006.[95] Copyright (2006) Annual Reviews. 58

Figure 2.3 **Overview of laser-assisted pulling protocol to generate double-barrelled nanopipettes.** Capillaries are aligned to grooves in the laser puller and clamped in place. A CO2 laser is then directed at the capillary to heat up the glass. The laser puller then sequentially progresses through the five parameters, according to pre-defined programme settings. Line 1 gently heats and pulls the capillary to yield a conical mid-section, which is further heated and pulled to generate nanopipettes in line 2. 61

Figure 2.4 **Bespoke carbon deposition set-up. A** A double-barrelled nanopipette is secured onto a moving holder and attached to a tygon tube, connected to a butane inlet. The holder is moved to position the tip of the nanopipette inside a single-barrelled capillary, connected to an Argon inlet. Gas flows are regulated as the **B** tips and **C** length of the nanopipette is heated to enable pyrolytic carbon deposition and yield carbon electrodes. 62

Figure 2.5 **Nanotweezer functionalisation.** Nanotweezer (NT) surfaces were functionalised with polyT linkers, in order to increase binding of cytoplasmic RNA, which comprise a complementary poly-adenine (polyA) tail. NTs were incubated in the

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Aromatase inhibition
AC	Alternating current
AFM	Atomic force microscopy
aL	Attolitres
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
AMV	Avian myeloblastosis virus
bp	Base pairs
cDNA	Complementary DNA
CO2	Carbon dioxide
Ct	Cycle threshold
d	Inter-electrode distance
DABCYL	4-((4-(dimethylamino)phenyl)azo)benzoic acid
DC	Direct current
DCIS	Ductal carcinoma in situ
DdRp	T7 DNA-dependent RNA polymerase
DEL	Delay
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DMEM	Dulbecco's modified eagle medium
E	Amplification efficiency
E2	Oestradiol
ER	Oestrogen receptor
ETs	Endocrine therapies
F	Functionalised
FCS	Foetal calf serum
FIL	Filament
HeLa	Henrietta Lacks
H&E	Hematoxylin and eosin
FISH	Fluorescence in situ hybridization
FFM	Fluid force microscopy
FFPE	Formalin-fixed paraffin-embedded
fL	Femtolitres
FLASH-seq	Full-length single-cell RNA sequencing
fps	Frames per second
FUCCI	Fluorescent ubiquitination-based cell cycle indicator
Gem21-MCF-7	Geminin-mCherry p21-GFP MCF-7
GFP	Green fluorescent protein
GFP_{high}	GFP high
GFP_{low}	GFP low
GΩ	Giga-ohm

G&T-seq	Genome and transcriptome sequencing
HPAECs	Primary human pulmonary artery endothelial cells
I	Current
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
I_{lim}	Limiting current at electrode in presence of excess supporting electrolyte
I-V	Current-voltage
LAMP	Loop-mediated isothermal amplification
LCIS	Lobular carcinoma in situ
LNA	Locked nucleic acid
LPM	Litres per minute
LPS	Lipopolysaccharide
LSV	Linear sweep voltammetry
M	Molarity
MCF7	Michigan cancer foundation 7
MCF-7-nucGFP	Nuclear GFP-tagged MCF-7 cells
mCherry+	mCherry positive
mCh+GFP+	mCherry and GFP positive
MIND	Mouse intraductal
MRD	Minimal residual disease
mM	Millimolar
NASBA	Nucleic acid sequence-based amplification
NEB	New England Biolabs
NND	Nearest neighbour deconvolution
nt	Nucleotides
NT	Nanotweezers
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
pHe	Picolitre
pM	Extracellular pH
pL	Pico molar
PolyA	Poly-adenine
PolyT	Poly-thymine
PUL	Hard pull
RCA	Rolling circle amplification
RNA	Ribonucleic acid
RT	Reverse transcriptase
RTed	Reverse transcribed
RTion	Reverse transcription
$Ru(NH_3)_6)Cl_3$	Ruthenium (III) hexamine chloride
R^2	Goodness of fit

sci	Single-cell combinatorial indexing
scRNA-seq	Single-cell RNA sequencing
SERDs	Selective oestrogen receptor degraders
SERMs	Selective oestrogen receptor modulators
SICM	Scanning ion conductance microscopy.
SNR	Signal to noise ratio
S3-ATAC-seq	Symmetrical strand sci ATAC-sequencing
T	Tail
TNBC	Triple negative breast cancer
T75	*75* cm^3 cell culture flask
TIRF	Total internal reflection fluorescence
UF	Unfunctionalised
UMI	Unique molecular identifiers
V_{pp}	Peak-peak voltage
V	Voltage applied to electrodes
VASA-seq	Vast transcriptome analysis of single cells by dA-tailing sequencing
VEL	Velocity
WE	Working electrode
WM	White media
λ-DNA	Lambda-phage DNA
6-FAM	6-carboxyfluorescein
µM	Micromolar

CHAPTER 1
INTRODUCTION

1.1 Cancer and its Hallmarks

Cancer is a disease characterised by the uncontrolled growth of cells, which constantly evolve to develop local and distal lesions of various sub-types, possessing unique characteristics, both within individual tumours and at a patient level. At present, cancers are the leading cause of premature deaths worldwide.[3] According to the Global Cancer Observatory, 19.3 million new cancer cases and close to 10 million cancer deaths were recorded in 2020 alone, with the greatest cancer incidence attributed to breast cancer, diagnosed in approximately 2.3 million women worldwide, followed by lung, colorectal and prostate cancers.[4] Furthermore, cancer related deaths are predicted to increase to close to 20 million annually by 2060.[5] It is important to note that mortality is not necessarily a linear measure, as this can occur many years after diagnosis, However, the global healthy emergency and need for more effective therapies to eradicate this disease becomes evident when considering the increased risk of developing cancer with age, especially those of epithelial origin and driven by endocrine secretion, and the projected rise in cancer incidence in years to come.[4,6]

Cancers generally arise from the dysregulation of key genes and processes that maintain normal cellular function, growth and division. In transforming from a healthy to a cancerous state, cells can acquire somatic mutations, in a complex multi-step process, which subsequently drive their characteristic proliferative phenotype, although this can also be induced by inherited mutations in cancer predisposition genes.[7] These genetic alterations can occur in oncogenes, or tumour-suppressor genes, disrupting normal function and evasion of otherwise tightly regulated cell cycle control. As a cancer increases in size, mutations within cells tend to accumulate, resulting in significant genetic drift within the cell population and in vast heterogeneity.[8] Given this dynamic nature, while numerous therapies have been developed to combat this group of diseases, treatment is often hindered by the emergence of resistance, as cancers evolve from their original genetic states. This is complicated further by cells which are able to escape from their primary site and form distal lesions, in a process called metastasis. During this process of adaptation, cancers progress through a series of alterations, eventually rendering them capable of evading therapy and enabling eventual disease recurrence. These features that characterise cancers have been collated by Hanahan and Weinberg, resulting in the generation of a collection of 'Hallmarks of Cancer', which summarise the key driving mechanisms adopted by cancers to survive and thrive (figure 1.1).[9] These include the ability to do the following: **(1)** replicate infinitely and **(2)** proliferate without exogenous signalling, with the ability to bypass key cell cycle checkpoints; **(3)** evade cell death (apoptosis); **(4)** resist the effects of growth suppressors; **(5)** induce and sustain the formation of new blood vessels (angiogenesis); and **(6)** invade tissue, and metastasise at distal sites. This list was modified approximately a decade later to account for two new enabling characteristics, which facilitate acquisition of the original hallmarks – including **(1)** generation

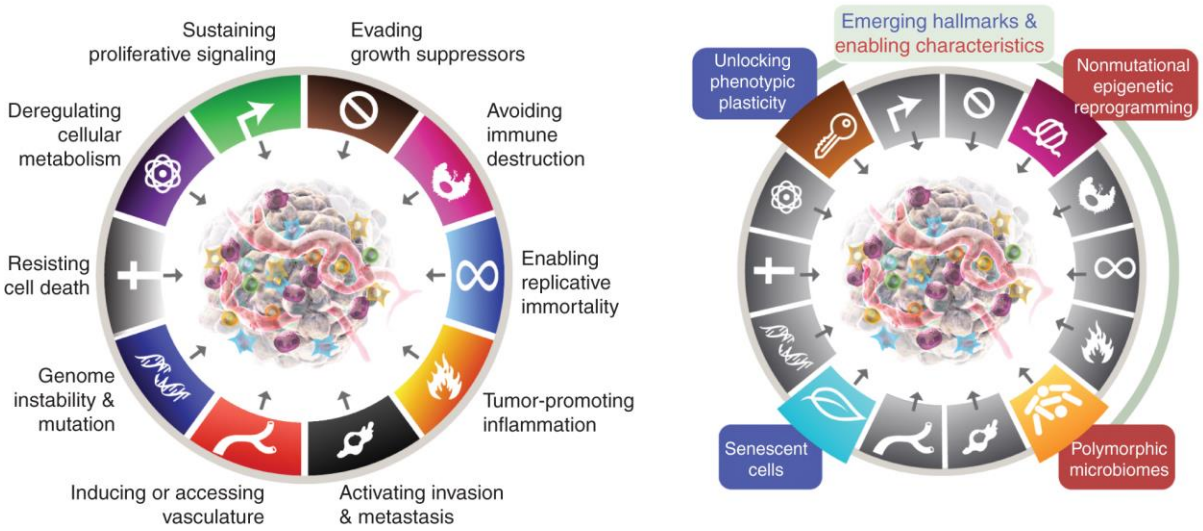


***Figure 1.1* Hallmarks of Cancer as summarised by Hanahan, 2022,** *highlighting the 14 main mechanisms, including enabling characteristics and new emerging hallmarks, that drive tumourigenesis. Figure taken from D. Hanahan et al., 2022.*[10]

of random mutations caused by genetic instability and **(2)** inflammation-driven tumourigenesis - and two new emerging hallmarks, which could also play a crucial role in cancer development - **(7)** resisting apoptosis by evading immune control and **(8)** dysregulation of cellular metabolism to sustain proliferation.[11] However, much like cancer itself, this list of hallmarks is ever evolving, as new research continues to uncover the complexities of this family of diseases, standing at 14 hallmarks as of 2022.[10]

1.2 Breast Cancer

Breast cancer is the most prevalent malignancy diagnosed in women worldwide, accounting for approximately 30 % of total female cancers.[4] Notably, since 1993, while the number of total cancer cases has risen in the UK, likely due to the implementation of more efficient screening programmes, the total proportion of breast cancer cases has remained stable and is projected to stay such even in 2035 (figure 1.2).[12] As with cancer, breast cancer is an umbrella term for a group of diseases, with different subtypes possessing different molecular, phenotypic and morphological characteristic features. However, of these, only 5-10 % are caused by a genetic predisposition, with mutations classified according to their population frequency and risk: *high-penetrance mutations* are rare but associated with high risk while *low-penetrance mutations* which are relatively omnipresent, but present little risk of developing disease.[13,14] The most commonly known of these high-penetrance mutations, occur in the tumour-suppressor genes, BRCA1 and BRCA2, wherein small nucleotide (nt) insertions or deletions result in the translation of truncated proteins, resulting in up to 70 % of women with this mutation developing breast cancer by the age of 80.[15] Risk in the remaining 90-95 % of tumours is

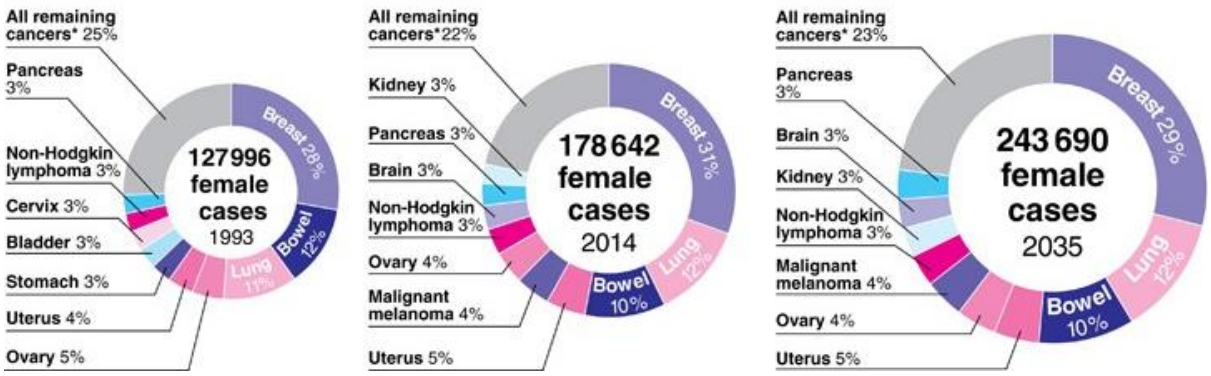


***Figure 1.2** Summary of historical number of cancer cases in the UK in 1993 and 2014, and projected number of cases in 2035.The size of the doughnuts corresponds to the number of total cases diagnosed, with the proportions of the various types of constituent cancers represented on the individual charts. For all three years shown, breast cancer cases comprised the largest percentage and remain stable around 30%. Figure taken from C. R. Smittenaar et al., 2016.*[12]

conferred by somatic mutations in low-penetrance genes, such as *TP53*, *PIK3CA* and *MYC*.[16] Furthermore, given the hormone-dependence of breast development, as will be described in section **1.2.1**, the risk of developing breast cancer is strongly influenced by prolonged hormone exposure. Key factors that contribute towards this include the early onset of menstruation, late menopause, late pregnancy, use of hormonal birth control and hormone replacement therapies.[17]

1.2.1 The Breast

Before delving into the different subtypes of breast cancer, it is necessary to provide a brief overview of the breast architecture. The breast is a bilateral organ, unique to and highly conserved across mammals, that undergoes significant morphological changes and development in response to hormone expression at key timepoints including puberty, menarche, pregnancy and menopause.[18] The mammary gland, responsible for the production and delivery of milk after birth, is housed within the breast and is comprised of a network of 15-20 lobes, each with a series of smaller protrusions called lobules and terminal bulb-like glands, connected *via* ducts (figure 1.3). These ducts are formed of a double-layered epithelial cell structure, with an internal polarised luminal epithelial layer surrounded by an outer basal myoepithelial layer. The aforementioned structures are supported by adipose tissue and interspersed with a complex network of blood and lymphatic vessels, which not only supply the breast with key nutrients and facilitate lymphatic drainage, but also serve as a primary means for the dissemination of cancer cells throughout the body. Breast and mammary development are known to be a multi-step process, largely initiated at puberty, as hormone expression begins to fluctuate. At this stage, the duct termini enlarge to yield their characteristic bud morphology, which initiates the proliferative cycle of breast maturation. The ducts continue to grow every menstrual cycle until they

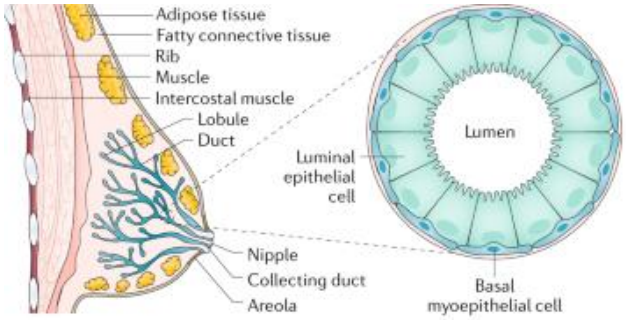


***Figure 1.3* Structure of the breast and duct architecture.** *The mammary gland is comprised of numerous lobular structures connected via ducts, which adopt a double-layered epithelial structure. A complex network of blood and lymph vessels surround this structure, which is housed within the fatty tissue of the breast. Figure reproduced with permission from N. Herbeck et al., 2019.*[18] *Copyright (2019) Springer Nature.*

have penetrated the fat pads before branching out to yield their final branched structure.[19] This branching is further enhanced during pregnancy, wherein prolactin secretion also increases, in preparation for lactation. Interestingly, studies have shown that early pregnancy imparts a protective effect against the development of breast cancer, as pregnancy results in the long-term depression of prolactin secretion.[20]

1.2.2 Breast Cancer Subtypes and Therapeutic Implications

Breast cancers are generally characterised using two distinct classification techniques, according to their hormone receptor status (more common) or histologically. Four different hormone-receptor subtypes exist, depending on the expression of the oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth receptor 2 (HER2), each accompanied with unique risks of disease progression and varying aggressiveness and response to treatment (figure 1.4 A).[21] Cancers expressing ER and PR (ER+/PR+) are classified as being of a luminal subtype, those expressing solely HER2 (ER-/PR-/HER2+) are classed as non-luminal, HER2+ cancers, and those that express none of the three markers are referred to as triple negative (ER-/PR-/HER2-; TNBC).[22] Luminal-type cancers can be further subdivided depending on expression PR and the proliferation marker, Ki67, with luminal A having high PR and low Ki67 expression, resulting in slow growth, and luminal B having low PR and high Ki67 expression, thus exhibiting high proliferation potential. Distinguishing between these two subtypes is especially important when determining treatment regimens, as luminal A cancers tend to be more responsive towards hormonal therapies and subsequently have better overall prognosis. In general, ER+ breast cancers are treated with anti-oestrogenic therapies, which either inhibit oestradiol synthesis (aromatase inhibitors, AI), degrade the ER (selective oestrogen receptor degraders, SERDs), or block the ER preventing ligand binding (selective oestrogen receptor modulators, SERMs).[23] As HER2+ cancers are characterised by the significant overexpression of the HER2 receptor on epithelial cell surfaces, therapy entails treatment with the monoclonal antibody, trastuzumab, which interferes with the key dimerization step that initiates downstream proliferative signalling.[24] In contrast, due to the lack of receptor expression and distinct molecular targets in TNBC, there are no targeted therapies available for this subtype. Given that only 20 % of patients diagnosed with TNBC respond to standard chemotherapy, this subtype has the worst prognosis of all aforementioned breast cancer subtypes.[21]

Classification according to the histological subtypes considers the morphology, invasiveness of the tumour mass and where it originates (figure 1.4 B).[25] Cancers that originate in the glandular lobes of the breast are referred to as lobular tumours, while those that occur in the milk ducts are known as ductal tumours, which, when in their non-invasive states are referred to as lobular carcinoma *in situ* (LCIS) and ductal carcinoma *in situ* (DCIS), respectively. However, both of these carcinomas are able to expand and invade tissue beyond their sites of origin, although occurring more frequently with invasive ductal carcinoma (IDC; in about 40 % of tumours) than in invasive lobular carcinomas (ILC; 5-15 % of tumours).[26,27] While these two molecular subtypes differ in terms of morphology, with IDCs adopting more fibrous characteristics, both are able to metastasise *via* the bloodstream and lymph nodes, resulting in the formation of distal lesions.

1.2.3 The Oestrogen Receptor

Oestrogen signalling has previously been shown to contribute towards numerous cancer hallmarks, resulting in onset of more aggressive diseases, including evasion of cell death, inflammation- driven

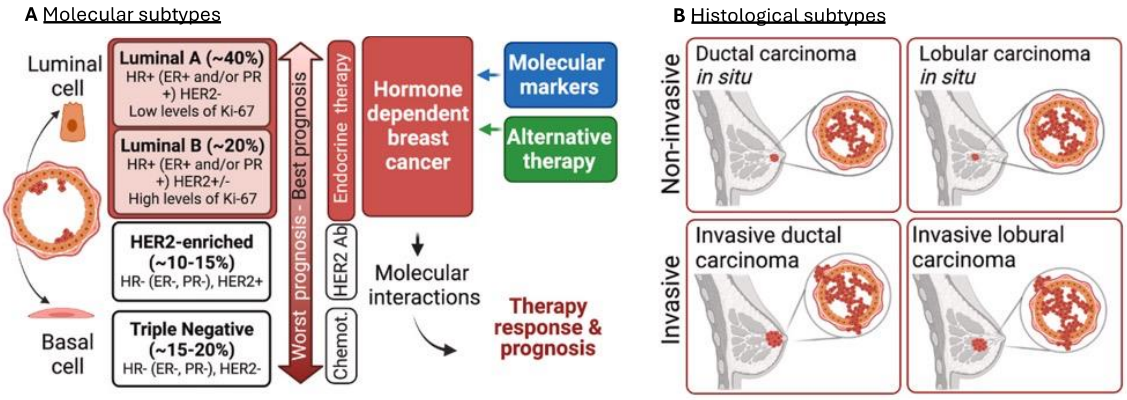


Figure 1.4 ***Breast Cancer Classification according to A molecular and B histological subtypes.*** ***A*** *Four main molecular subtypes exist, each with varying oestrogen (ER), progesterone (PR) and human epidermal receptor 2 (HER2) expression. Prognosis varies with receptor expression, with those expressing ER and PR (luminal-type) having the best prognosis and triple negative having the worst. Standard therapeutic methods are also shown.* ***B*** *Cancers can also be defined according to where they originate in the breast, either in the ducts (ductal carcinoma in situ) or in the lobes (lobular carcinoma in situ). Both are capable of expanding outside of the ductal and lobular threshold and invading nearby walls – invasive ductal and invasive lobular carcinoma. Figure taken from T. Kunstič et al., 2023.*[25]

tumourigenesis and promotion of invasion and metastasis.[28–30] These processes are particularly significant in breast cancer, given 70 % of diagnosed cases adopt some, if not total, reliance on oestrogenic compounds, whose expression is modulated by the ER, for tumourigenesis.[25] The importance of oestrogen in regulating mammary function has been demonstrated in both mice models and humans, wherein lack of ERα or mutations in the gene encoding ERα, *ESR1*, resulted in stunted differentiation and maturation.[31,32] However, this same dependency that drives normal breast development, is also responsible for carcinogenesis, promoted by prolonged exposure. While two ER isoforms exist, α and β, only the α sub-type is present within the mammary gland and has implications on its development. Oestrogens, predominantly 17β-oestradiol (E2), bind to ERα, causing dimerization which enables its translocation into the nucleus, wherein binding to promoter regions of oestrogen response elements induces transcription (genomic) and downstream signalling (non-genomic).[33,34] It is through this signalling pathway that expression of essential oestrogen-responsive genes is modulated. Furthermore, ERα binding to chromatin can be facilitated by a set of key transcription factors, termed pioneer factors and include genes such as *FOXA1* and *GATA3*.[35] These function by independently binding to condensed chromatin and decompacting them, thus facilitating binding of other transcription factors, including ERα. *FOXA1*, in particular, is of notable importance in breast cancer, as it has been shown to contribute towards the induction of key cell cycle activator genes, including *CCND1*, which promotes cell proliferation.[36]

1.2.4 Endocrine Therapies and Resistance

Given the notable E2 dependency of ER+ breast cancers, therapeutic intervention mostly relies on preventing E2 synthesis, blocking effects of E2 on certain cells or degrading the ER, termed aromatase inhibitors (AI), selective ER downregulators (SERDs) and selective ER modulators (SERMs), respectively. Combined, these constitute a class of therapies, called endocrine therapies (ETs). AIs, including Anastrazole, Letrozole and

Examestane, are administered to both pre-and post-menopausal women, and function by preventing E2 synthesis by blocking aromatase, the key enzyme used to convert its precursor hormone, androgen, into E2.[37] As oestrogen production is still rife in pre-menopausal women, upon AI administration, the hypothalamic-pituitary axis compensates by increasing gonadotropic releasing hormone (GnRH) secretion, which stimulates the release of luteinising and follicular stimulating hormones, subsequently stimulating the ovaries to produce more oestrogen. Therefore, to avoid this negation of the effects of AI, pre-menopausal women are also administered GnRH agonists, which induce temporary menopause, suppress ovarian function and thus prevent this increased oestrogen production.[38] SERDs, including Fulvestrant, competitively binds to the E2-binding site in the ER, wherein they exert their antagonistic action by preventing ER dimerisation and its subsequent nuclear translocation.[39,40] In doing so, transcriptional activity is completely halted, and the ER rapidly degrades. SERMs, such as Tamoxifen, also bind to the ER, although exert their action by causing a conformational shift that blocks the receptor coactivator site, thus preventing E2 binding.[41] A schematic of their modes of function is shown in figure 1.5.

Patients are generally treated with five years of adjuvant therapy, post-surgical resection. Nonetheless, treatment, in up to 50 % of patients with ER+ breast cancer, will fail due to the development of resistance, with risk of relapse increasing at a steady rate from 5 to 20 years post-treatment.[42,43] This can be increased to up to seven years for patients at intermediate risk of relapse, and up to ten years for patients with a high risk of recurrence.[44] To further delay this onset of resistance, patients are generally treated with a combination of the aforementioned and CDK 4/6 inhibitors such as Ribociclib and Abemaciclib. Results from the recent NATALEE trial, which compared the combination of non-steroidal AIs with either Ribociclib or placebo revealed an improvement in invasive disease-free survival of over 3 % with 3 years of treatment, alongside improvements in distant disease-free survival and recurrence-free survival.[45] Additionally, the MONARCH 3 trial, which explored Abemaciclib or placebo with a non-steroidal AI, revealed an improvement in median progression-free survival by 13.4 years.[46] However, the overarching large risk of recurrence highlights the need to obtain a more comprehensive overview of the specific processes underlying this shift from response to relapse. Resistance can be classified as either *de novo* or acquired. *De novo* resistance, which, by contributing to minimal residual disease (MRD) wherein very few cells can remain in the body even after treatment, can lead to the emergence of more aggressive, acquired drug resistance. Studies have revealed that this resistance can be conferred by the overexpression of certain genes, such as CYP19A1 - which encodes for aromatase, enabling increased E2 production even in the presence of AI - or by mutations, resulting in decreased enzymatic activity - as with CYP2D6 which converts Tamoxifen into its active component, endoxifen.[47,48] Further mutations in key genes encoding the ER, including in the ligand-binding domain of *ESR1*, have also been shown to significantly reduce the efficacy of ETs, in up to 20 % of metastatic ER+ disease.[49] While mutations are the cause of resistance in certain individuals, a large clinico-genomic analysis conducted by Razavi *et al.* revealed that only 40 % of resistance is conferred by mutations in key driver genes.[50] The remaining 60 % could not be associated with any genetic drivers, suggesting non-genetic factors, including epigenetic processes, may be at play. This was further backed up in a similar, but much smaller study by Yates *et al.*, which found that 55 % of metastatic lesions did not have genetic drivers, although those that did, were not always unique to the metastatic setting.[52] However, while both of these studies identified the existence of non-genetic drivers of metastasis, they both focused solely on panels of cancer-associated genes, and as such, *de*

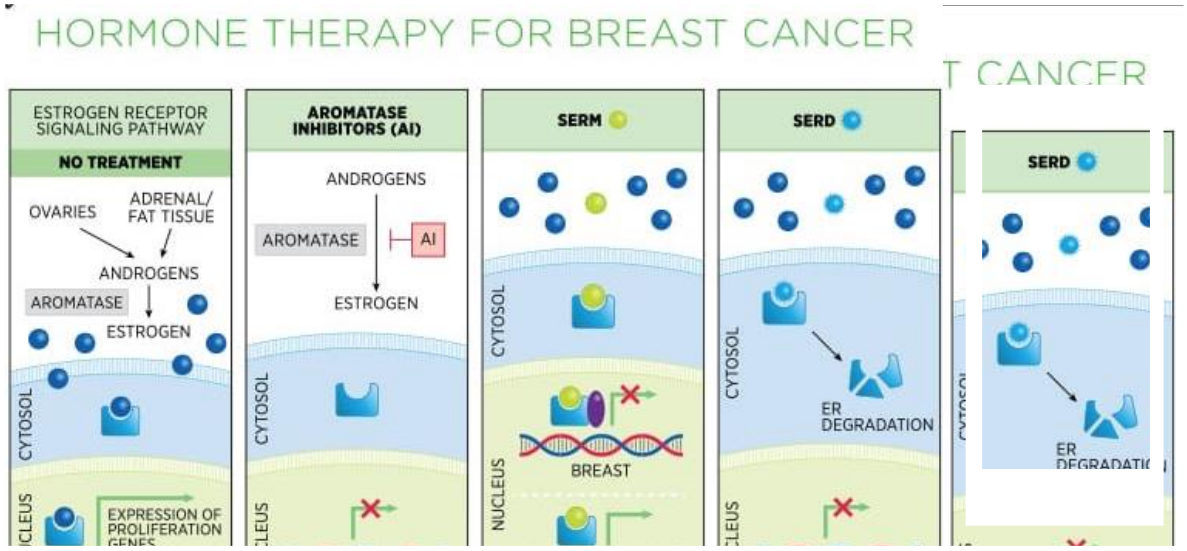


***Figure 1.5* Schematic overview of the three main classes of endocrine therapies, and their cytosolic or nuclear targets along the oestrogen receptor (ER) signalling pathway.** *Aromatase inhibitors, which are administered to post-menopausal patients, inhibit function of aromatases, which convert androgen precursors to oestrogen (E2). In doing so, E2 concentrations within the body are essentially depleted, preventing binding to the ER and subsequent expression of proliferation genes. SERMs essentially function as allosteric inhibitors, which bind to the ER and prevent E2 binding by inducing a conformational shift. When bound to the ER present in the breast, this induces the recruitment of transcriptional corepressors, which in turn prevents expression of proliferation genes. SERDs function similarly, in that they also bind to the ER, but prevent ER dimerisation and subsequent translocation into the nucleus, resulting in ER degradation. Figure taken from Endocrine Therapy in Breast Cancer: Making Sense of the "Word Salad" (web).*[51]

novo mutations outside of this core panel could have been missed. Notably, even though both metastatic and primary or matched healthy samples were compared in these studies, they are inherently limited by the lack of dynamic measurements. Given the vast heterogeneity that exists within cancers, effectively tracking the molecular mechanisms that underlie the evolutionary pathways requires sampling of more than a primary sample and a metastatic lesion. This complexity is

further compounded by the fact that MRD often remains undetectable, and that cells that are ultimately responsible for distant metastases depart from the primary tumour at very early stages of cancer evolution, leading to significant phenotypic drift over time.[53]

While therapeutic intervention has proven successful in kerbing initial disease progression and offering apparent curative treatment, metastasis remains a major challenge. Owing to the large heterogeneity and sub-cellular diversity attained by cells as they genetically diverge from the cell-of-origin, metastatic disease, which can emerge even decades after initial diagnosis, remains largely incurable.[54] This is reflected in the survival probability of females with breast cancer, which drops from 95.1 % one-year post diagnosis to 80.2% and 68.4 % five- and ten-years post diagnosis, respectively.[55] Combined, therapeutic failure and development of resistance, highlights the need to further develop our understanding of the processes that underlie and regulate progression to metastasis in breast cancer.

1.2.5 Evolution of Hormone-Dependent Breast Cancer: Dynamics and Dormancy

Breast cancer is a complex disease with diverse subtypes, marked by both inter- and intra-tumoral heterogeneity, making its study highly challenging. To better understand critical bottlenecks in this disease, we will investigate the evolution of hormone-dependent breast cancer in response to therapy, focusing on the dormant cancer life cycle (figure 1.6 A). Prior to diagnosis, cancer cells divide and grow exponentially within the body and disseminate from the primary site to yield micro-disseminated lesions, resulting in the formation of a largely

heterogenous mass. Adjuvant endocrine therapies are then administered for up to five years post-surgical resection of the primary tumour mass, which induces a therapeutic bottleneck. Certain cells, such as those that have disseminated and now occupy distal regions, or those with a pre-adapted survival fitness, are able to survive this, and enter a so-called hibernation state, called dormancy.[56] In this state of MRD, cells, notably stalled in the cell cycle, undergo modifications with time, that confer the ability to develop resistance to and evade therapy, eventually leading to the onset of metastasis. Given that these metastatic lesions are derived from cells that already possess innate heterogeneity, and have previously been shown to disseminate early in the processed of tumorigenesis, transcriptional reprogramming in the transition from a dormant to an awakened, metastatic phenotype, results in the generation of a highly dynamic and heterogenous population landscape.[56,57]

The mechanisms that underlie evolution during dormancy are unclear but have recently been proposed to be the result of epigenetic reprogramming by a number of groups.[58–60] An example of this has been outlined in a recent Cancer Discovery paper from our group, which studies the long-term evolution of MCF-7 and T47-D cells in response to AI and Tamoxifen treatment.[2] By transcriptionally barcoding the founder population, with varying barcode frequencies highlighting heterogeneity in replicative fitness in the different lineages, dominant lineages could be identified and compared across different analytical timepoints. This notably revealed that cellular replicates diverge considerably over time, even when seeded simultaneously from the same progenitor pool, demonstrating variability in their adaptation to treatment. Whole-genome sequencing confirmed the absence of *de novo* mutations, pointing instead to epigenetic mechanisms as key drivers of adaptation and eventual therapy resistance. The study found that dormancy exhibits marked epigenetic reprogramming, with a distinct increase in histone marks associated with heterochromatin (H3K9me2, H3K27me3 and H4K20me) and a reduction in acetylation marks (H3K4me3 and H3K9/14ac) during dormancy entrance (figure 1.6 B). This condensed DNA state partially reverses during early and late metastatic progression but remains distinct from the untreated epigenome, suggesting a potential role in promoting cellular awakening. Furthermore, awakening dynamics were found to be distinct across different replicates, occurring stochastically and with different winner barcode populations. Many cases of "failed awakening" were also observed, wherein cells initially began proliferating but quickly regressed (the concept of this is explored further in chapter **5**). Dormancy was therefore defined to be an unstable, treatment-induced state while awakening from this state is marked by asynchronous dynamics, driven by processes involving marked epigenetic erosion and rewiring.

Taken alone, this study highlights the need for non-disruptive sequencing techniques, to fully decipher the processes that underlie and regulate dormancy and resistance. While possessing the capacity to probe cellular dynamics at multiple timepoints, these must ensure that the sampled cell remains intact and viable for future measurements and thus enabling more longitudinal sampling.

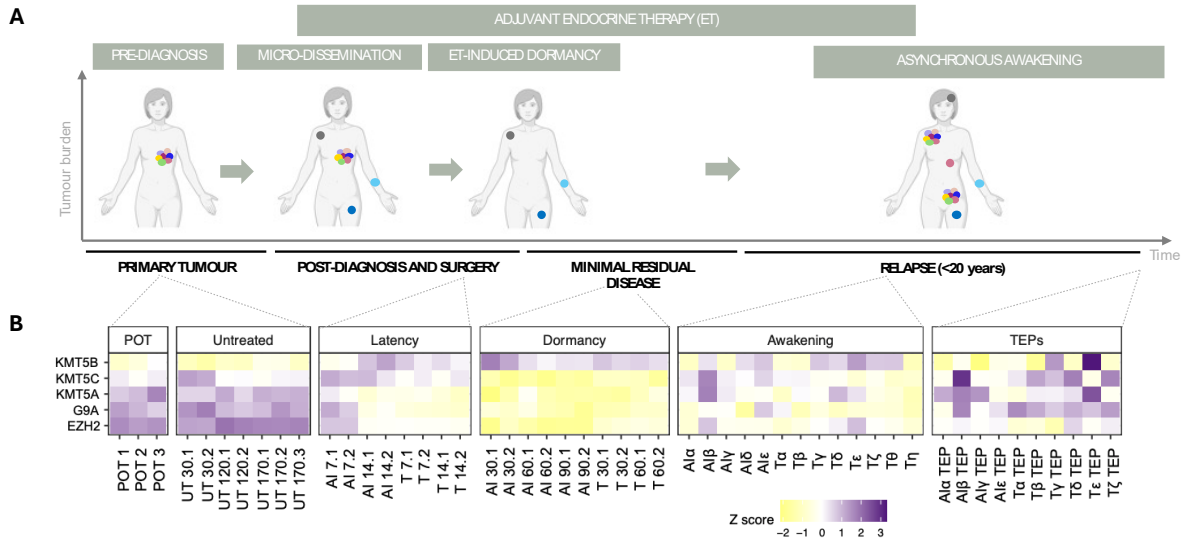


***Figure 1.6** Evolution of Hormone-Dependent Breast Cancer. **A** Patients with hormone-dependent breast cancer are generally treated with up to five years of adjuvant endocrine therapy, administered post-surgical resection of the primary tumour. These induce a transient stalling in the cell cycle, causing persistor cells and micrometastatic lesions to enter a state of dormancy where they remain clinically invisible for up to 20 years. Behind the scenes, cells can adapt to this therapeutic intervention by acquiring gain-of-function mutations or reprograming of their epigenetic machinery, enabling them to evade therapy and awakening asynchronously. **B** More and more studies are now addressing an epigentic contribution to this reversal of dormany. A recent study by Rosano et al. highlighted repressive chromatin states in dormant samples, across both aromatase inhibitor (AI) and Tamoxifen (T) treated MCF-7 and T47-D cells. These marks were notably reversed upon awakening (early relapse) and in terminal end points (TEPs; late relapse), both with different characteristics to their untreated counterparts. Figure reproduced from Rosano et al., 2024.[2]*

1.3 Dynamic single-cell studies

One significant drawback of current single-cell analytical and sequencing techniques used to explore cells' transcriptomes, mutational burden and chromatin architecture, is the need to remove cells from sometimes long-term culture, and subject them to harsh lysis conditions to access cellular material. This results in only static "snapshots" from each extraction point, from which past and future trajectories, can only be inferred. The effect of this is especially pronounced in the study of epigenetics, which is governed by transient processes. Current sequencing techniques only allow for epigenetic modifications to be studied at certain timepoints. Further timepoint studies must then be carried out in parallel replicates. Piecing together a complete overview of cellular dynamics therefore relies on the assumption that results obtained from these sequential experiments accurately reflect continuous cellular changes. However, given the inherent heterogeneity within cell populations, particularly in cancer cells, this assumption is unreliable, as due to the strong influence of evolutionary pressures, it is likely that subsequent measurements may diverge from earlier states. This variability can result in significant discrepancies, making it challenging to map an uninterrupted trajectory of cellular evolution over time, highlighting the need for more dynamic single-cell analytical systems. Notably, these must have the capacity to operate in a broad spectrum of conditions, depending on the cells native culture and sub-cellular conditions. Furthermore, as analyte concentrations can vary not only between different cell types but also different treatment conditions (with enzymatic reactions occurring at a micromolar (μM) to millimolar (mM) range, and biosensing on a sub-picomolar (pM) range), systems must exhibit stability within these environments for prolonged durations.[61] When working in a live-cell system and conducting dynamic measurements, maintaining viability is of paramount importance. Extractions must either be conducted on a femtolitre scale or require pre-concentration of material without extracting any cytoplasmic volume. This places a significant concentration limit on these techniques, often requiring high operating voltages or involving

prolonged exposure – both of which can induce undesired stress of heating, ultimately impacting the accuracy of read-outs or harming the cells themselves.

1.3.1 Existing sequencing techniques: the transition from bulk to single cell

The elucidation of the three-dimensional structure of DNA by Watson and Crick in 1953 opened the doors to the limitless possibilities of nucleic acid sequencing.[61] Despite this breakthrough, developing DNA sequencing techniques proved challenging due to the inherent similarity between nucleotide bases, resulting in a 15-year delay before the first DNA sequence was achieved.[62] Significant progress was however made in the late 1970s, through the development of two new DNA sequencing techniques: the chemical sequencing method developed by Maxam and Gilbert, and the dideoxy technique (better known today as Sanger Sequencing), by Sanger, Nicklen and Coulsen, the subsequent automation of which was crucial for the completion of the Human Genome project.[63,64] However, DNA sequencing machines utilising these techniques typically generated reads just under one kilobase in length.[65] It took another 18 years to sequence the first complete bacterial genome, a milestone that spurred the sequencing of other genomes, including those of eukaryotes. Further advancements included the development of pyrosequencing, which notably offered the advantage of using natural nucleotides instead of the heavily modified nucleotides required in chain-termination methods, such as Sanger Sequencing.[66] While automation enabled massive parallelisation of sequencing reactions and significantly increased throughput, pyrosequencing was still limited to generating reads of only 400 – 500 base pairs in length. Introduction of paired-end sequencing techniques by Illumina further enhanced the accuracy of output data, aiding the identification of spliced exons and fused genes. Variations of these DNA sequencing techniques include chromatin immunoprecipitation sequencing (ChIP-seq) and assay for transposable-accessible chromatin with sequencing (ATAC-seq), which probe the interaction between proteins and DNA and genome-wide chromatin openness, respectively, enabling genome-wide mapping of the chromatin regulatory landscape at high-resolution. However, in their most rudimentary formats, these techniques only provide bulk measurements, producing a single, average signal and masking heterogeneity within complex sample populations. This has partially been addressed by development of a new third generation of sequencing techniques, focusing mainly on single-cell level detection. The most renowned of these is nanopore sequencing (now commercialised by Oxford Nanopore Technologies). Building on previous advances made in detection and quantification using nanopores, this technique utilises the principles of detection using electrophoresis and current blockade, to yield a raw read accuracy of 99 %.[67] In general, single cell sequencing resolution is attained by dissociating cells into their individual cellular components. To ensure that the identity of these single cells can be retained and distinguished between during downstream analysis, cells are uniquely barcoded. This enables multiple cells to be processed in a single-reaction and sequenced together, with high-throughput. Whole-genome amplification techniques have also been developed in parallel, enabling the fragmentation and amplification of the complete genome of single-cells, with minimal error rates.[68]

A similar trend can be observed in the evolution of RNA-sequencing techniques to probe cells' transcriptomes (figure 1.7). Interestingly, sequencing efforts initially focused on RNA, as it avoided the complexity of a complementary strand, could be mass-produced in vitro, and enzymes for its processing were readily accessible.[65] However, limitations in analytical techniques at the time, resulted in a temporary slowing in progress. Nonetheless, by combining existing knowledge of the structure of DNA, the notion that RNA contained a different nucleotide base and generating RNA fragments, Robert Holley and colleagues successfully managed to generate the first complete nucleic acid sequence of alanine transfer RNA from *Saccharomyces cerevisiae* in 1965.[69] Despite this success, RNA sequencing was not developed until 2007.[70] Two types of bulk RNA sequencing techniques exist:

mRNA-only and whole- transcriptome sequencing.[71] These approaches can use both single-end and paired-end sequencing, with paired-end allowing for longer reads. This extended sequencing capability not only reveals insights into the molecular mechanisms that underlie diseases and their progression, but also provides information on gene fusion, splice variants, point mutations, somatic and acquired mutations. This heterogeneity is often masked, and as such, low-frequency variations,

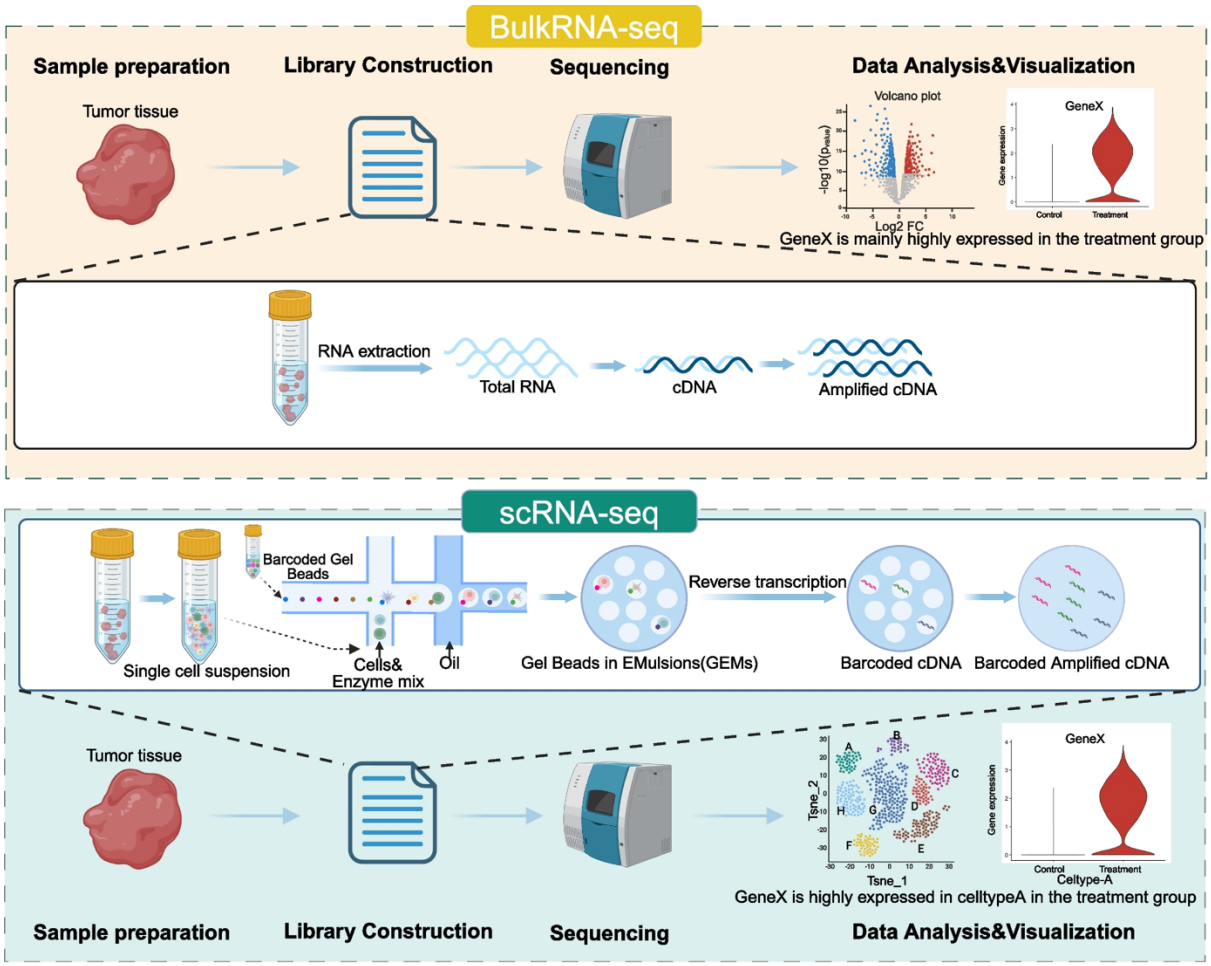


***Figure 1.7* Overview of differences between bulk and single cell RNA sequencing techniques.** *In both cases, samples (shown here as tumour tissue) are extracted from their native environments. library can be generated from these samples without the need for prior pre-processing. However, bulk sequencing captures only an average signal, leading to a loss of granularity and preventing the association of specific features with individual cells or cell types. Single cell library preparation is somewhat more tedious given the need to generate a single cell suspension and uniquely barcode each cell to enable downstream deconvolution of the cells' identities. This has been optimised using microfluidic techniques, by 10X Genomics. In doing so, the transcriptional profile of single cells can be dissected and associated to individual cells, offering substantially more detailed gene expression profiles and identification of features unique to rare sub-populations. Figure taken from M. Hagemann-Jensen et al., 2020.*[72]

including those that are unique to therapy resistant populations to go undetected, prompting the development of single cell technologies a mere two years after bulk RNA sequencing was introduced. As with single-cell DNA sequencing, single-cell RNA sequencing requires the generation of a heterogenous single cell pool and barcoding to enable downstream deconvolution. RNA from cells must first be reverse transcribed into its complementary DNA (cDNA) and amplified to yield sufficient sample for efficient detection before coupling with a whole transcriptome amplification technique. Rapid advancements in this field have led to the development of techniques such as Smart-seq3, which, much like its predecessors, enables high-resolution single-cell transcriptome profiling, albeit with increased accuracy in capturing full-length transcripts, enabling improved detection of low-abundance transcripts

and increased resolution.[72] Development of the 10X genomics chromium system has further revolutionised this field and its applications, providing the ability to analyse the gene expression profiles of up to 30,000 cells in a single assay. Further variations include nascent RNA sequencing, which, by tagging nascent RNA, enables differentiation between nascent and total RNA, allowing a measure of active transcription and gene activity in cells.[73] Unfortunately, current sequencing techniques are riddled with limitations, most notably the need for samples to be removed from their native growth conditions and subsequent dissociation which can have negative implications on the transcriptional profiles being explored.

The resolution of sequencing techniques has improved drastically since their first discovery in the 1970s. Initially limited to bulk sampling, which provided only an average population signal, thus masking phenomena unique to certain subpopulations, these methods can now analyse individual cells and distinguish between single cells at both the RNA and DNA levels. Further advancements have led to spatial technologies involving fluorescence *in situ* hybridisation (FISH), including RNAScope and seqFISH+, and more recently Nanostring's CosMx and GeoMx, which allow for the tagging and quantification of many thousands of RNA transcripts in fixed cells.[74,75] However, given their *in situ* nature, these only provide snapshot measurements, which significantly limits their use in complex, dynamic systems with inter- and intra-cellular heterogeneity. While advancements in sub-cellular sampling methods offer promising avenues in this regard (discussed further below), these must be paired with improvements in next-generation sequencing sensitivity, as the amount of extracted material is notably reduced at this scale. Sub-cellular sampling also introduces additional challenges, particularly as the impact of dynamic processes and diffusion can no longer be disregarded, which is especially evident when exploring enzymatic reactions, transcription, and reversible processes. Nonetheless, success has been attained in sampling just a portion of the cells cytoplasm.[76] Although numerous sub-cellular sampling techniques are currently in development, only a few, which are relevant to this study and showing the most promise will be discussed in more depth.

1.3.2 Subcellular sampling using nanopipettes

A large proportion of sub-cellular sampling techniques have been derived from nanopipettes (NPs) - that is, tapered quartz capillaries with nano-sized apertures, generally fabricated using laser-assisted pulling. The first successful cellular sampling and subsequent sequencing was reported by Actis *et al.* in 2013 (figure 1.8 A).[76] Herein, they extracted RNA and mitochondria from Human Fibroblasts and HeLa cells, respectively, using a combination of previously optimised techniques: (i) scanning ion conductance microscopy (SICM), for spatial control of the NP with regard to the cell; (ii) extraction using electrowetting.[76,77] A more comprehensive overview of SICM is provided in section **3.1.2**, which in essence relies on the NP being pre-filled with an electrolyte solution, the flow of which upon application of a voltage bias, results in a measurable ion flow. Moving the NP over a sample results in variations in the resulting current trace due to current blockage, indicative of the topography of the sample. An electrowetting modality can be incorporated into this platform by ensuring that the electrolyte is of an organic nature.[78] Upon immersing into an aqueous solution, a liquid-liquid interface is generated at the NP aperture. Application of a negative voltage across this interface results in a change in the surface tension of the organic medium, causing the aqueous solution to flow into the NP. The direction of flow is reversed when the sign of the voltage is reversed. Of the 50 nanobiopsies conducted as part of this study, 70% yielded detectable cDNA concentrations from approximately 1 % of the cells volume, as observed by post-PCR gel electrophoresis. Whole transcriptome RNA libraries were subsequently generated from the cytoplasmic extracts and mapped to the human reference genome while mitochondria were analysed using next-generation sequencing techniques. Notably, two different isoforms of the cytoplasmic Poly-A binding protein (*PABC1*) were identified in three individual aspirations, while analysis of the most overexpressed Gene Ontologies revealed notable variation in genes related to

translation-related machinery and metabolism. Furthermore, analysis of the mitochondrial DNA sequencing results highlighted the presence of heteroplasmic variants. Combined, this study highlights the ability to not only extract and sequence minute amounts of RNA and mitochondrial DNA at a sub-cellular level, but also differentiate between the presence of different gene isoforms and monitor mitochondrial mutation rates, while notably maintaining cell viability. This platform has since been expanded for use in combination with matrix-assisted laser desorption/ionisation-mass spectrometry, yielding characterisation and chemical analysis of extract components.[79] Additionally, this technique involved extraction using positive pressure, as opposed to a potential bias, which enabled more controlled sample collection affording increased spatial control.

The applications of NPs are vast. Previously employed to measure the electrophysiology of synaptic boutons and to trap molecules within lipid bilayers, combining their sampling capabilities with additional sensing modalities further expands their potential uses.[80,81] This can be achieved by increasing the number of compartments within the pipette, transitioning from a single-barrel to a multi-barrel configuration. Examples of this include scanning electrochemical conductance microscopy (SECM), enabling the detection of hydrogen ions, to measure extracellular pH (figure 1.8 B).[82] These pH sensors were fabricated by cross-linking two pH sensitive molecules, glucose oxidase and poly-l-lysine at the tip of a NP. Permeability of differentially charged ions across this membrane is favoured at different pHs, with negative ions favoured in low pH conditions and positive ions at high pH conditions. This platform was subsequently used to map the topography and measure pH gradients

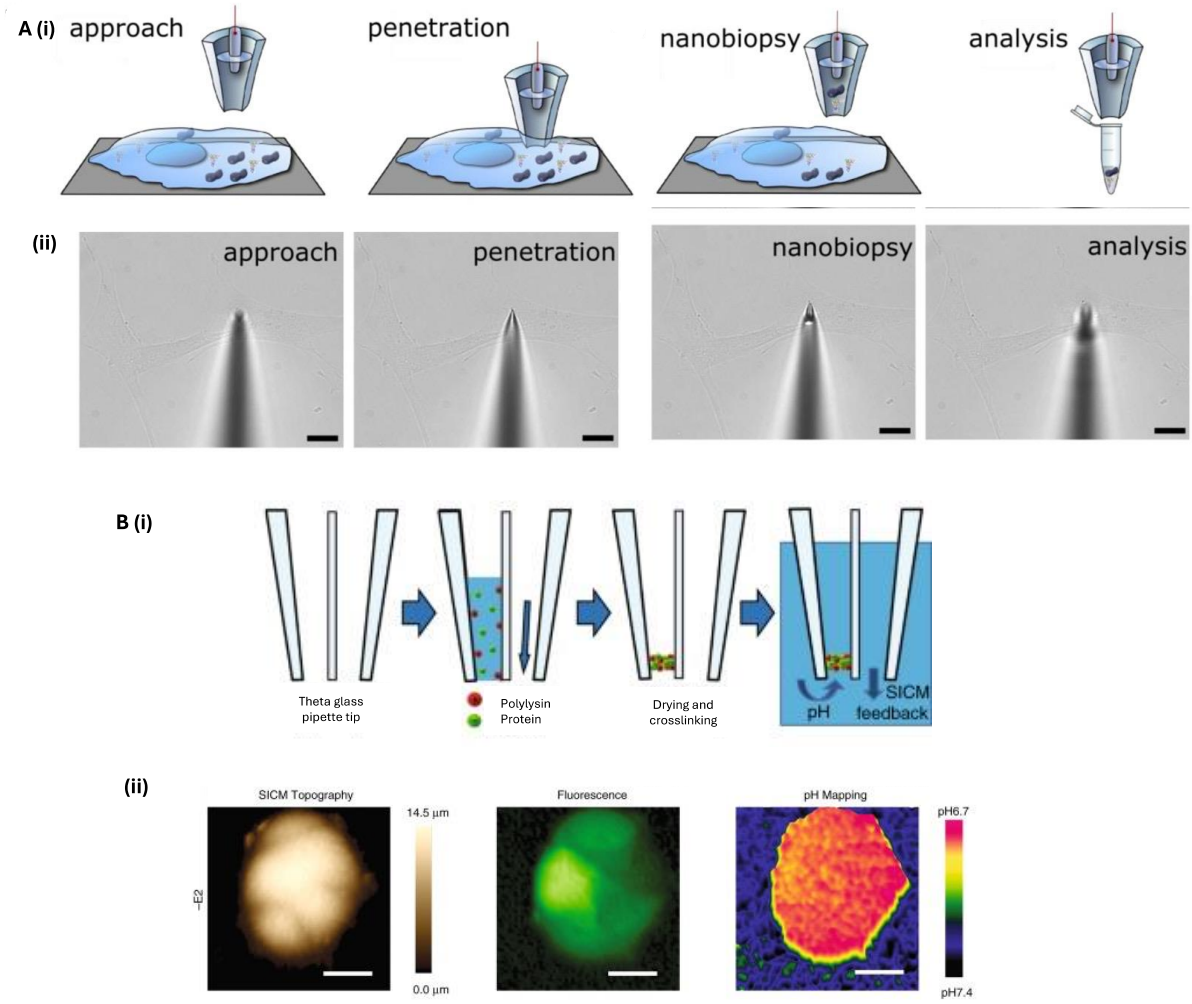


Figure 1.8 ***Various NP modalities.*** ***A*** *Schematic of a single cell nanobiopsy, showing (i) an illustration of approach to the cell, penetration using the NP, aspiration of material via electrowetting and subsequent analysis. (ii) Corresponding optical micrograph images are shown, highlighting the NP out of the plane of the cell upon approach, penetration of the cell membrane, aspiration and final retraction from the cell. Scale bars represent 25 µm. Reprinted with permission from A. Shevchuk et al., 2001.*[76] *Copyright 2014 American Chemical Society.* ***B*** *NPs can be conjugated with chemical moieties to enable electrochemical measurements in addition to topographical mapping.* ***(i)*** *An example of the chemical conjugation of a glass pipette with poly-L-lysine and protein to enable pH measurements in combination with SICM feedback is shown.* ***(ii)*** *Resulting SICM topography, fluorescent imaging of the cell and mapping of the extracellular pH of an MCF-7 cells, under oestrogen deprived conditions, is shown. Figure adapted from Y. Zhang et al., 2019.*[82] *Copyright (2019) Springer Nature.*

at the surface of MCF-7 cells that had been previously engineered to express GFP in the presence of CD44 expression. Interestingly, a pH gradient was identified only for cells cultured in oestradiol-deprived conditions, with those cultured in oestrogen-supplemented conditions showing no variation in pH, further highlighting the notion of inter-cellular heterogeneity. Notably, while NPs offer a minimally invasive platform for sub-cellular sampling, while maintaining cell viability, owing to the small aperture sizes, certain studies have reported that in order to enable efficient cytosolic extract, the cell must be penetrated at least 3 µm below the cell membrane.[83] As such, this technique cannot be used in combination with flat cells. Fine spatial control can also be imparted by mounting the NPs onto micromanipulators, enabling not only cytoplasmic sampling but also single organelle manipulation. However, while subsequent sequencing of nanobiopsy extracts has been successful, this technique may be limited by its extraction of cytoplasmic fluid, which may require further purification and inevitable loss of the already little amounts of material, prior to further library preparation, in turn making low abundance transcript detection difficult.

1.3.3 Atomic Force Microscopy-based nanobiopsies

Another prevalent sub-cellular sampling technique involves the use of atomic force microscopy (AFM) probes. AFM is a powerful imaging technique that provides the ability to map the surface of an object with unprecedented sub-nanometre (nm) resolution and signal-to-noise ratios.[84] The AFM probe comprises a tip at the end of a cantilever that can respond to oscillatory cues. By controlling the forces acting between the tip and a sample, AFM can contour surfaces as the probe is moved across the sample. Variations of this technique include operation within aqueous solutions, dynamic mode - wherein the cantilever oscillates at its resonance frequency to minimise collision with the sample - and high-speed AFM, which acquires images up to 1000 times faster than standard AFM, allowing for the visualisation of dynamic biological processes (figure 1.9 A). Notably, in dynamic mode, upon approaching the sample, this resonance frequency and its amplitude changes, enabling a feedback loop to be generated, with the probe tip only touching the sample at the very end of its downward trajectory, minimising friction between the tip and the sample - especially beneficial when working with fragile biological samples.[84] Integrating AFM with fluid chambers has expanded its capabilities, enabling not only imaging but also probing and manipulation of materials for a range of applications.

The first report of using an AFM-based technique for probing cellular contents was made in 2003, by Osada *et al.*[85] *β-ACTIN* expression in extracts obtained from rat-fibroblast-like VNOf90 cells and mouse osteoblast-like MC3T3-E1 were explored independently using reverse-transcriptase polymerase chain reaction (RT-PCR). In summary, the AFM tip was approached to the surface of the cells and indented beyond the lowest point in the cantilever's downward trajectory to penetrate the cell. The tip was inserted close to the nucleus and mRNAs were extracted by non-specific surface contact. Of the 189 assays conducted in both cell models, 182 were positive for *β-ACTIN* expression, albeit to varying degrees. Notably, cells were maintained within their standard culture vessels and conditions throughout the whole process, highlighting that it is in fact possibly to conduct nanobiopsies from samples and detect sufficient material, while maintaining cell viability. Samples which did not yield *β-ACTIN* expression were postulated to result from poor positioning of the AFM tip within the cell. Nonetheless,

this technique was the first of its kind to enable mRNA detection within living cells in their native physiological environment, while ensuring minimal cell damage and high reproducibility.

A novel application, which has now been incorporated into a commercial platform, builds upon this technique of extracting material from cells, but does so using hollow cantilevers enabling cytoplasmic sampling (figure 1.9 B). This technique, which is aptly termed fluid force microscopy (FFM), allows controlled extraction and delivery of substances and even organelles, to and from single cells, under physiological systems.[86] Pressure-driven micro-channelled silicon nitride probes are incorporated into the AFM tip, generally of pyramidal geometry and approximately 400 nm in diameter (figure 1.9 C).[87] This was first successfully explored in 2016 by Guillaume-Gentil *et al*. As with AFM, the tip, whose aperture size can be modulated for precise control over extractions, is lowered into the cell until it has been penetrated, although material is now extracted by application of a positive pressure and collected within the hollow cantilever channel. The post-extraction viability of HeLa cells was monitored using fluorescence and revealed that even after the extraction of approximately 2.9 pL of cytoplasm, cells subjected to extractions behaved almost identically to adjacent non-extracted cells, with both counterparts exhibiting the same rounding and cell division dynamics. Numerous extractions, wherein the authors extracted up to 90 % of the cytoplasm and 20 % of nuclear volume revealed 82 % and 86 % cell viability after a single extraction, with cells reversing back to their normal sizes within five days. While this does highlight the resilience of cells and their ability to recover from the removal of such large proportions of their volume, this could have significant implications on their gene expression profiles, which was not explored in this context. More recent advancements, made by the same group, has resulted in a technique called Live-seq, combining the capabilities of FFM with subsequent RNA sequencing using Smart-seq2 from as little as 1 pg of total RNA.[88] Increased sensitivity was attained by **(i)** pre-preloading the FFM probe with lysis buffer, significantly reducing the time between extraction and initiation of processing, **(ii)** supplementing the aforementioned lysis buffer with RNase-inhibitors and **(iii)** avoiding sample dilution as far as possible to minimise RNA degradation. Notably, in addition to probing and confirming the same metrics as in their previous study with regard to cell recovery after extraction, Chen *et al*. sampled the same single cell at two distinct timepoints, with only four hours separating the sequential probing – one of the earliest examples of such. Macrophage-like RAW cells sampled before and after stimulation with lipopolysaccharide (LPS), and extracts sequenced with

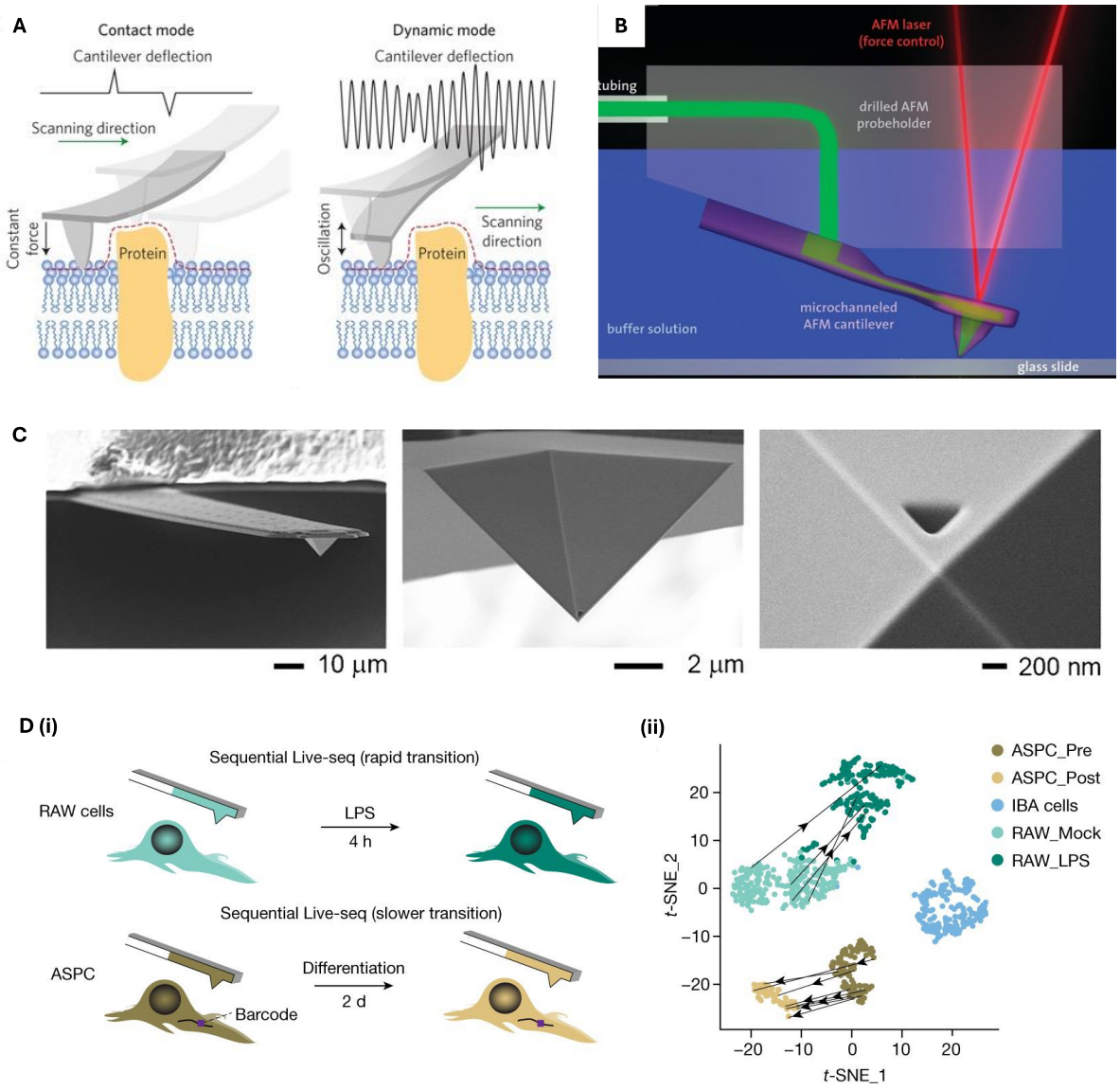


Figure 1.9 ***Overview of Atomic Force Microscopy (AFM) and variations, enabling single cell sampling. A*** *AFM probes are scanned over samples generate high resolution images with high signal-to-noise ratios. Cantilever deflection can be modulated in two ways, either in contact mode, wherein the cantilever is in constant contact with the sample, or dynamic mode, where the cantilever oscillates at resonance frequency and distance between the sample and tip is regulated by a feedback loop. Use of AFM in dynamic mode minimises damage to soft samples. Figure reproduced with permission from Y. Dufrêne et al., 2017.[83] Copyright (2017) Springer Nature.* ***B*** *Schematic showing a variation of AFM, fluid force microscopy (FFM). The cantilever tip is hollowed out to incorporate a microchannel, enabling extraction of cytoplasmic fluid from the cell. Figure adapted with permission from A. Meister et al., 2009.[86] Copyright (2009) American Chemical Society.* ***C*** *Scanning electron micrograph images of the FFM cantilever and probe are shown. The probe comprises a 400 nm pyramidal tip with a nano-scale aperture for extraction of fluid from cells. Figure adapted with permission from O. Guillaume-Gentil et al., 2016.[87] Copyright (2016) Elsevier.* ***D*** *Live-seq was born from the integration of FFM and Smart-seq2.* ***(i)*** *Two different cell types were sequentially probed to assess dynamics either on a short (top) or long (bottom) timescale.* ***(ii)*** *Extracts were sequenced and trajectories of probing timepoints were mapped. Figure taken from W. Chen et al., 2022.[88]*

Smart-seq2 (figure 1.9 D). Of the ten cells samples, only 40 % exceeded the stringent quality control cut-offs. Despite this, high-quality reads that mapped correctly to their cell state clusters could be produced, which uncovered that *NFKBIA* expression contributes to phenotypic heterogeneity in LPS-stimulated macrophages (also validated using conventional single-cell RNA sequencing). Overall, this approach enables the linking of a cell's current state to downstream phenotypic states, effectively mapping its dynamic trajectory. It allows for minimally invasive, sequential probing of single cells, albeit at low efficiency. Nonetheless, Live-seq provides an alternative means of coupling single-cell extractions with RNA sequencing techniques, while maintaining cell viability.

1.3.4 Single molecule trapping and nucleic acid extraction using nanotweezers

Despite the relative novelty of sub-cellular sampling techniques, extraction methods are slowly moving away from using the traditional single-barrelled nanopipette set-up. Notably, use of dielectrophoretic (DEP)-based techniques for extraction of sub-cellular material and single-organelle manipulation has become more popular in recent years. A novel example of this technique being used was reported in 2017 by Naddapuram *et al*, who developed carbon-coated double-barrelled nanopipettes, termed nanotweezers (NTs), to manipulate and extract single molecules from solution and cells.[1] A more detailed overview of DEP, how it varies from electrophoresis, and how it can be used to trap molecules are provided in sections **2.1.1** and **2.1.2**, respectively. In brief, these DEP-based NTs comprise two individually addressable barrels, separated by a nanosized septum (figure 1.10 A). Generated in a similar fashion to nanopipettes, laser-assisted pulling parameters can be carefully controlled to the decrease the inter-electrode distance to sub-20 nm. Carbon is deposited onto the inner surfaces of the nanopipettes using pyrolysis to form a conductive surface, which upon establishing a closed circuit and application of an alternating current (AC) voltage, results in the generation of large electric field gradients (on the order of 10^{28} V^2 m^{-3}; figure 1.10 B). The DEP force then acts on any polarisable material in its near vicinity resulting in its trapping and concentration at the aperture tip, which is notably lost back into solution as the DEP force acting across the electrodes is turned off. In this work, the authors demonstrated the ability of NTs to conduct single-cell biopsies by extracting DNA and RNA from the nuclei of primary human pulmonary artery endothelial cells (HPAECs) and demonstrated single organelle manipulation by extracting single mitochondrion from primary rodent hippocampal neurons. Individual nanobiopsies were conducted from the nuclei and cytoplasm of HPAECs, to extract DNA and RNA, respectively (figure 1.10 C). DNA was qPCR amplified for a target sequence in *45S* ribosomal DNA, while mRNA was initially reverse transcribed to cDNA, which was subsequently qPCR amplified for two low abundance transcripts, *ETS-1* and *KLF-2*, and one high abundance transcript *β-ACTIN* (figure 1.10 D). Both *45S* and *β-ACTIN* were detected reliably well from the DNA and mRNA extracts, respectively. However, due to the inherent low abundance of *ETS-1* and *KLF-2* and the low capture volumes of the NT platform, these were not detected from the mRNA extracts. Cell viability was also monitored after these biopsies to rule out any structural damage as a result of NT penetration. Combined, this highlights the generation of an alternative minimally invasive, subcellular sampling technique with high spatial resolution all while maintaining cell viability. The NT platform is unique in its ability to isolate polarisable material from cellular compartments without the need to extract cytosolic or nuclear volume. Coupled with the functionality of DEP, which is able to generate extremely large electric fields without unwanted heat generation (explored further in section **2.1.1**), this offers a significant advantage over nanopipette- and AFM-based techniques, by minimising resultant stress responses. The work outlined in this section will be built upon in **Chapter 2.**

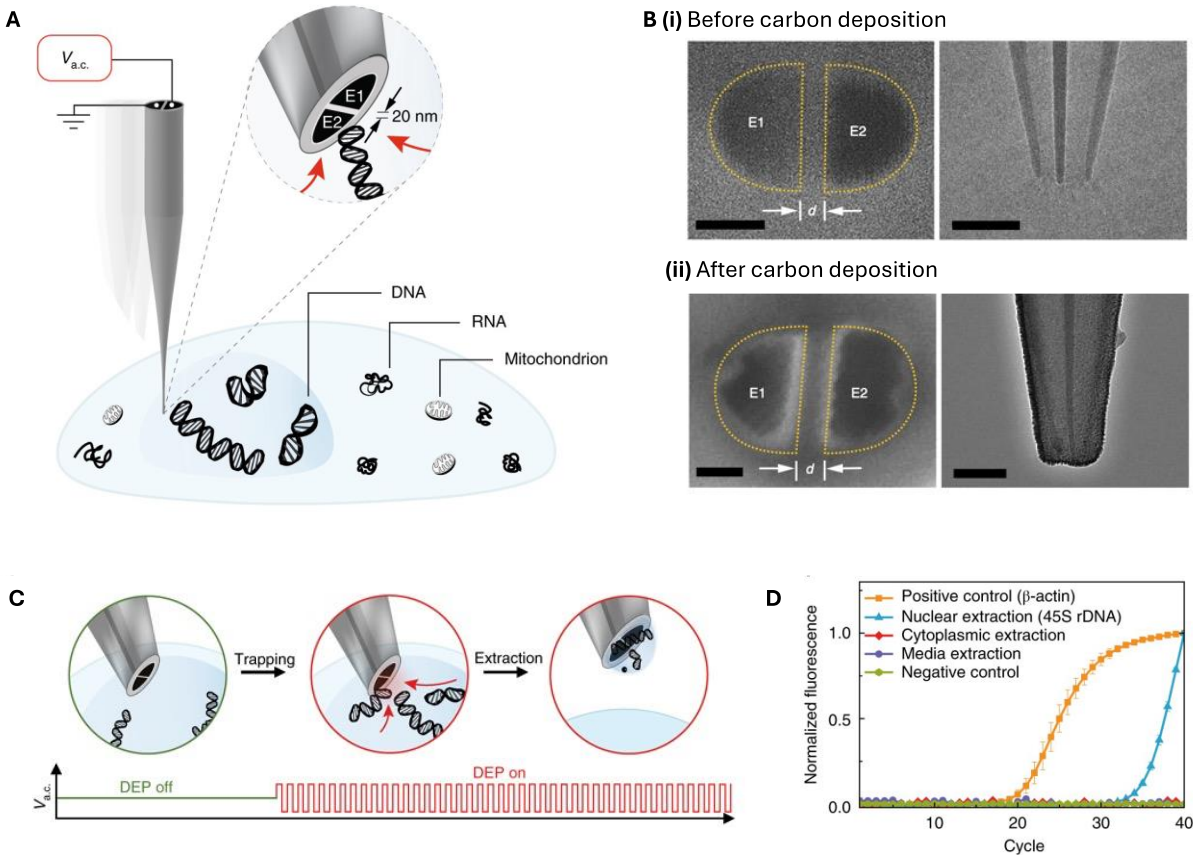


Figure 1.10 **Overview of the nanotweezer platform.** ***A*** *Application of an a.c. voltage through the nanotweezer (NT) results in the generation of a large dielectrophoretic force, sufficiently to capture polarisable material, at the tip.* ***B*** *Scanning electron microscopy (left) and transmission electron microscopy (right) images of the NTs* ***(i)*** *before and* ***(ii)*** *after carbon deposition. The carbon coating can be clearly visualised in both images as a dark haze coating the quartz surfaces. The electrode gap is shown as d. Scale bars, 20 nm (left) and 100 nm (right).* ***C*** *Trapping of molecules is a reversible process. DEP must therefore be kept on for the duration of the trapping experiment and beyond, until the captured material is ready to be dispensed into the reaction vessel.* ***D*** *DNA extracts captured from the nuclei of primary human pulmonary artery endothelial cells were qPCR amplified for a sequence contained with 45S ribosomal DNA (rDNA). qPCR amplification is shown alongside a positive control (β-ACTIN) and various controls. Figure adapted with permission from B. P. Nadappuram et al., 2019.[1] Copyright (2019) Springer Nature.*

While still requiring immense optimisation to yield the sensitivity of existing single-cell RNA and DNA sequencing techniques, these sub-cellular probing techniques show great promise towards the study of single-cell omics with unprecedented resolution, challenging current detection limits. Facilitating the repeated interrogation of the same single cell, will allow for a more comprehensive overview of dynamic processes, including epigenetic regulation, that has been postulated to underlie disease progression and development of resistance to therapeutic intervention, a key issue in hormone-dependent breast cancer.

CHAPTER 2
OPTIMISING NANOTWEEZERS FOR ENHANCED NANOBIOPSY EFFICIENCY

2.1 Graphical Summary

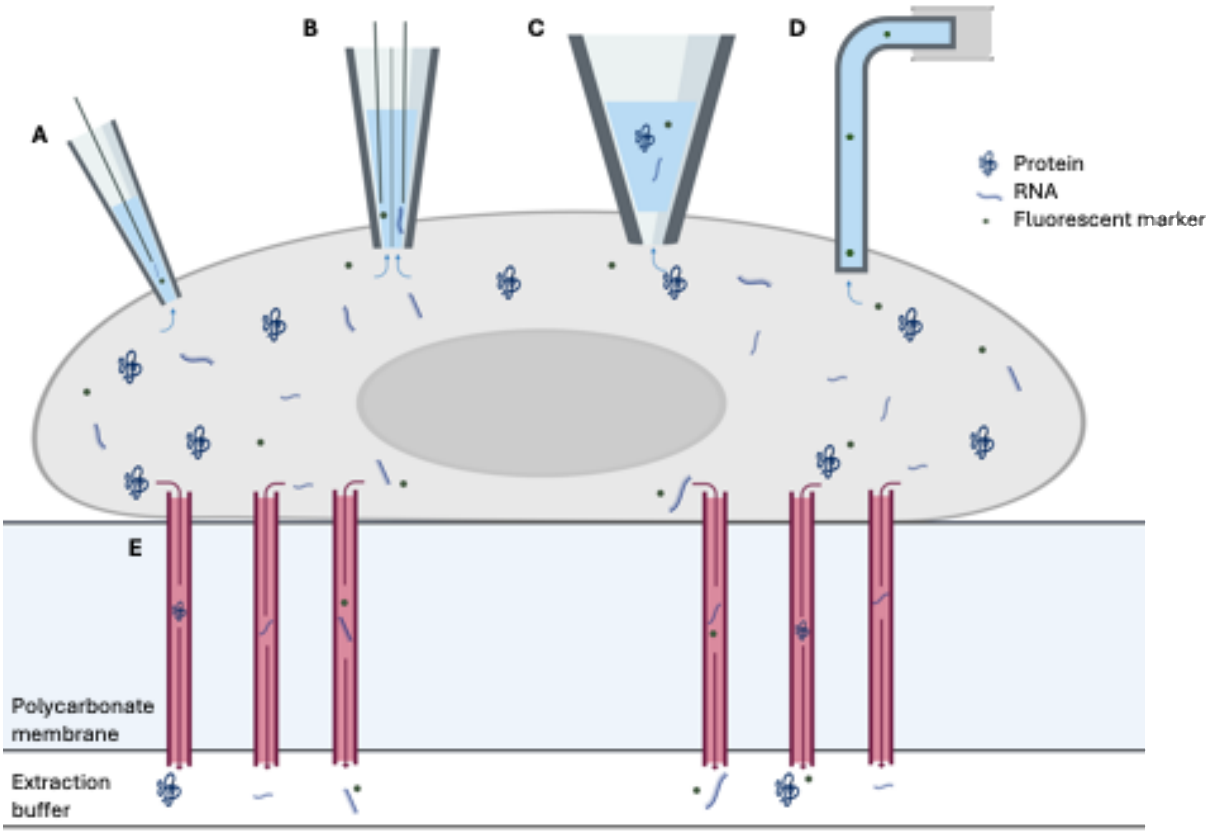


Table 2.1 Key features of the five nanobiopsy techniques shown in figure 2.1, enabling extraction of sub-cellular contents from live cells.

	A	B	C	D	E
Technique	*Nanopipettes*	*Nanotweezers*	*Fluid Force Microscopy*	*Nanotube endoscopy*	*Nanostraws*
Tip material	Glass nanopipette	Double-barrelled glass nanopipette	Hollow atomic force microscope tip	Gold nanoparticle coated carbon nanotube	Hollow alumina tube
Probe Diameter	100 nm	50 – 100 nm	400 nm	50 – 200 nm	150 nm
Volume	50 fL	N/A, pre-concentration	5 pL	attolitres	7 % extraction efficiency
Driving force	Electrophoresis, pressure	Dielectrophoresis	Pressure	Electrophoresis, pressure	Pressure, electroporation
Molecular targets	DNA, RNA, fluorescent markers	RNA, DNA, organelles, fluorescent markers, nanoparticles	RNA, proteins, fluorescent markers	Ca^{2+}, fluorescent markets, nanoparticles	RNA, proteins, fluorescent markers
Acquisition time	5 seconds	5 - 30 seconds	5 minutes	1 – 2 minutes	2 - 10 minutes

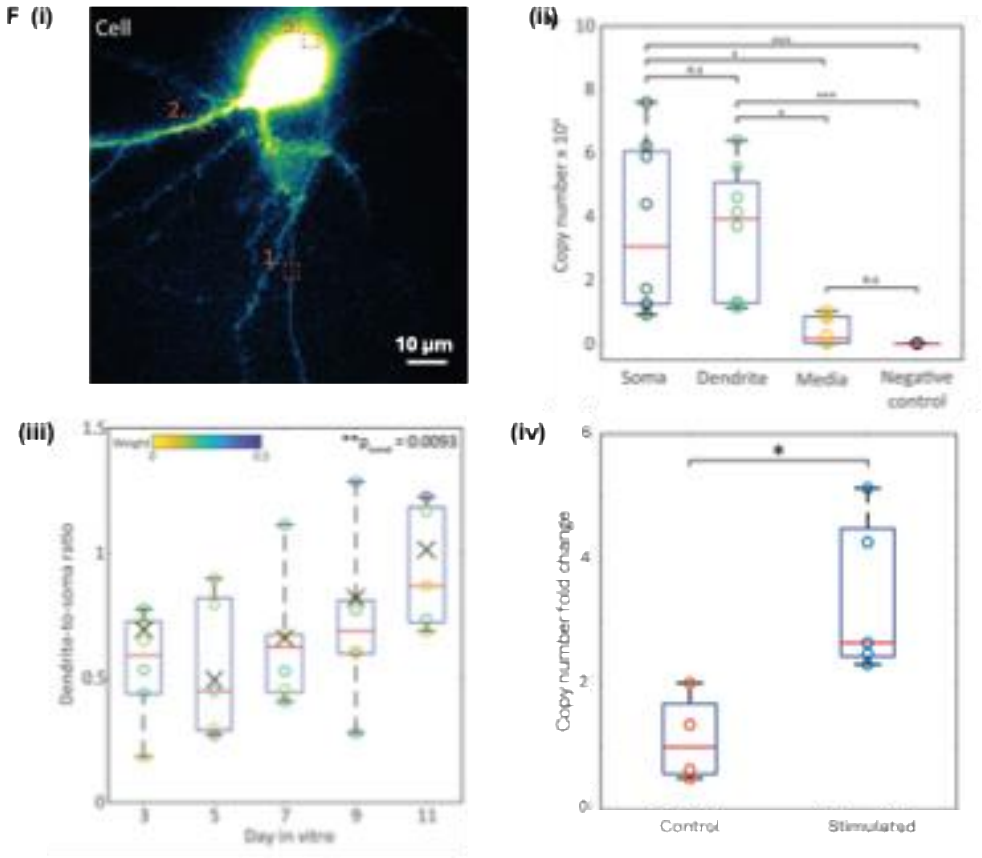


Figure 2.1 ***Graphical summary of five nanobiopsy techniques enabling extraction of sub-cellular content from live cells.*** *This figure illustrates five nanobiopsy techniques that have been developed in recent years to facilitate minimally invasive extraction of cytoplasmic or nuclear content from single, living cells:* ***A*** *Nanopipettes – fine-tipped glass capillaries that use electrophoretic or pressure-driven control to aspirate subcellular material;* ***B*** *Nanotweezers – carbon-coated double-barreled, fine-tipped glass capillaries that use dielectrophoresis to isolate and retrieve cellular content;* ***C*** *Fluid-force microscopy – a hybrid technique that combines atomic force microscopy with microfluidics to enable the mechanical puncture and extraction of defined volumes of subcellular material using defined forces;* ***D*** *Nanotube endoscopy – gold nanoparticle-coated carbon nanotubes inserted into cells that act as channels for molecular transfer and* ***E*** *Nanostraws – vertically aligned hollow alumina nanostructures, integrated into the membrane of cells to facilitate passive and electroporation-driven molecular transport from live, adherent cells.*[1,89–92] *Notably, by ensuring cells are kept alive throughout sampling, these techniques have revolutionised longitudinal sampling of individual cells within heterogenous populations. While sub-cellular sampling techniques have only recently been coupled with downstream omics applications such as single-cell transcriptomics, the underlying engineering of these methods has been evolving since the early 2000s. Each technique offers unique advantages in terms of spatial precision, volume control, throughput and compatibility with downstream molecular analysis. However, successful application of these techniques in cellular contexts is heavily dependent on the biological context and cell type, requiring careful optimization of sampling parameters. Key features of the aforementioned techniques are outlined in* ***table 2.1****. This thesis focuses on techniques* ***A*** *and* ***B****: with nanotweezers explored in chapter 2 and nanopipettes, coupled with SICM and electroporation, explored in chapter 3. Fluid force microscopy (****C****) is referenced for comparison, and techniques* ***D*** *and* ***E*** *are not referenced further in this work, though their recent advancements are acknowledged herein in the broader context of nanobiopsy development. Despite promising early results, reproducibility remains a significant challenge due to the heterogeneity of live-cell systems and the ultra-low input levels involved. Throughout this thesis, difficulties encountered during sampling and amplification are critically examined, and potential solutions are proposed to improve both the robustness and analytical utility of nanobiopsy-derived samples. Panel F illustrates advanced applications of nanotweezers in live hippocampal neurons, as demonstrated by Sahota et al. i) EGFP-filled hippocampal neurons were sequentially samples using DEP at three subcellular regions: (1) axons, (2) dendrites and (3) somas (nanobiopsy sites marked in orange). (ii) 18S rRNA was successfully amplified from biopsied material via RT-qPCR and copy numbers of 8 replicates were quantified by region against no template controls. (iii) Longitudinal tracking capabilities were demonstrated by measuring MAP1A expression six neurons over 11 days, with repeated sampling at both soma and dendrites. This was supplemented by the analysis of CAMK2A expression at dendritic spines before and after NMDA/glycine stimulation (iv). While this study highlights the versatility of this platform and the ability to target different spatial compartments and track changes with time, as discussed further in chapters 2 and 3, this application is limited by low RNA yield, sampling bias due to transcript localisation and reliance on RT-qPCR for signal detection. Top figure panel created with Biorender. Figures adapted with permission from S. G. Higgins et al., 2017 and Sahota et al., 2025.*[93,94] *Copyright (2017) Science & (2025) ACS Nano.*

2.2 Introduction

The work in this chapter builds on the NT platform developed by Nadapurram *et al.*, and expands its use to MCF-7 cells, with an aim to enable dynamic tracking of their transcriptome in response to therapy.[1] By mimicking aromatase inhibition (AI) in MCF-7 cells *in vitro*, and probing dynamic processes, a more thorough understanding of adaptive processes in cancer cells could be obtained, allowing for identification of possible druggable targets or processes. This technique was chosen due to its ability to pre-concentrate material upon application of relatively small voltage bias, which is advantageous when conducting longitudinal experiments, where cell viability is key.

2.2.1 Dielectrophoresis

When particles are placed in electric fields, their motion can be controlled by two distinct forces: electrophoresis (EP) and dielectrophoresis (DEP).[95] EP forces (or Coulomb forces) surface from the interaction of charged particles, with charge q, and the surrounding electric field, $\vec{E}$. The resulting coulombic interaction, $\vec{F}_{EP}$, can be described using the following equation:

$$\vec{F}_{EP} = q\vec{E} \tag{2.1}$$

However, where the net charge of the molecule is zero, or when the particle is an alternating field, where the current averages zero, the EP forces acting on the particle disappear. In cases where q=0, other forces, including DEP, are able to act on the particle. The concept of DEP was first mentioned in literature 1951 by Pohl and is defined as the motion of a dielectric particle, that is, a particle with inducible polarizability, in a non-homogenous electric field.[96] In its simplest form, the DEP force, $\vec{F}_{DEP}$, acting on a particle with an induced dipole moment, $\vec{p}$, in an electric field with gradient $\nabla\vec{E}$, can be defined as follows:

$$\vec{F}_{DEP} = (\vec{p}.\nabla)\vec{E} \tag{2.2}$$

The induced dipole moment, $\vec{p}$, of a dielectric can be further broken down, to consider for the induced polarizability, α, per unit volume, V, permitted the particle is isotropically, homogenously and linearly polarisable:

$$\vec{p} = \alpha \mathrm{V}\vec{E} \tag{2.3}$$

α can adopt values ranging between +1.0 and -0.5, with the sign of the value determining sign of the DEP force and thus direction of motion.[97] When α is positive, the induced dipole moment is aligned with the electric field, and a positive DEP is exerted on the particle, with the opposite being true when α is negative. The effect of the electric field gradient can be further visualised in figure 2.2. If the two poles of the particle sit within the same field, no net force is exerted on the particle, ∇E=0 and thus no motion is induced. However, if the poles sit in different fields, the net force exerted on the particle can adopt any non-zero value, and the particle will move up the electric field gradient.

Dielectrophoresis is especially advantageous given its ability to generate large electric fields upon application of relatively low voltages. A study by Bahai and Bailey deduced that $\vec{F}_{DEP}$ scales with the applied voltage, V, and the inter-electrode distance, L, as shown in equation **2.4**:[98]

$$\vec{F}_{DEP} \propto \frac{V^2}{L^3} \tag{2.4}$$

Therefore, varying electrode geometry by effectively miniaturising systems to decrease L, stronger and more concentrated electric fields can be generated, more sensitively than can be achieved by increasing V.[99] Furthermore, given that Joule heating effects, ΔT, resulting from the intense electric fields generated at the electrodes, scale with L as:[100]

$$\Delta T \sim L^2|E|^2 \quad \textbf{(2.5)}$$

decreasing L will inevitably decrease unwanted Joule heating effects. This is especially beneficial when working with cells, as Joule heating effects when the electrode is in contact with the membrane or within cytoplasm can result in generation of undesired side-reactions which may ultimately affect cell viability and observed transcriptomic measurements.

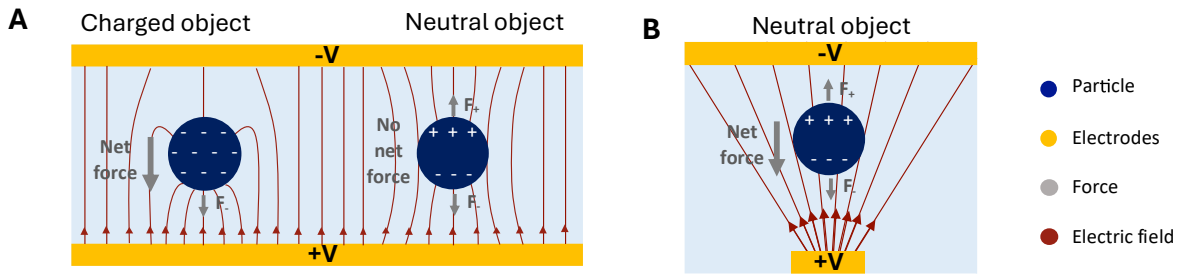


Figure 2.2 ***Electrophoretic and dielectrophoretic forces exerted on charged and neutral species in uniform and non-uniform electric fields.*** ***A*** *A charged particle in a uniform electric field experiences EP forces, while the neutral particle will not experience a net force.* ***B*** *However, when in a non-uniform electric field, the neutral particle experiences an overall net force, DEP, as the magnitude of forces acting on the induced dipoles is unequal. Charges are shown in white on the particles. Figure adapted with permission from J. Voldman, 2006.*[95] *Copyright (2006) Annual Reviews.*

2.2.2 Dielectrophoretic Trapping

DEP can be used in a wide array of systems, predominantly for spatially-controlled molecular manipulation and selective enrichment of target.[97] DEP-based systems typically comprise two individual electrodes which can be modified depending on the desired purpose, including carbon and gold coatings, which have previously been used for the trapping of double-stranded DNA (dsDNA).[102–104]. Carbon coating surfaces of quartz capillaries is extremely beneficial, especially when working with temperature-sensitive systems, as DEP field gradients of comparable magnitude can be generated with the application of 10-times less voltage bias as compared to electrodeless platforms (i.e. platforms wherein electrodes are placed outside the working fluidic channels).[105]

Given that the direction of F_{DEP} is dependent on the square of the field gradient (equation **2.4**) as opposed to the direction of the field, DEP can be induced in the presence of both direct (DC) and alternating (AC) current fields. However, especially when working with biomacromolecules, AC fields are most commonly used, as they have been shown to minimise (i) EP forces exerted on particles and (ii) any electrochemical reactions (including joule heating effects and faradaic reactions) at the electrodes, thus reducing any physiological impact on cells.[105] AC field waves can adopt various waveforms, although sinusoidal and square waveforms are generally used for DEP trapping experiments. In both cases, waves oscillate or jump between two voltages, with the difference between the minimum and maximum value termed the peak-to-peak voltage (V_{pp}), which defines the strength of the electric field gradient generated across the electrodes and thus the F_{DEP} (equation **2.4**). In a study comparing the trapping efficiency of sinusoidal and square waveforms, Contrerase-Dávila *et al.* showed, both computationally and experimentally, that square waveforms are able to generate much larger forces than with sinusoidal waveforms, even when applying the same V_{pp} across electrodes.[106] As such, square-waveforms were adopted as the waveform shape of choice in these experiments, enabling the generation of extremely large capture volumes without the need of large operating voltages.

2.3 Materials & Methods

2.3.1 Cell Culture & RNA Extraction

2.3.1.1 Routine Maintenance

MCF-7 cells (cat no. HTB-22, ATCC) were cultured in Dulbecco's Modified Eagle Medium with 1.0 g/L glucose, without L-glutamine (DMEM, Lonza) supplemented with 10% foetal calf serum (FCS, First Link), 2 mM L-glutamine, 100 units/mL penicillin and 100 units/mL streptomycin (Sigma). Given that MCF-7 cells are oestrogen-receptor positive (ER+), culture media was also supplemented with 10^{-8} M 17-β oestradiol (E2, Sigma) before incubation in a humidified, 5% CO_2 environment at 37 ˚C to replicate native growth conditions.

Cells were serially passaged every 2-3 days, or whenever culture vessels reached approximately 80% confluence. Culture media was removed from flasks and cells were briefly washed twice with 1x phosphate-buffered saline (PBS). Cells were treated with 0.5% Trypsin-EDTA (Gibco; approximately 2 mL for a 75 cm^3 (T75) flask) and incubated in a humidified, 5% CO_2 environment at 37 ˚C for 3 minutes, or until cells had completely detached from the flask. 2 mL of culture media was added to flasks to neutralise the trypsin and cells were pelleted at 1200 rotations per minute (rpm) for 4 minutes. The supernatant was discarded and cells resuspended in fresh media before being re-seeded into new flasks for maintenance.

Cells cultured and maintained in this way will be referred to as untreated MCF-7 cells for the rest of this work.

2.3.1.2 Generation of Nuclear GFP-Expressing MCF-7 cells

MCF-7 cells were labelled using Incucyte® NucLight Green Lentivirus Reagent (Sartorius) according to manufacturer's guidelines to generate a nuclear-restricted green-fluorescent protein-expressing cell line (MCF-7-nucGFP). Cells were maintained in culture as in **section 2.3.1.1**, above.

2.3.1.3 RNA Extraction

RNA was extracted from MCF-7 cells using the Trizol/Chloroform technique and purified using the RNEasy Mini Kit (Qiagen). Cells were detached from their culture vessels and pelleted by centrifugation at 1200 rpm for 4 minutes. After discarding the supernatant, the pellet was washed by briefly resuspending in PBS and pelleted by centrifugation at 3000 rpm for 5 minutes. The pellet was resuspended in 1 mL TRIzol (Invitrogen) and left to incubate at room temperature for 5 minutes, before adding 200 µL chloroform, vortexing briefly and incubating for a further 5 minutes. The sample was then centrifuged at 12 000 rpm at 4 ˚C for 20 minutes. The upper aqueous phase was carefully transferred to a clean centrifuge tube and gently mixed with 1 volume 70% ethanol before transferring to a RNEasy Mini Column (Qiagen). RNA was subsequently extracted following the protocol for the RNEasy Mini Kit (Qiagen) and eluted into 20 µL nuclease-free water containing 20 U of Superase-In RNase inhibitor (Invitrogen).

The final RNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific).[107] RNA was considered pure if the ratio of sample absorbance at 260 nm vs sample absorbance at 280 nm (A260/280) ~ 2.0. Deviations of the A260/280 from this value, specifically when lower, suggests the presence of contaminants that absorb closer to 280 nm than 260 nm. The resulting RNA was diluted into aliquots of 1 µg and stored at -80°C until ready for use.

2.3.1.4 Generation of GFP-tagged RNA

RNA extracted from MCF-7 cells was GFP-tagged using the Ulysis™ Alexa Fluor™ 488 Nucleic Acid Labelling Kit (Invitrogen) according to manufacturer's guidelines. The resulting labelled RNA was then purified using a Micro Bio-Spin® P-30 column (Bio-Rad).

2.3.2 Nanotweezer Fabrication

Two types of nanopipettes were generated for the purpose of this work. Experiments involving the different nanopipettes are separated into two chapters to avoid confusion between techniques. Work conducted with single-barrelled nanopipettes for patch-clamp experiments are discussed in **Chapter 3**, while work conducted with double-barrelled nanopipettes for NT experiments is discussed in **Chapter 2**.

Double-barrelled nanopipettes comprise two chambers separated by a quartz septum. Specifications of the double-barrelled capillaries are as follows: quartz capillary with inner filament. Inner diameter = 0.9 mm, outer diameter = 1.2 mm, length = 10 cm (WAR-QTF120-90-100, Friedrich & Dimmock).

2.3.2.1 Nanopipette fabrication

Double-barrelled capillaries were plasma cleaned (PDC-001, Harrick Plasma) for 30 minutes to remove any organic residue and contaminants. Each capillary was then aligned individually in the groove of the puller bar, parallel to the laser's angle of incidence. Two pipettes were then generated from each capillary by CO_2 laser-assisted pulling (P-2000, Sutter Instruments) using protocols shown in table 2.2. Two-line protocols were used, with the first line gently heating and pulling to generate a conical mid-section, and the second line heating further and pulling to yield two nanopipettes. An overview of the pulling protocol for double-barrelled nanopipettes is shown in figure 2.3. It is worth noting that instrument parameters are highly temperature and humidity dependent and therefore protocols may require fine tuning on other pullers or locations. Each line of the protocol comprises five parameters, which allows for fine control over pipette geometry and diameter. Unless specified, all units are arbitrary.

HEAT Refers to the output power of the CO_2 laser and thus the amount of energy that is applied to the capillary. Values range from 0-999. Typical HEAT values for borosilicate glass and quartz are 350 and 700-900 respectively.

FILAMENT (FIL) Specifies the scanning pattern that the laser adopts to apply HEAT to the capillary. Values range from 0-15, where each value corresponds to a certain longitudinal length and scan rate, thus enabling careful control of the distribution of heat. FIL values of 3 and 4 were used to generate all pipettes, corresponding to scan lengths of 4.5 and 6.5 mm, respectively.

VELOCITY (VEL) Defines the velocity at which the puller carriage is moving prior to the final PULL. As the value of velocity is determined by the viscosity of the glass, which in turn is dependent on how much the glass has been heated, the velocity can be used as a measure of glass temperature. Therefore, the glass temperature can be used as a trigger for the final hard PULL. Values range from 0-255, with lower values typically used for generating patch pipettes and higher values for micropipettes.

DELAY (DEL) Refers to the time between the deactivation of the laser and the start of the hard PULL. Values range from 0-255 ms. When DEL > 128 ms, the hard PULL is executed after laser deactivation. The larger the delay, the more the capillary is able to cool down prior to execution of the PULL, and therefore the pipette will have a smaller taper and larger tip.

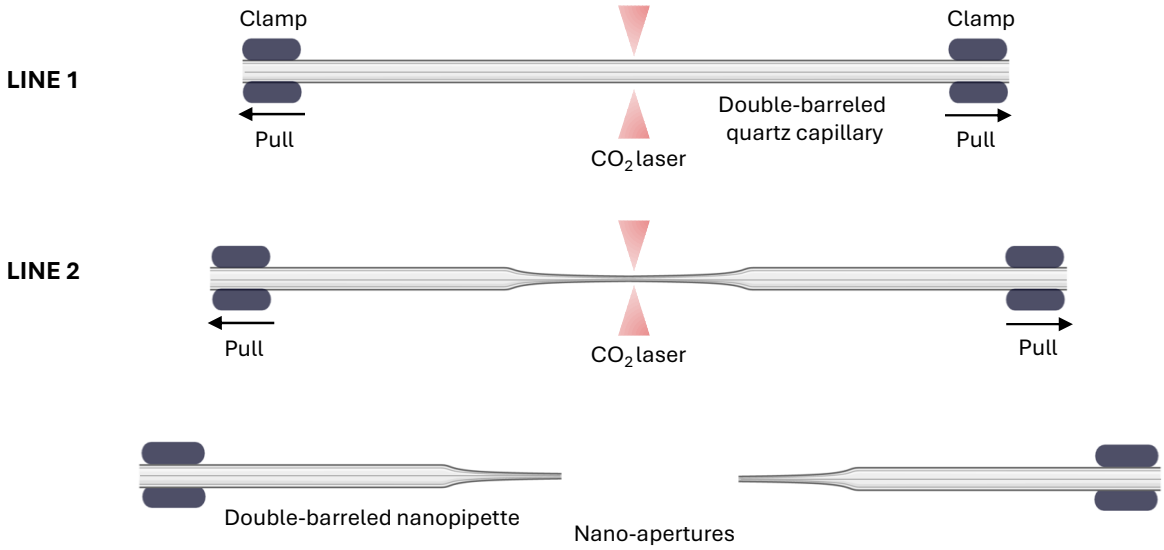


***Figure 2.3* Overview of laser-assisted pulling protocol to generate double-barrelled nanopipettes.** *Capillaries are aligned to grooves in the laser puller and clamped in place. A CO2 laser is then directed at the capillary to heat up the glass. The laser puller then sequentially progresses through the five parameters, according to pre-defined programme settings. Line 1 gently heats and pulls the capillary to yield a conical mid-section, which is further heated and pulled to generate nanopipettes in line 2.*

***Table 2.2* Two-line programmes used to generate double-barrelled nanopipettes using the P-2000 laser puller.** *Two different programmes were initially used for the pulling of double-barrelled pipettes, here referred to as (A) Programme 1 and (B) Programme 2, with only pipettes generated using Programme 2 used for nanobiopsies.*

A. Programme 1

	Heat	Fil	Vel	Del	Pul
Line 1	850	4	30	160	100
Line 2	860	3	20	140	160

B. Programme 2

	Heat	Fil	Vel	Del	Pul
Line 1	850	4	30	160	100
Line 2	875	3	20	120	165

PULL (PUL) Controls the force of the pull. Values range from 0-255, with each step corresponding to a variation of 4 mA of current through the pull solenoid.

In general, higher values of HEAT, VELOCITY and PULL yield pipettes with longer taper lengths and smaller pore diameters. Expected pulling times for the programmes used range between 4-5 seconds. Any pipettes with pulling times outside of this range were discarded. Where possible, all pipettes were pulled on the day of experimentation or functionalisation. Line 2 parameters were varied in the double-barrel programme to yield smaller and more uniform apertures. All nanopipettes used for nanobiopsies were generated using programme 2.

2.3.2.2 Carbon Deposition

Nanopipettes were processed further to generate carbon electrodes. Carbon was pyrolytically deposited onto the inner surface of each of the double-barrelled nanopipettes to generate electrodes using a bespoke setup

(figure 2.4). The nanopipette was secured onto a moveable holder and the open end attached to a butane cannister (Campingaz) *via* a Tygon tube (Saint-Gobain). The holder was slowly moved to carefully insert the tapered end into the open end of a second capillary (single-barrelled), already connected to an argon outlet. Argon and butane flows were regulated by gas flow valves. The system was flushed with argon for 15 seconds to displace any oxygen and ensure uniform carbon deposition. The rate of argon flow was decreased to ~ 0.1 litres per minute (LPM) and butane flow turned on to either 0.5 or 0.15 LPM. The tip of the nanopipette was then heated for 20 seconds using a butane blow torch (RS Components). The blow torch was run up and down the length of the nanopipette to ensure uniform carbon coverage and generate a conductive surface. A summary of all deposition conditions used for this study is shown in table 2.3. Once ready for use as NTs, copper wire was threaded into each barrel to generate electrical contact. NTs used in this way are termed 'unfunctionalised' (UF), and will be referred to as such throughout the rest of this work.

2.3.2.3 Nanotweezer Characterisation

2.3.2.3 (a) Optical Characterisation

Unfunctionalised NTs (UF NTs) were characterised optically using Scanning Electron Microscopy by Binoy Paulose Naddapuram and Annie Sahota (Edel/Ivanov Labs) to determine optimal deposition conditions: argon flow = 0.1 LPM, butane flow = 0.15 LPM, time heating tip = 20 seconds. All further pyrolytic depositions were therefore conducted using these conditions.

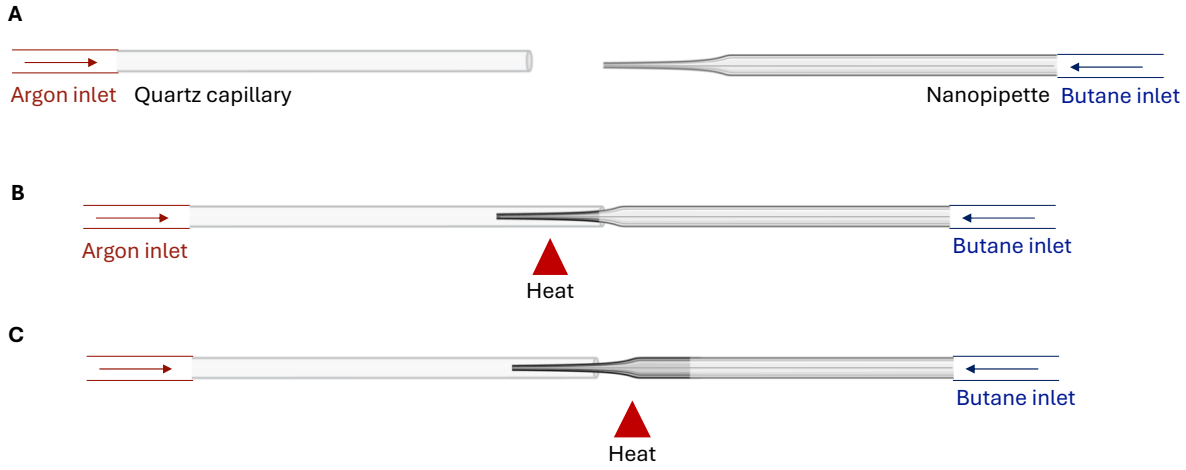


***Figure 2.4** **Bespoke carbon deposition set-up. A** A double-barrelled nanopipette is secured onto a moving holder and attached to a tygon tube, connected to a butane inlet. The holder is moved to position the tip of the nanopipette inside a single-barrelled capillary, connected to an Argon inlet. Gas flows are regulated as the **B** tips and **C** length of the nanopipette is heated to enable pyrolytic carbon deposition and yield carbon electrodes.*

Table 2.3** **Argon and butane flow parameters used to deposit carbon onto inner surfaces of double-barrelled capillaries.

Argon Flow (LPM)	0.1	
Butane Flow (LPM)	0.05	0.15
Time heating tip (s)	20	

2.3.2.3 (b) Electrochemical Characterisation

NTs were periodically electrochemically characterised, using linear sweep voltammetry (LSV), to determine variation in puller function, ensure uniformity in carbon deposition and calculate aperture size. Using ruthenium (III) hexamine chloride ($Ru(NH_3)_6Cl_3$) as a redox mediator, LSVs were run at a sweep speed of 100 mV s^{-1} between voltages ranging from -0.5 and 0.5 V, using an Ag/AgCl electrode.

2.3.2.4 Poly-thymine Functionalisation

In order to increase the efficiency of mRNA capture, the quartz surface of the NTs was conjugated with poly-thymine (polyT) linkers. PolyT should theoretically bind to the complementary poly-adenine tail of mRNA molecules, thereby enabling increased mRNA trapping, subsequent extraction and improved detection. PolyT functionalisation was conducted in two main steps: vapour silanisation (**2.3.2.4 (a)**) to yield a functionisable surface and oligomer functionalisation (**2.3.2.4 (b)**).

All incubations were conducted in 1mL screw-top autosampler vials with PTFE/Silicon septum caps to ensure NTs could be suspended and incubated securely, with each NT functionalised in its own vial. An overview of the functionalisation process is outlined in figure 2.5.

2.3.2.4 (a) Vapour Silanisation

200 µL of 3-(trimethoxysilyl)propyl methacrylate (silane) was heated at 95 °C for 20 minutes in closed vials to generate silane vapour. NTs were then suspended over the surface of the silane and the vials incubated in a dessicator for 2 hours to enable surface silanisation.

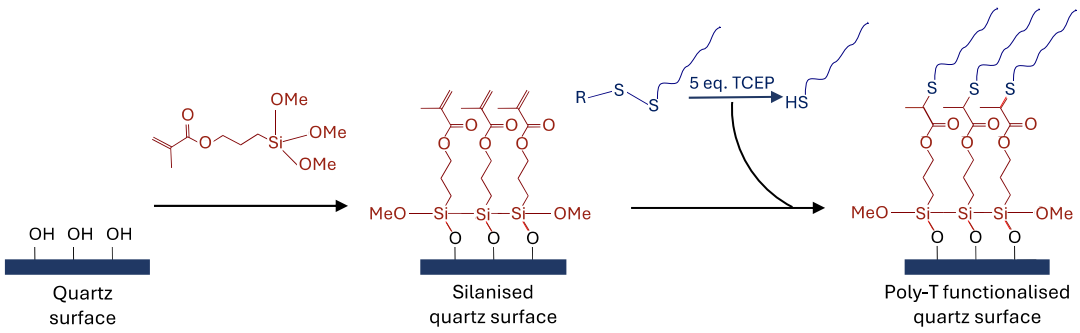


***Figure 2.5* Nanotweezer functionalisation.** *Nanotweezer (NT) surfaces were functionalised with polyT linkers, in order to increase binding of cytoplasmic RNA, which comprise a complementary poly-adenine (polyA) tail. NTs were incubated in the presence of 3-(trimethoxysilyl)propyl methacrylate vapour (red) for two hours to silanise the surface. Separately, the disulphide cap of the polyT oligomer (blue) was reduced using 5 equivalents of tris(2-carboxyethyl)phosphine. The silanised NTs were then carefully suspended in the reduced oligomer mixture for at least 12 hours prior to use.*

2.3.2.4 (b) Oligomer Functionalisation

In parallel, the bioconjugation mix required for oligomer functionalisation was prepared. The polyT oligomer was designed to have a terminal 5' protecting thiol to increase stability - 5Thio/GA CCT CCT ACT ATT

TTT TTT TTT TTT TTT T (IDT DNA Technologies). The disulfide group was therefore reduced prior to bioconjugation, by adding 500 pM polyT oligomer to 2.5 nM tris(2-carboxyethyl)phosphine (TCEP), in a total volume of 500 µL nuclease-free water and incubating at room temperature for 2 hours. A large excess of TCEP (5 equivalents) was used to ensure maximal disulphide reduction.

2.3.2.4 (c) Bioconjugation

Once silanised, the NTs were transferred and submerged into the bioconjugation mix and left to incubate for at least 12 hours. PolyT-conjugated NTs are referred to as 'functionalised' for the rest of this work. Functionalised NTs were used within one week of generation to minimise risk of surface oligomer degradation.

2.3.3 Nanobiopsies

2.3.3.1 Microscope and Nanotweezer Set-up

NTs were mounted onto a motorised micromanipulator (Patchstar, Scientifica) capable of movement in x-, y- and z-directions. This was placed above a motorised stage (Prior Scientific) on an inverted microscope (IX71, Olympus). Copper wire (0.25 mm, Goodfellows, UK) was thread through the back end of each barrel of the NT and connected to a function generator (TG2000, TTi), which was used to generate an a.c. bias between electrodes. The function generator had a frequency range of 0.001 Hz to 20 MHz, resolution of up to 1 MHz, and an amplitude range of between 5 mV_{p-p} and 20 V_{p-p} [108]. A coupling 488 nm laser (Melles Griot) was connected to a Total Internal Reflection (TIRF) module (cellTIRF, Olympus), to improve the signal-to-noise ratio (SNR), *via* an optical fibre. Light was reflected to the sample by a 488 nm dichroic mirror, through a 60x water objective (UPLSAP0 60xW, Olympus), employed in epifluorescence mode, allowing light to both illuminate the sample and for emitted light to be collected. Images were captured using an sCMOS camera (Andor Marana) cooled to -30 °C to enable extremely low noise acquisition and a 95% quantum efficiency. The whole platform was mounted on an optical table (PTM51509, Thorlabs) to minimise interference from vibrational noise. A schematic of the microscope set-up is shown in figure 2.6.

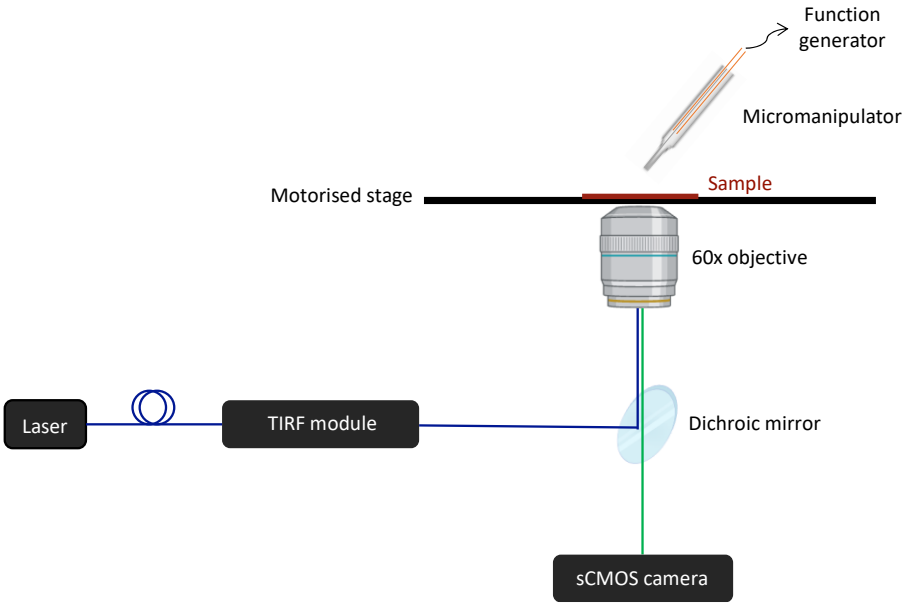


Figure 2.6 ***Microscope set-up for trapping experiments.*** *A coupling 488 nm laser and TIRF module are connected via an optical fibre. Light is reflected through a 488 nm dichroic mirror onto the back of a 60x W objective, and subsequently onto the sample, which is placed on a motorised stage. Light emitted from the sample is collected in the opposite way, through the 60x objective, transmitted through the dichroic mirror and finally detected by a sCMOS camera. A nanotweezer (NT) is mounted onto a motorized micromanipulator above the sample. Copper wires are thread into each of the barrels of the NT and connected to a function generator, which enables generation of an a.c. bias for trapping. Figure adapted with permission from B. P. Nadappuram et al., 2019.[1] Copyright (2019) Springer Nature.*

2.3.3.2 Dielectrophoretic Trapping

Extracting material from solution and cells involve largely similar processes with regard to application of the dielectrophoretic force and extraction of trapped material but varies in terms of positioning the NT. Sections **2.3.3.2 (a)** and **2.3.3.2 (b)** outline variations in positioning the NT in solution and cells, respectively. All RNA trapping experiments discussed in this thesis were conducted by Anthony Monteza-Cabrejos (Edel/Ivanov Group), with both UF and F NTs.

2.3.3.2 (a) In Solution

A droplet containing 0.6 pM GFP-tagged RNA in 100 mM KCl solution was placed onto a coverslip on the microscope stage. The NT was carefully mounted into the micromanipulator, perpendicular to the coverslip and slowly lowered into the droplet. The function generator was turned on to generate a DEP force with a square waveform and frequency of 1 MHz. $V_{p\text{-}p}$ was varied between 5 and 20, to validate optimal trapping conditions. The NT was held in solution for 60 seconds before withdrawal. Extracts were collected by inserting the NT tips into PCR tubes containing 5 µL of nuclease-free water and 5 U of Superase-In RNase inhibitor (Invitrogen) and crushing the tip. The function generator was only turned off once the sample had been collected.

2.3.3.2 (b) In Cells

MCF-7-nucGFP were seeded into 35 mm, high Grid-500 culture µ-dishes (Ibidi), at a density of 5 x 10^3 cells/mL one day prior to conducting nanobiopsies. The dish was placed onto the microscope stage and the NT positioned at a slight angle in the micromanipulator to facilitate cell penetration. Cells were first visualised in brightfield mode to allow for accurate NT positioning, before switching to epifluorescence mode to allow visualisation of the GFP-tagged nuclei. The NT was carefully moved to penetrate the cell membrane, being careful not to pierce the nucleus. Material was extracted as in section **2.3.3.2 (a)**, but with waveform magnitude of 15 $V_{p\text{-}p}$ and NTs held in solution for 15 seconds.

2.3.3.2 (c) Variations: Pooling

In order to increase the amount of material extracted during nanobiopsies, the same cell was probed multiple times, (i) two, (ii) three and (iii) five times. Each extraction was conducted with a new NT, and the samples pooled together into the same PCR tube as in **2.3.3.2 (b)**. PCR tubes were kept on ice until all biopsies had been pooled, before snap-freezing.

2.3.4 Real-time Quantitative Polymerase Chain Reaction

A ProFlex™ PCR System (Applied Biosystems) was used for all reverse transcription steps and all qPCR reactions were conducted on a QuantStudio 3 Real-Time PCR System (Applied Biosystems).

2.3.4.1 Sample Processing

Initially, 10 functionalised and 10 unfunctionalized nanobiopsy extracts were subjected to RT-qPCR using iScript cDNA synthesis kit (Bio-Rad) and SYBR Select qPCR Master Mix (Applied Biosystems). Reagent combinations were optimised as described in section **2.3.4.2**, after which all further nanobiopsy sample processing was conducted using LunaScript® RT Mix and Luna® qPCR Master Mix (protocol outlined in section **2.3.4.2**).

2.3.4.1 (a) cDNA synthesis – iScript Reverse Transcriptase

To each 5 µL nanobiopsy sample, 1 µL iScript reverse transcriptase (RT), 5 µL reaction buffer and 10 µL nuclease-free water were added. cDNA was then synthesised according to the following protocol: *priming* - 25 °C for 5 minutes; *reverse transcription:* 46 °C for 20 minutes; *enzyme inactivation:* 95 °C for 1 minute. Due to potential sample loss during pipetting steps and as all 20 µL of sample was required for downstream qPCR reactions, 2 µL nuclease-free water was added to each cDNA sample. Samples were stored at -20 °C for a maximum of 3 days prior to use in qPCR.

2.3.4.1 (b) qPCR – SYBR Select qPCR Master Mix

Three oestrogen-responsive targets (*ESR1*, *TFF1* and *GREB1*) were explored for each sample of interest using *GAPDH* as a housekeeping gene. Reactions were set up in MicroAmp Fast Optical 96-well 0.1 mL plates (Applied Biosystems). Each 20 µL cDNA sample was split across four wells, with 5 µL cDNA in each. Each reaction required the addition of a master mix containing 10 µL SYBR Select qPCR Master Mix and 4 µL nuclease-free water, alongside a mix gene-specific forward and reverse primer at a final concentration of 2.5 µM. Sequences of primers used for qPCR experiments are shown in table 2.4. Plates were carefully vortexed and spun in a CVP-2 microplate centrifuge vortex (Grant-Bio) and qPCR was conducted using the following programme: *Enzyme Activation*: 50 °C for 2 minutes, heated to 95 °C at a rate of 1.6 °C/s, hold at 95 °C for 10 minutes; *cDNA Amplification*: 40 cycles of denaturation at 90 °C for 15 seconds and annealing at 60 °C for 1 minute, with fluorescence acquisition at the end of every cycle; *Melt Curve*: Increasing temperature from 60 °C to 90 °C at a rate of 0.1 °C/s.

***Table 2.4** Forward and reverse primer sequences used for the three genes of interest (ESR1, TFF1, GREB1) and the housekeeping gene of choice, GAPDH, for all qPCR experiments in this chapter.*

Gene	Forward Primer Sequence	Reverse Primer Sequence
ESR1	5'- CAGGTGCCCTACTACCTGGA-3'	5'- ACTGGCCAATCTTTCTCTGC-3'
β-ACTIN	5'- CACCATGGAGAACAAGGTGA-3'	5'- TGACACCAGGAAAACCACAA-3'
GREB1	5'- ATCAGCTGCTCGGACTTGCTG-3'	5'- TGAGCTCCGGTCCTGACAGATG-3'
GAPDH	5'-GGTCACCAGGGCTGCTTTTA-3'	5'-TTCCCGTTCTCAGCCTTGAC-3'

2.3.4.2 RT-qPCR Reagent Optimisation

Given the poor qPCR detection of the nanobiopsy extracts using the reagent combination in section 2.2.4.1 (as discussed in section 2.3.3), the efficiency and detection limits of two different reagent kits were compared to determine whether detection could be improved using alternate reagents. This was done using RNA extracted from MCF-7 cells using the TriZOL/Chloroform method as outlined in section 2.2.1.3 and testing for *GAPDH*, using the primer sequences in table 2.5. RTs tested were: iScript RT Supermix (Bio-Rad) and LunaScript® RT Mix (NEB); and qPCR master mixes tested were: SYBR Select Master Mix (Applied Biosystems) and Luna® Universal qPCR Master Mix (NEB).

Experiments were conducted by serially diluting 1 µg of RNA, such that starting samples contained between 1 µg and 0.1 pg RNA, using a 1:10 dilution factor. All dilutions were reverse transcribed, amplified and *GAPDH* expression was measured using the following reagent combinations: (a) iScript RT Supermix and SYBR Select Master Mix; (b) iScript RT Supermix and LunaScript® qPCR Master Mix; (c) LunaScript® RT Mix and SYBR Select Master Mix; and (d) LunaScript® RT Mix and LunaScript® qPCR Master Mix. An overview of the reagent combinations is shown in table 2.5, section **2.4.4**. Samples were processed with iScript RT Supermix and SYBR Select Master Mix according to the protocols outlined in sections **2.3.4.1 (a)** and **2.3.4.1 (b)**, respectively. LunaScript® RT and Luna® Universal Master Mix protocols are outlined below.

The protocol was initially optimised using *GAPDH* primers only, before validation with *ESR1*, *GREB1* and *TFF1* primers (see table 2.4 for primer sequences). Standard curves were generated for all reagent combinations, from which amplification efficiency (E) and correlation coefficients (R^2) were calculated for all primer combinations.

2.3.4.2 (a) cDNA synthesis – LunaScript® RT Mix

4 µL LunaScript® RT and buffer mixture was added to 5 µL sample and made up to 20 µL with nuclease-free water. cDNA was synthesised as follows: *priming*: 25 °C for 2 minutes; *reverse transcription:* 55 °C for 10 minutes; *enzyme inactivation:* 95 °C for 1 minute. After cDNA synthesis, 80 µL nuclease-free water was added to the reaction mix. All qPCR reactions were conducted using the diluted samples.

2.3.4.2 (b) qPCR – Luna® qPCR Master Mix

0.25 µM forward and reverse primer mix and 10 µL Luna® Universal qPCR master mix was added to each 5 µL sample to be tested. The reaction was made up to a final volume of 20 µL with nuclease-free water. Plates were carefully in a CVP-2 microplate centrifuge vortex (Grant-Bio) and qPCR was conducted according to the following programme: *initial denaturation:* 95 °C for 60 seconds; *cDNA Amplification:* 40 cycles of denaturation at 95 °C for 15 seconds and template extension at 60 °C for 30 seconds, with fluorescence measured at the end of each cycle; *melt curve:* Increasing temperature from 60 °C – 95 °C at a rate of 0.1 °C/s.

2.3.5 RNAScope

Given the difficulties in detecting the four genes of interest from the nanobiopsy samples, Multiplex RNAScope was conducted to determine **(i)** whether gene localisation affected the ease of trapping during nanobiopsies and **(ii)** whether gene expression varied as significantly as RT-qPCR values suggested.

The RNAScope assay was conducted using the RNAscope® Multiplex Fluorescent Reagent Kit v2 (ACD-Bio). Using proprietary double-Z probes, this kit enables the spatial visualisation of up to four distinct RNA targets, depending on microscope capabilities, in fixed cells as punctate dots (figure 2.7). [109] In this study, the three-target kit was used. Gene-specific probes hybridise to the corresponding RNA targets in cells. Each Z-probe comprises three sequences: an 18-25-base sequence complementary to the target RNA, a spacer sequence and a 14-base tail sequence. Two of these Z- probes hybridise directly adjacent to each other on the target RNA sequence, yielding the double-Z configuration, which provides a 28-base site for the pre-amplifier to hybridise. Up to 20 double-Zs bind to the target RNA, covering a 1 kb region. The pre-amplifier provides a 20-base platform for amplifier binding, with each amplifier sequencing providing a further 20 binding sites for each of the label probes. Combined, this double-Z-probe design and sequential hybridisation steps enable highly sensitive and specific read-outs by minimising off-target probe hybridisation.

The assay was expanded from untreated MCF-7 cells, to include timepoints that would enable the study of variations in gene expression in response to treatment during entrance to dormancy and in response to transient changes in treatment conditions. Culture conditions were adjusted to mimic AI. A summary of all the timepoints explored is summarised in figure 2.8.

2.3.5.1 Experimental Set-up

1.2 x 10^5 MCF-7 cells were initially seeded in T75 flasks for day 7, 14, 21 and 28 AI timepoints in standard maintenance conditions – DMEM containing 10% FCS, 2 mM L-glutamine, 100 units/mL penicillin and 100 units/mL streptomycin. In parallel, 50 x 10^3 MCF-7 cells were seeded into each well

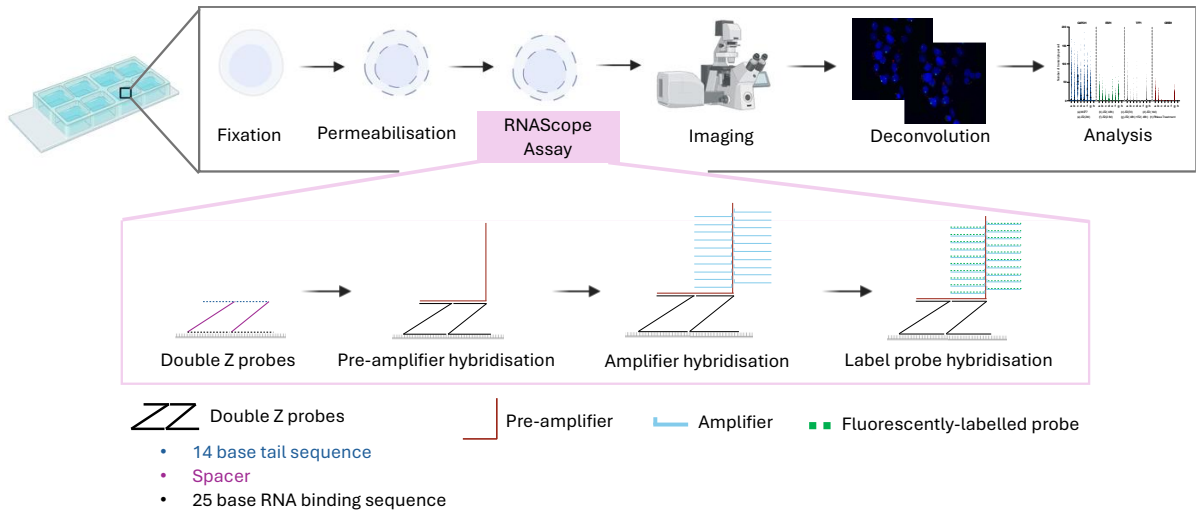


***Figure 2.7* RNAScope Overview.** *Cells are initially removed from culture, fixed and permeabilised prior to conducting the RNAScope Assay. Two Z probes bind next to each other on the RNA target, to form a double-z structure. A pre-amplifier sequence then hybridises to the top of the double-Z structure, providing binding sites for the amplifier sequence. Labelled probes (Opal 520, 570 and 650) are then able to bind to the amplifier, enabling visual readouts of genes of interest. Images are then deconvolved to increase image resolution prior to analysis. Figure adapted from F. Wang et al., 2012 and ACD-Bio.*[109,110]

of six two-well Nunc™ Lab-Tek™ II Chamber Slides (Thermo Fisher Scientific). Cells were treated with 10^{-8} M E2 and incubated in a humidified, 5% CO_2 environment at 37 °C for 48 hours. Two of the chamber slides were then removed from culture (D0), and the remaining chamber slides and flasks were washed twice with 0.5 mL and 5 mL 1x PBS, respectively, and returned to culture in phenol red free DMEM supplemented with 10% charcoal stripped FCS (First Link), 2 mM L-glutamine, 100 units/mL penicillin and 100 units/mL streptomycin. Cells were also deprived of oestrogen to mimic AI. After a further 48 hours, two more chamber slides were removed from culture (*48hAI*) and the remaining slides were supplemented with 10^{-8} M E2 for a further 48 hours (*48hAI+48hE2*) prior to fixation. Flasks were maintained in culture with weekly media changes. 48 hours prior to removal from culture, cells were detached from flasks and 50 x 10^3 MCF-7 cells were seeded into each well of two two-well Nunc™ Lab-Tek™ II Chamber Slides. Cells were left to adhere in the chamber slides for 48 hours prior to fixation.

In parallel, a second T75 was cultured in the same way for all AI timepoints. Cells were collected at the respective timepoints and RNA was extracted according to the Chloroform/TriZOL method as outlined in section 2.2.1.3.

2.3.5.2 Cell Fixation

When cells were ready for collection, growth media was removed, and the chambers disassembled. Slides were briefly submerged in 1x PBS and transferred to a 4% (wt. %) paraformaldehyde (PFA) solution in PBS (Thermo Scientific) and incubated at room temperature for 15 minutes. All slide submersions were conducted in Coplin staining jars (Ref: MIC6000, SLS). The slides were then gently rinsed twice with 1 x PBS. Slides were subjected to a series of 5-minute incubation steps at room temperature in 50 % (v/v) ethanol, 70 % (v/v) ethanol and 100 % ethanol, before a final 10-minute incubation in 100 % ethanol to dehydrate cells prior to storage at -20 °C until ready for processing.

2.3.5.3 Cell Rehydration

Cells were rehydrated with a series of 2-minute incubations in 70 % (v/v) ethanol, 50 % (v/v) ethanol, and a 10-minute incubation in 1x PBS, all at room temperature. Cell rehydration and all subsequent experimental steps were carried out for all treatment timepoints at the same time.

2.3.5.4 Cell Pre-treatment: Fixation and Permeabilisation

Prior to pre-treatment and hybridisation, one well from the of the D0 timepoint was treated with 0.05 mg/mL RNase A diluted in 50 mM Tris-HCl (pH 8.0) for an hour at 37 °C to determine whether

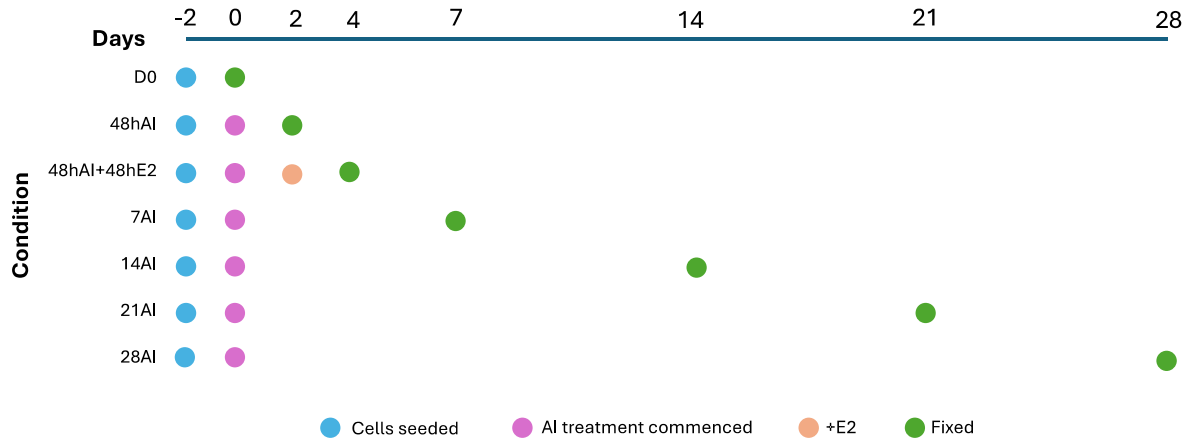


Figure 2.8 ***Overview of cell treatment conditions prior to RNAscope Fluorescent Multiplex v2 Assay.*** *MCF7 cells were cultured in seven different ways – (a) standard oestradiol supplemented (+E2) conditions for 48 hours (D0); (b) oestradiol deprived (-E2) conditions for 48 hours (48hAI); (c) 48h AI followed by 48h +E2 (48hAI+48hE2); (d) 7 days -E2 (7AI); (e) 14 days -E2 (14AI); (f) 21 days -E2 (21AI) and (g) 28 days -E2 (28AI). All conditions were seeded into two-well chamber slides 48 hours prior to fixation.*

spots detected on the individual images were true signals. This well was hybridised with a *GAPDH* probe only.

Cells were treated with a few drops of RNAScope Hydrogen Peroxide (ACD-Bio) to cover the entire slide and left to incubate for 10 minutes at room temperature. Slides were tapped to remove excess reagent and rinsed twice by submerging in a Coplin jar containing distilled water, with frequent agitation. All further rinsing steps, unless otherwise specified, were conducted in the same manner. Using an Immedge hydrophobic barrier pen (H-4000, Vector Labs), a hydrophobic barrier was drawn around each well and allowed to dry before rinsing in 1x PBS. A few drops of Protease III (ACD-Bio), enough to cover each well, diluted 1:11 in PBS, was applied to each slide and left to incubate for 10 minutes at room temperature. Excess liquid was removed, and slides were washed twice in 1x PBS.

2.3.5.5 Multiplex RNAScope Assay

For all conditions, each slide was hybridised in two ways. Probes were assigned to each of the three channels: *GAPDH*-C1, *ESR1*-C1, *TFF1*-C2 and *GREB1*-C2. The following probe mixes were prepared, by adding 1 volume of the C2 probe to 50 volumes of the C1 probe. One well from each chamber slide was treated as follows: *Well 1 (probe mix a): GAPDH*-C1, *TFF1*-C2; *well 2 (probe mix b): ESR1*-C1, *GREB1*-C2. Unless otherwise specified, all steps outlined below were followed by two brief rinses in 1x Wash Buffer (ACD-Bio), with gentle agitation, and incubated in a HybEZ™ II Oven (ACD-Bio) at 40 °C. Initially, a few drops (approximately 75 µL) of probe mix were pipetted onto each of the respective wells and incubated for 2 hours. Targets were then amplified in three rounds of hybridisation steps. Each round comprised addition of a few drops of RNAScope® Multiplex FL v2 Amp reagent (AMP 1, AMP 2 or AMP 3; ACD-Bio), incubation in a HybEZ™ II Oven at 40 °C and two washes. AMP 1 and AMP 2 were incubated for 30 minutes, and AMP 3 for 15 minutes.

Opal fluorophores (Opal 520, 570 and 650, Akoya Biosciences) were reconstituted in 75 µL DMSO before diluting 1:750 in TSA buffer (ACD-Bio). Each fluorophore was then uniquely assigned to a channel, and thus an individual target, with channels being processed in a step-wise manner. A few drops of HRP-C1 were added to the slides and incubated for 15 minutes. Approximately 150 µL of the assigned Opal fluorophore was added to the slides, and again incubated for 30 minutes, before addition of a few drops of HRP-blocker and another 30-minute incubation. This was repeated for HRP-C2 and HRP-C3 with their respective fluorophores. Finally, nuclei were stained with a few drops of DAPI (ACD-Bio). A few drops of Prolong Gold Antifade Mountant (Thermo Fisher Scientific) were added to each slide and a coverslip was placed on top, being careful not to trap any air bubbles. Slides were stored at in the dark at 4 °C, before imaging the following day.

2.3.5.6 Widefield Imaging

All slides were imaged using a Zeiss Axio Observer Inverted Widefield Microscope in the Facility for Imaging by Light Microscopy at Imperial College London. The widefield microscope employs a CoolLED pE-4000 illumination system. Using the Zeiss ZenPro software, three LED channels were set-up at specified wavelengths: 365 nm (DAPI), 470 nm (FITC) and 555 nm (Cy3). Cells were located and focus adjusted using the 5x objective with the DAPI channel selected. A 4x4 tile was created and a preview scan was conducted, again using the DAPI channel. Once a satisfactory preview scan was obtained, the objective was changed to 40x oil. Preview quality was verified by 'live' imaging the slide using all channels of interest, selecting regions on the preview, assessing overlap with the preview and adjusting focus as relevant. Z-stacks, comprising 3 planes, with 0.6 µm step size, to cover a total depth of 1.2 µm, were defined, again using the most focused plane as the central reference. Positions were selected and images acquired with all three channels activated. Exposure times and laser intensity were adjusted according to visualised signal intensity.

2.3.5.7 Image Deconvolution

Widefield files were deconvolved using Huygens (Scientific Volume Imaging, licence provided by the Facility for Light Microscopy at Imperial College) to improve image resolution. Wavelengths of imaged channels and refractive index of mountant used (Prolong Gold; 1.47) were specified prior to running the 'Deconvolution Wizard', using standard parameters. Images were corrected for lamp jitter, where necessary, and saved as 16-bit TIFFs, with one slice assigned per channel on a linked scale.

2.3.5.8 Analysis: FIJI and QuPath

TIFF files were opened in FIJI (Fiji is Just ImageJ) for initial image processing. Each image file, acquired as a z-stack, was converted into 2-D image by generating a maximum Z-projection, such that all imaged pixels were projected onto the same plane. Maximum Z-projection files combined into a single file, with all channels superimposed and composite images were saved as TIFFs. This process was converted into a macro for ease of processing (code shown in supplementary **2.2**).

Composite maximum projection TIFFs were then opened in QuPath for further processing, involving nuclei and cell membrane detection and transcript quantification per cell, for each treatment condition (figure 2.9). Nuclei and membrane detection were conducted using the in-built 'cell detection' module on the DAPI channel (channel 1). No adjustments were made to the recommended thresholds - requested pixel size: 0.5 µm, cell expansion: 5 µm, with intensity split by shape and cell expansion including cell nucleus. Transcripts were quantified using the 'subcellular spot detection' module. Thresholds for channels containing transcripts and minimum spot size were determined by measuring pixel intensity of transcripts on the image and spot area, respectively. Transcript counts were plotted and a Kruskal-Wallis test was conducted to determine significant differences between different treatment conditions in Graphpad Prism v10.

2.4 Results & Discussion

2.4.1 Nanotweezer Fabrication: Optimisation & Characterisation

Contributions:

- *Generation of NT pool for optical characterisation was conducted by Zoe Kwan, Annie Sahota, Anthony Monteza Cabrejos (all Edel/Ivanov Group) and Debjani Saha.*
- *SEM imaging was carried out by Annie Sahota and Binoy Nadappuram (both Edel/Ivanov Group).*

Nanopipettes were generated by the laser-assisted pulling of double-barrelled quartz capillaries. SEM imaging of these nanopipettes reveal semi-elliptical apertures, separated by a 20 nm insulating septum, at the end of a tapered tip (figure 2.10 (a)). On average, aperture diameters of nanopipettes used in experiments ranged from approximately 30-50 nm, resulting in tips with total diameters of approximately 100 nm (figure 2.10 (b), (c)(ii)). Aperture sizes were carefully regulated by ensuring that quartz capillaries were aligned in the same way in the laser puller and by using the same pulling protocol. However, previous nanobiopsy experiments conducted by other members of the Edel group,

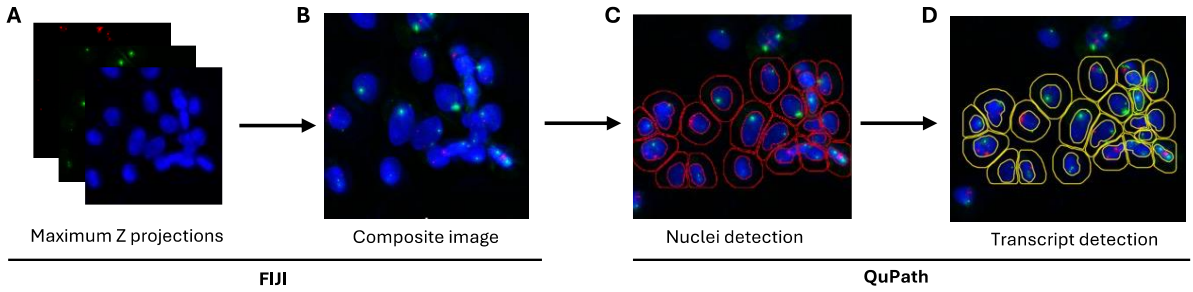


***Figure 2.9* Image processing pipeline.** *Once images had been deconvolved, **A** maximum Z-projections were generated in FIJI before **B** creating a composite image, wherein all channels could be visualized in a single image, in the same plane. Composites were saved as TIFFs and opened in QuPath for further processing. **C** Nuclei were detected and membrane boundaries estimated and annotated (red). **D** Transcripts (red and green) were then detected within the annotations (yellow) using 'subcellular detection' and measured values for intensity and size.*

revealed difficulties in extracting detectable quantities of sub-cellular material for downstream processing and analysis (data not shown). As such, various techniques were adopted to improved extraction yield. Initial modifications involved varying parameters in line 2 of the pulling protocol, specifically: increasing heat, decreasing delay and increasing pull (table 2.2). By decreasing the delay time, the quartz has less time to cool down in between steps (also affected by increased heat value), which in turn results in an increased taper length and smaller tip. Aperture size is further decreased by increasing the value of the hard pull. The effect of this is twofold. Firstly, by generating longer tips with smaller apertures, the inter-electrode distance can be decreased. Recalling **equation 2.4**, decreasing the inter-electrode distance increases the DEP force generated between the two electrodes when conducting nanobiopsies with minimal residual joule heating effects, thus enabling increased RNA trapping. Secondly, use of smaller NTs minimises stress exerted on cells during cell penetration and extraction, enabling cell membranes to recover more efficiently. Pore sizes of the resulting nanopipettes from the two pulling protocols, post carbon deposition, were measured using linear sweep voltammetry (figure 2.10 (c)).

Carbon was pyrolytically deposited on the inner surfaces of nanopipette barrels from butane, with a constant stream of argon being flushed through the system to displace any residual oxygen, using a bespoke carbon deposition set-up (figure 2.4). The rates of argon and butane flow were carefully regulated to allow further control over aperture size. Optimisation of the carbon deposition process entailed variation in the rate of flow of butane, testing two different rates - 0.5 and 0.15 LPM. Optical characterisation, using SEM, revealed most uniform and reproducible carbon deposition using a butane flow rate of 0.15 LPM. Figure 2.10 (b) shows comparison of NTs generated with both butane flow rates and a broken NT tip. Aperture sizes were also periodically validated using LSV to ensure proper puller function and successful carbon deposition, by measuring the reduction of RuHx between -0.5 and 0.5 V. Given that NTs exhibit ohmic behaviour in certain voltage ranges, where the ionic current, I, is proportional to the voltage bias, V, applied, the aperture radius, r, can be calculated from the linear region of the LSV trace, with the assumption that apertures were of circular geometry. However, as seen from figure 2.10 (b), apertures adopt semi-elliptical geometries, and thus LSV aperture sizes can only be considered rough approximations. The aperture radius, r, was calculated using the following equation:[111,112]

$$r = \frac{i_{ss}}{4nFDC} \qquad \textbf{(2.6)}$$

where:

i_{ss}: steady-state reduction current (A)

n_e: the number of electrons transferred during reduction = 10

F: Faraday constant = 96485.3365 C mol^{-1}

D: diffusion constant of $Ru(NH_3)_6^{3+}$ in 100 mM KCl = 8.5 x 10^{-10} cm^2 s^{-1} [112]

c: bulk concentration of redox active species, $Ru(NH_3)_6^{3+}$ = 0.1 mM

Calculated aperture sizes ranged from 0.07 to 345 nm. NTs with aperture sizes greater than 50 nm and those with unequal pore sizes were generally indicative of broken tips and uneven carbon

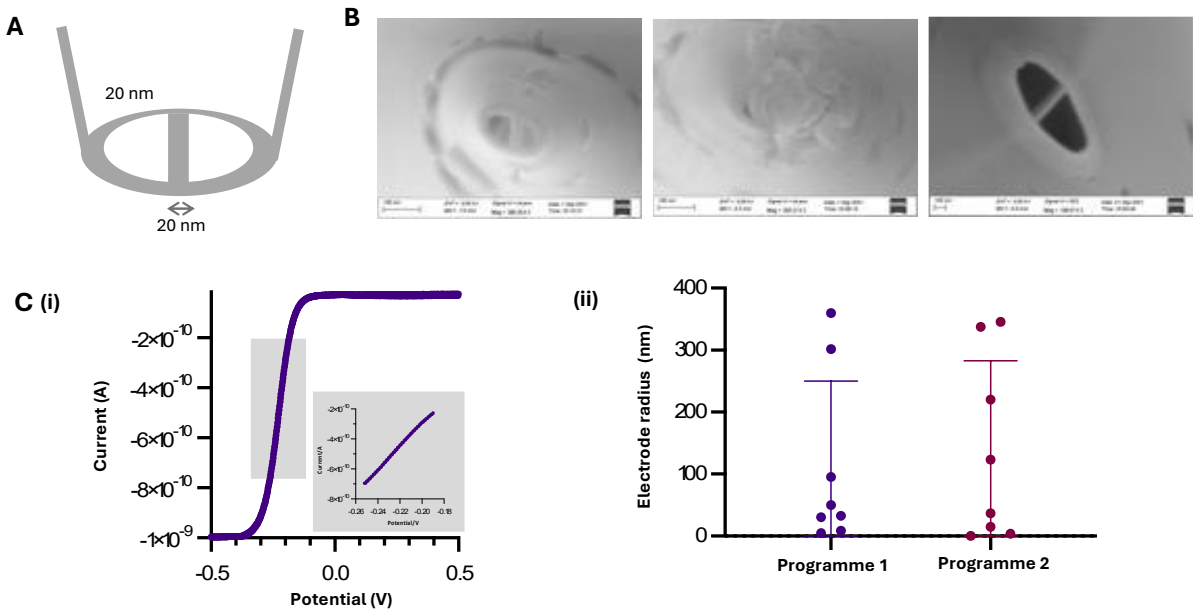


Figure 2.10 ***Nanotweezer fabrication: Optimisation and Characterisation*** ***A*** *Schematic overview of NT (NT) dimensions.* ***B*** *SEM images of an optimally carbon coated NT (butane flow = 0.15 LPM), an overdeposited NT (butane flow = 0.5 LPM) and a broken, carbon-coated NT tip (butane flow = 0.15 LPM) (L-R). Scale bars represent 100 nm.* ***C (i)*** *NTs were electrochemically characterized by linear sweep voltammetry with 0.1 mM $Ru(NH_3)_6Cl_3$ in 100 mM KCl between -0.5 and 0.5V. Inset (grey): Aperture sizes were calculated from the linear region of the voltammograms.* ***(ii)*** *A large variation was observed in pore sizes, both between different pulling protocols and within protocols. NTs with non-symmetric apertures, or aperture sizes above 50 nm were discarded.*

deposition, respectively, and were discarded. Small electrode areas and inter-electrode gaps are essential parameters to consider when working with dielectrophoretic forces.[95] While NTs with large apertures or unequal pores could theoretically still be used for nanobiopsy experiments, larger aperture sizes would result in increased heat generation and undesired electrochemical reactions, subsequently exerting significantly more stress during cell penetration and withdrawal. As the ultimate target of this platform is to enable the sequential probing of single cells over time, it is of paramount importance that the cells are **(i)** kept alive, and are **(ii)** not subjected to unnecessary stress, in this case, both in the form of heat and membrane stress. Therefore, variation of tip size, by decreasing delay and increasing pull, is an attractive technique by which to increase trapping force without increasing cytotoxicity.

2.4.2. Validating nanotweezer functionalisation

Contributions:

- *All functionalisation validation experiments were conducted by Anthony Monteza-Cabrejos, Edel/Ivanov Lab, Imperial College London.*

In order to further increase the trapping capacity of the NTs, the quartz surfaces were functionalised with broad-spectrum linkers. With the target of interest being cytoplasmic mRNA, polyT was chosen as a binding moiety, complementary to the polyA tails of mRNA. Functionalisation capacity was characterised by measuring accumulation of GFP-tagged RNA, from solution, at the tips of the NTs. Trapping capacities for 10 F and 10 UF NTs were measured across a range of V_{pp} – 5, 10, 15 and 20 V. In the presence of an electric field gradient at the tip of the NT, GFP-tagged RNA was attracted to the tip of the NT. RNA accumulation was visualised as a fluorescent signal, which was quantified using FIJI. Signal intensity was found to increase the longer the DEP force was applied for, with material dissipating as the DEP force was turned off (see supplementary **2.1** for an example trace). This

overall increase in fluorescence was observed after DEP was turned on, and with F NTs, confirms both successful NT functionalisation and that mRNA is in fact able to bind to the polyT linkers *via* their terminal polyA tail.

The difference between initial fluorescence signal (t = 0 s) and signal after 60 seconds of DEP being on (t = 60 s) were compared both (i) between F or UF NTs at various voltages and (ii) across F and UF NTs as the same voltage (figure 2.11). A significant trapping advantage could be identified with the F NT as compared to UF, as the poly-thymine tails promote binding of poly-adenine capped mRNA, with up to a 3-fold variation in change in fluorescence signal (at 15 V_{p-p}). Furthermore, very little change in fluorescence signal was observed upon increasing the applied voltage with UF NTs, with

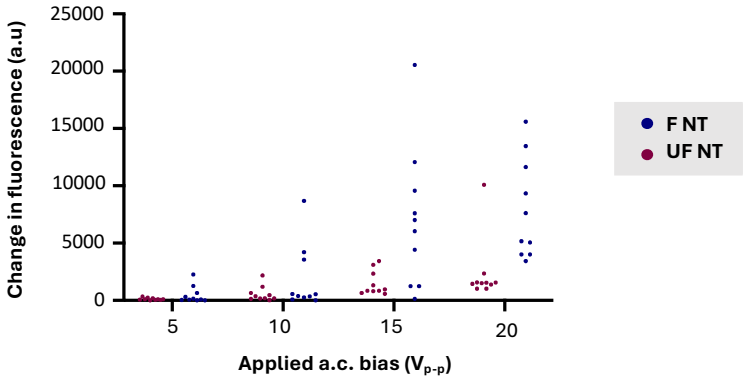


***Figure 2.11** **Validating trapping ability of nanotweezers in solution.** Both functionalised and unfunctionalised nanotweezers (NTs) were used to capture AF488-tagged mRNA from solution. Average fluorescent signals at the tips of the NTs were quantified across different applied a.c. biases (V_{p-p}). Data generated by Anthony Monteza Cabrejos, Edel/Ivanov Labs, Imperial College London.*

only a 2000 a.u. increase from 5 V to 15-20 V_{p-p}, as compared to a change of up to 22,000 a.u with the F NTs. Increasing the magnitude of the V_{p-p} does seem to increase the amount of RNA captured between 0-15 V_{p-p}, but shows minimal difference in trapping capacity between 15 V_{p-p} and 20 V_{p-p}. As such, all further experiments were conducted at 15V, to minimise effects of Joule heating effects. Variability in mean fluorescence values between NTs of same functionalisation condition and voltages could possibly be due to differences in NT aperture shape and variation in carbon coating. It is also worth noting that while mRNA trapping can be visualised at the tips of the NTs in these validation experiments, mRNA in the cell line used is not fluorescently labelled, and therefore cannot be visualised or quantified during extraction. This is further complicated by the inability to determine both the length and conformation of the captured mRNA, particularly the degree of coiling, which affects how readily the genes of interest can be accessed, thereby complicating gene detection. Given its closed-loop nature, RNA, with the accumulation of sufficient torsional stress, can form numerous topological structures, beside the standard secondary and tertiary folding, including coils, loops and knots.[113] As the RNA deviates from its linear primary structure, critical regions such as the terminal poly-A tails - which serve as the trapping target in functionalised NTs – may become partially or fully inaccessible, further reducing trapping efficiency and target detection sensitivity, Additionally, translation can occur on free-floating ribosomes within the cytoplasm or on ribosomes attached to the endoplasmic reticulum. This spatial separation of translation machinery may further influence the localisation and accessibility of specific mRNA species during sampling, adding another layer of complexity to effective transcript capture and detection.

2.4.3 Nanobiopsies: Part 1

Contributions:

- *Nanobiopsy experiments were conducted by Anthony Monteza-Cabrejos, Edel/Ivanov Lab, Imperial College London.*
- *All sample processing and analysis was conducted by Debjani Saha.*

All nanobiopsy experiments were conducted on nuc-GFP-MCF7 cells seeded into a 35 mm, high Grid-500 culture µ-dish (Ibidi) at a density of 5 x 10^3 cells/mL to facilitate cell tracking. In the first instance, five cells were probed with F NTs and 5 with UF NTs at a V_{pp} = 15 V. The same cells were probed a second time two days later. Extracts were collected by inserting the NT tips into PCR tubes containing 5 µL of nuclease-free water and 5 U of Superase-In RNase inhibitor (Invitrogen) and crushing the tip. To minimise loss of trapped RNA further, the function generator was only turned off once the sample had been collected. Samples were immediately snap-frozen, reverse transcribed on the same day using iScript RT and cDNA was stored at -20 °C until all samples were ready for qPCR processing. Each cDNA aliquot was split into four to test for four genes of interest - *ESR1*, *TFF1*, *GREB1* and *GAPDH*. Given the ultimate aim of this technique in this work is to enable the tracking of dynamic events in breast cancer as it adapts to therapy, with a focus on AI, three oestrogen-responsive genes were chosen as targets. While no studies using dormant cells were conducted as part of this thesis, detecting these genes in untreated MCF-7 cells, provides good insight into how efficiently genes that are downregulated in response to AI would be detected.

Initial experiments revealed very little amplification, if at all, for both F and UF NTs (figure 2.12 A), with only 11.25% (9 out of 80 wells), yielding positive amplification. Of these, only two, three and one F NT samples were amplified for *GAPDH*, *ESR1* and *GREB1*, respectively, and one and two UF NT samples were amplified for *ESR1* and *TFF1*, respectively (figure 2.12 B). For genes where both F and UF NTs yielded detectable amplification, F NTs yielded an overall larger percentage of positive amplifications as compared to the UF counterpart. Interestingly, for *ESR1*, the Ct values obtained for the F NTs were all a few cycles higher than for the UF samples (Ct(F, *ESR1*) = 35.6, 37.8, 37.2; Ct(UF, *ESR1*) = 34.9). Whether F NTs were in fact capturing more material than UF NTs could not be determined from this experiment alone, as no trend in Ct values could be identified. Furthermore, as mRNA was not fluorescently tagged, exctractions could not be quantified. Notably, all samples exhibiting positive amplification had Ct values above 34, very close to the widely accepted threshold of 35, above which signals are generally considered to be false positives. However, in this case, given the small amounts of material extracted from the cells, it was difficult to determine whether there were issues with the reagent combination or if these Ct values were truly representative of positive

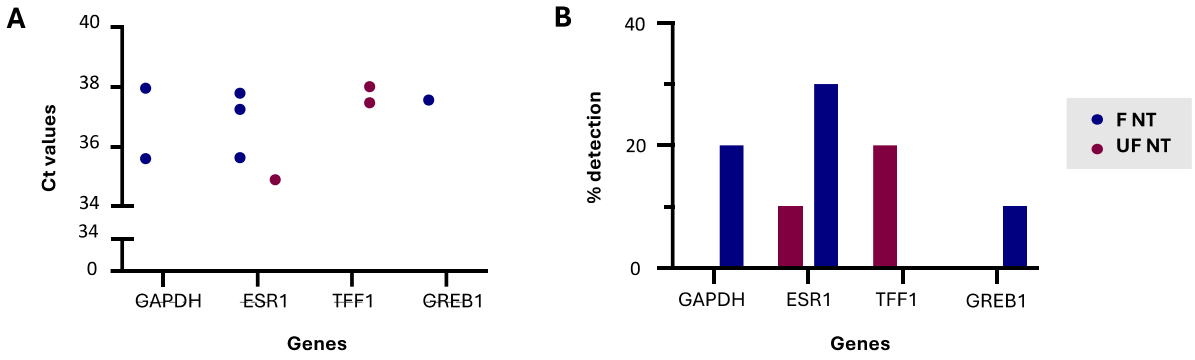


***Figure 2.12** Scatter plot and bar graph showing A **Ct values** and B **percentage of nanobiopsy samples resulting in detectable Ct values for both unfunctionalised (UF NT, blue) and functionalised (F NT, purple) nanotweezers.** RNA was reverse transcribed using iScript RT (Bio-Rad) and qPCR was conducted using SYBR Select (Applied Biosystems) using GAPDH, ESR1, TFF1 and GREB1 primers. **A** No Ct values below 34 cycles were obtained, and **B** less than 30% of samples resulted in detectable amplification.*

amplification, as no additional controls were tested alongside initial nanobiopsy extracts. Therefore, for the purpose of this experiment, melt curves of positively amplified samples were analysed collectively and considered true amplifications if melt curves were comparable. Any positively amplified samples with melt curves exhibiting non-specific amplification (either exclusively or in addition to a true amplification peak) were considered a false amplification and were omitted from analysis.

GAPDH is known to have relatively consistent expression both across cells and treatment conditions. With an abundance of functions in the cell, *GAPDH* expression is controlled by a complex interplay of different mechanisms, albeit all resulting in the synthesis and regulation of the same transcript.[114–116] Combined with its ubiquitous overexpression in cancers and its cytosolic abundance, *GAPDH* makes for a relatively reliable housekeeping gene for this cancer cell model and analytical platform.[116,117] Compared to the other genes of interest, all of which are oestrogen-dependent, not only should it have the highest expression in untreated MCF-7 cells, it should also be consistently detected in all nanobiopsy extracts. However, as seen in figure 2.12, not only was there no *GAPDH* detection from UF NTs, but only 20% of F NTs yielded detectable amplification. Furthermore, Ct values were comparable, if not higher than those of the other lower abundance genes. Difficulty in detecting *GAPDH* at this stage proves problematic for further dormancy experiments, wherein expression of oestrogen-responsive genes, which are already less abundant than *GAPDH* in untreated conditions, would be further decreased. Therefore, given the high Ct values (>34) and either lack of or non-specific amplification observed in the melt curves of close to 90% of total samples, it was assumed that there was inadequate amplification and that regardless of amount of RNA trapped, all processing steps could benefit from using more sensitive reagents.

2.4.4 RT-qPCR Reagent Optimisation

Following initial RT-qPCR conducted with 20 samples (10 F and 10 UF), with only a small percentage of samples yielding detectable Ct values, the RT-qPCR reagent combination was optimised. Two RTs - iScript RT (Bio-Rad) and LunaScript® (NEB) - and two qPCR master mixes - SYBR Select (Applied Biosystems) and Luna® (NEB) – were tested in all possible combinations as outlined in table 2.5. The reagent combination was initially optimised using *GAPDH* primers only, before expanding the primer pool and validating the chosen combination with all four genes of interest. RNA extracted from MCF-7 cells using the combined TriZOL/Chloroform method was serially diluted 7-fold to yield a range of aliquots containing between 1 µg and 0.1 pg RNA to hopefully mimic nanobiopsy concentrations. Serially diluted RNA was reverse transcribed using iScript RT and LunaScript® RT before both cDNA sample sets were qPCR amplified using both SYBR Select and LUNA qPCR master mixes. Starting RNA amounts were diluted 80-fold throughout the whole RT-qPCR protocol. While the RNA extracted from cells during nanobiopsies cannot be quantified at this stage - due to difficulties in detection using qPCR, the inability to visualise RNA transcripts and variations in nanopipette tip size and volume - optimising the reagent combination using dilute samples, ideally as close to nanobiopsy concentrations as possible, is highly beneficial for proof-of-concept and downstream processing. Ct values and melt curves were analysed for each sample, as in section **2.4.3**, with extra attention paid to samples with Ct values greater than 35, although these generally corresponded to the most dilute samples and fell within a good linear correlation range (R^2) of RNA dilutions. Calculated R^2 values reveal greatest linear correlation for the iScript RT and SYBR Select reagent combination. However, Ct values for the same concentrations were found to be consistently lower for the LunaScript® RT and Luna® qPCR master mix combination, suggesting that this combination may in fact be more sensitive. Despite having a lower R^2 value of 0.9268, this is still indicative of a good linear correlation. Furthermore, this assumes a 100% efficient RNA:cDNA conversion ratio, which may not always be the case, resulting in this skew. Additionally, the LunaScript® RT and Luna® qPCR master

mix had a shorter total run time of approximately 1h30 as compared to nearly 2h15 for the iScript RT and SYBR Select qPCR Master Mix reagent combination. Once it had been determined that the LunaScript® RT and Luna® qPCR Master Mix was the more sensitive reagent combination, standard curves were generated for all genes of interest (figure 2.13 B). Good linear correlation was obtained for all genes, ranging between $R^2 = 0.9291$ for GREB1 to $R^2 = 0.9722$ for GAPDH (table 2.6).

2.4.5 Nanobiopsies: Part II

As determined from section **2.4.4**, increased sensitivity in detection could be achieved by using the LunaScript® RT and Luna qPCR Master Mix combination. Therefore, all further nanobiopsy

***Table 2.5** **Overview of RT and qPCR master mix reagent combinations tested during reagent optimisation.** R^2 values (goodness-of-fit) were calculated using semi-log, non-linear regression in GraphPad Prism v10. Graph from which R^2 values were calculated is shown in figure 2.14 A.*

Reagent Combination		▼ 1	■ 2	■ 3	● 4
Reverse Transcriptase (RT)	iScript	✓	✓		
	LunaScript®			✓	✓
qPCR Master Mix (MM)	SYBR Select	✓		✓	
	Luna® qPCR		✓		✓
R^2		0.96	0.9715	0.9119	0.9268

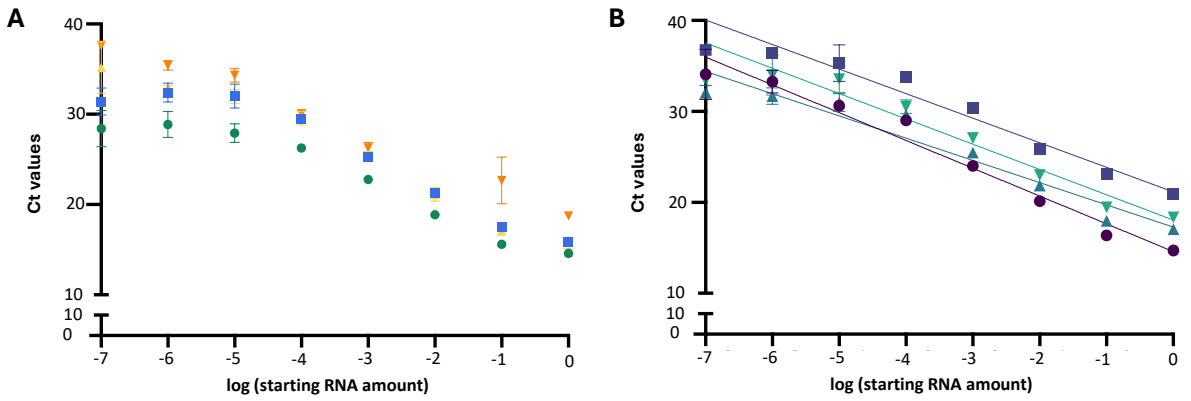


***Figure 2.13** **Standard curves for MCF-7 RNA, optimising the RT-qPCR reagent combination A** Four different reagent combinations (outlined in Table 2.5) were used to amplify serially diluted RNA in a 1:10 ratio (1 µg – 0.1 pg). RNA was 80x diluted from the starting amounts for qPCR reactions, using GAPDH primers only. **B** Once determined to be the optimal reagent combination, LunaScript® RT and Luna qPCR Master Mix were used to amplify and detect all four genes of interest – GAPDH, ESR1, TFF1 and GREB1. Mean Ct values are shown as icons, with standard deviations shown as error bars.*

***Table 2.6** **Summary of R^2 values for all four genes of interest – GAPDH, ESR1, TFF1 and GREB1.** R^2 values were calculated using semi-log, non-linear regression in GraphPad Prism v10. Graph from which R^2 values were calculated is shown in figure 2.15 B.*

Gene of Interest	● GAPDH	■ ESR1	▲ TFF1	▼ GREB1
R^2	0.9722	0.9445	0.9325	0.9291

processing experiments were conducted using this combination. Two additional samples were collected for every set of nanobiopsy experiment conducted: **(i)** a cell control, wherein a NT was inserted into the cell and held in place

for 15 seconds, without the application of an a.c. bias, and a **(ii)** media control, where a NT was positioned in the cell culture media in the sample vessel, and held in place, with a.c. bias applied for 15 seconds. These controls were taken predominantly to determine whether Ct values obtained from nanobiopsies were representative of true signals or were measuring noise. In cases where Ct values and/or melt curves of nanobiopsy extracts were similar to those of the controls, samples were disregarded from analysis. This could occur for one of many reasons, including, but not limited to, cellular material or debris being picked up by the poly-T linkers or quartz NT surfaces resulting in false positive signals, possible contamination, or primer amplification and dimer formation. Samples were processed alongside a 10 pg RNA control, as this was the lowest RNA concentration that was always reliably amplified during the RT-qPCR reagent optimisation experiments.

2.4.5.1 Single-Cell Nanobiopsies

All nanobiopsies from this section onwards were conducted as described in section **2.4.3**, however cDNA synthesis was conducted with LunaScript® RT mix and qPCR with Luna® qPCR Master Mix, as this was determined to be a more sensitive reagent combination than that initially used (section **2.4.4**). Nanobiopsies were repeated in the same manner as previously described (section **2.4.3**), using both F and UF NTs, ensuring that control samples were also acquired to increase reliability of RT-qPCR results. This included conducting nanobiopsies from the culture medium, inserting a NT into the cell but not applying any voltage through it, and using a 10 pg RNA and water RT-qPCR control. Nanobiopsies whose melt curves were similar to those of media controls, or differed significantly from the positive RNA control, were not considered for analysis.

Each bar in figure 2.14 (i), represented in a different colour, corresponds to a different day of experimentation, with the number of nanobiopsies conducted at each timepoint highlighted above the percentage of nanobiopsies resulting in gene detection in figure 2.14 (ii). Overall, approximately 26% of all samples yielded Ct values above 35, with very large variations in between nanobiopsy replicates conducted on the same day. While this is primarily due to different amounts of material being captured by each NT, this could be a result of uneven aperture sizes and irregular carbon deposition, which are not wholly accounted for in these experiments. Even with standardised pulling parameters, factors such as even minor slight misalignment of glass capillaries in the laser puller, as well as fluctuations in temperature and humidity, can lead to variations in NT pore size. Additionally, given that the carbon deposition process is largely manual, seemingly minor differences in the macroscopic procedure can translate into significant microscopic inconsistencies. Overall, very minimal difference in Ct values and percentage of total NTs yielding detectable amplification can be identified from these results, between F and UF NTs. The largest variation is observed for GREB1 and TFF1, wherein genes were detected in approximately 41% and 75% of nanobiopsies using F NTs as compared to 25% and 58% for UF NTs, respectively, with some detection observed close to 20 cycles with F NTs for GREB1. Furthermore, given the almost three-fold variation in fluorescence intensity observed when validating NT functionalisation using AF-488-tagged mRNA, this very slight trapping advantage in cells is quite poor and suggests that conducting nanobiopsies from cells is non-trivial. Interestingly, the lack of variation in GAPDH expression somewhat questions the reliability of these results. As a housekeeping gene, GAPDH expression should be considerably higher than that of the oestrogen-responsive genes, and should also be relatively consistently expressed across different cells. However, as seen in figure 2.14 (i), overall, GAPDH seems to have the highest average Ct values of the UF nanobiopsies, and no significant improvement to total % capture is observed upon using F NTs. Nonetheless, slightly lower Ct values for functionalised NTs suggest that there may be an advantage to functionalisation when conducting nanobiopsies from cells, although this may also stem from heterogeneity in gene expression, given that the F and UF extractions were not conducted

from the same cells. Additional complications may arise from the fact that the RNA collected via NTs undergoes approximately a 100-fold dilution — from the already minimal initial quantities — during the process

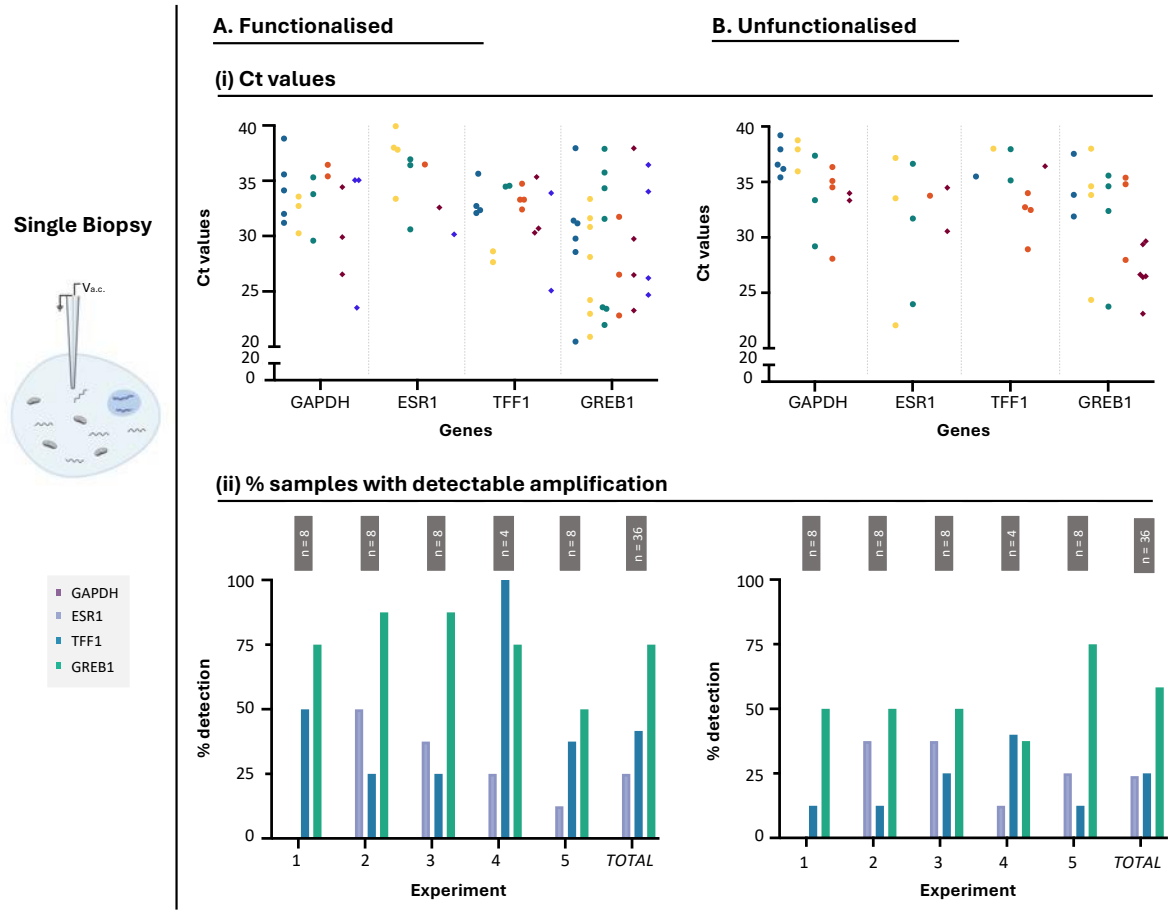


Figure 2.14 ***Ct values and percentage of total single nanobiopsy samples with detectable Ct values.*** *(i) Box and whisker plot with data points corresponding to Ct values obtained from nanobiopsies conducted with* ***A*** *functionalized and* ***B*** *unfunctionalised nanotweezers, processed using LunaScript® RT and Luna qPCR Master Mix. The percentage of nanobiopsy samples yielding detectable gene expression, both per experiment and as a total of all experiments, is shown in* ***(ii)****. For all graphs, each set of bar graphs corresponds to a different day of experiments, with day of experiment named arbitrarily. The number of nanobiopsies (n) conducted during each experiment are specified above graphs A(ii) and B(ii).*

from collection to qPCR due to reagent mixing. Moreover, since the amount of RNA initially captured cannot be directly quantified, the actual dilution could potentially be even greater.

This technique, therefore, presents a significant limitation in this context compared to other sub-cellular sampling methods such as FFM, which enable extraction and subsequent quantification of cytoplasmic volume. However, it is important to note that this technique does not extract any cellular material, and the NT tip is approximately 100 nm in diameter, thereby applying considerably less stress during membrane penetration. Alternative approaches could involve extending the trapping time to increase RNA capture (by saturating the tip). However, this introduces its own challenges, particularly the risk of elevated heat generation at the tip, which may compromise cell viability. Moreover, due to the inability to accurately quantify the volume extracted — which can vary between NTs due to differences in aperture size and carbon deposition — this remains a persistent limitation. Future work could focus on developing a more consistent NT pulling and carbon deposition process to achieve uniform pore sizes and electric field strengths. Additionally, engineering the cell line to fluorescently tag mRNA could enhance transcript visualization. However, this would necessitate highly sensitive microscopy to resolve subtle variations while avoiding pixel saturation.

2.4.5.2 Pooling multiple biopsies

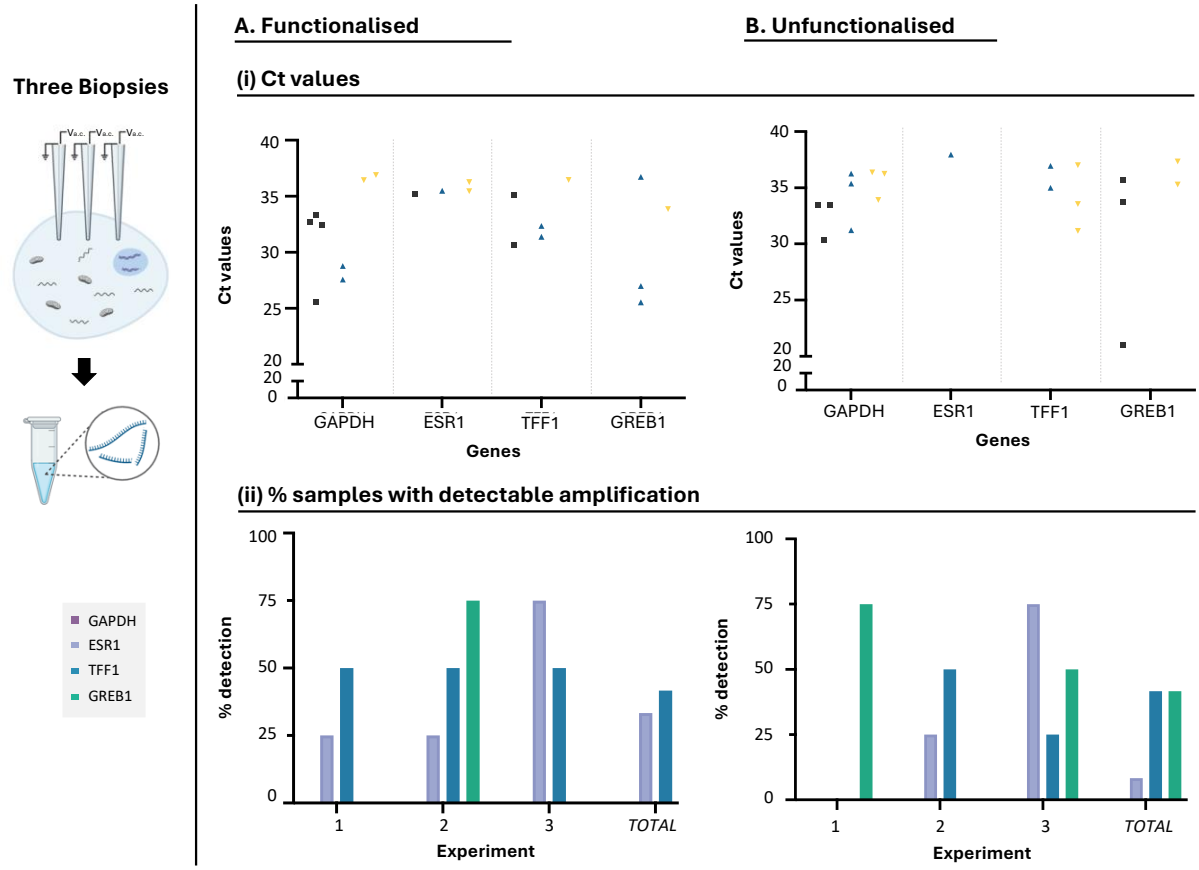
Given minimal improvement in nanobiopsy detection, it was decided to pool multiple nanobiopsies together into the same PCR tube, prior to conducting RT-qPCR, in an attempt to increase starting material and thereby push RNA concentration towards a detectable level. This was done by pooling (1) three nanobiopsies and (2) five nanobiopsies into the same tube. Notably, these nanobiopsies were conducted with separate nanotweezers, but from the same cell. While this pooling strategy is theoretically intended to enhance gene detection capabilities, it introduces the potential for variability due to inherent transcriptional heterogeneity between cells and induction of increased stress responses. Further precautions could have been taken in this regard, by using the same NT across pooled samples, however, RNA collection using the current technique requires the NT tip to be crushed into the PCR tube, which subsequently renders them unusable.

As can be seen in figure 2.15, a notable proportion of Ct values from both F and UF extracts were found to be above 35 cycles, for both three and five pooled nanobiopsies, indicating low levels of target gene expression. However, since all melt curves were analysed to confirm successful amplification, these data points were not excluded from analysis. Interestingly, despite theoretically containing three or five times more material than in single nanobiopsies, a significant difference in Ct values was not observed. Nonetheless, genes were detected in a larger overall percentage of samples, with GAPDH being detected in all samples on three occasions (figure 2.15 A, C and D (ii)). It is also worth noting that as additional nanobiopsies were performed on the same cell, signs of cellular stress became increasingly apparent, most notably in the form of membrane blebbing. This progressive stress response likely reflects cumulative mechanical strain induced by repeated membrane penetration and intracellular perturbation. In addition to compromising cell viability - with most cells unable to withstand more than three sequential samplings, suggesting a threshold beyond which homeostwatic mechanisms fail to compensate for the inflicted damage - this stress response may also alter the cell's transcriptomic profile. The activation of stress pathways and initiation of apoptosis can rapidly affect gene expression, potentially confounding downstream transcriptomic analyses by introducing stress-induced transcriptional changes that may not reflect the cell's original physiological state. As such, the lack of variation in Ct values from pooled samples could therefore be attributed to increased apoptotic gene expression as compared to our genes of interest.

While not explored in the scope of this study, material captured by NTs could have also been degraded by RNases present in media, which could be captured by NTs before membrane penetration, or cause degradation upon withdrawal of extracts. This is further complicated by the fact that some of the genes we are testing for, e.g. *TFF1,* are metabolic genes, and as such may start to degrade as soon as it is extracted from the cell, which could be accelerated in the presence of exonucleases and RNases in the culture media. PolyA tails on F NTs could also have been degraded prior to cell penetration, which could explain the similarity in gene expression across captures using F and UF NTs. Furthermore, while NP pulling parameters and the carbon deposition protocol was carefully controlled to achieve uniform aperture sizes, it is very difficult to accurately regulate this. As such, electric field gradient at the aperture tips varies, which, on a nanoscale, could result in variable nucleic acid capture. Additionally, qPCR primers used for amplification are all intron-spanning, and therefore only spliced gene variants can be detected. Optimisation of qPCR primers, including development of a primer mix comprising both intron and exon spanning primers for detection of both spliced and non-spliced genes, is essential. Combined, these results highlight the complexity of gene detection using the NT platform, opening avenues for further exploration and normalisation to ensure optimal, reliable, minimally invasive gene detection.

2.4.6 RNAScope

As seen from the nanobiopsy experiments, gene detection using NTs is non-trivial and extremely challenging, with less than 25% samples yielding detectable amplification with single biopsies in certain cases. Functionalisation and pooling multiple samples did not yield significant improvements and were compounded with additional challenges, including cells not tolerating probing beyond three sequential nanobiopsies, the inability to quantify extracted material and RNA degradation upon



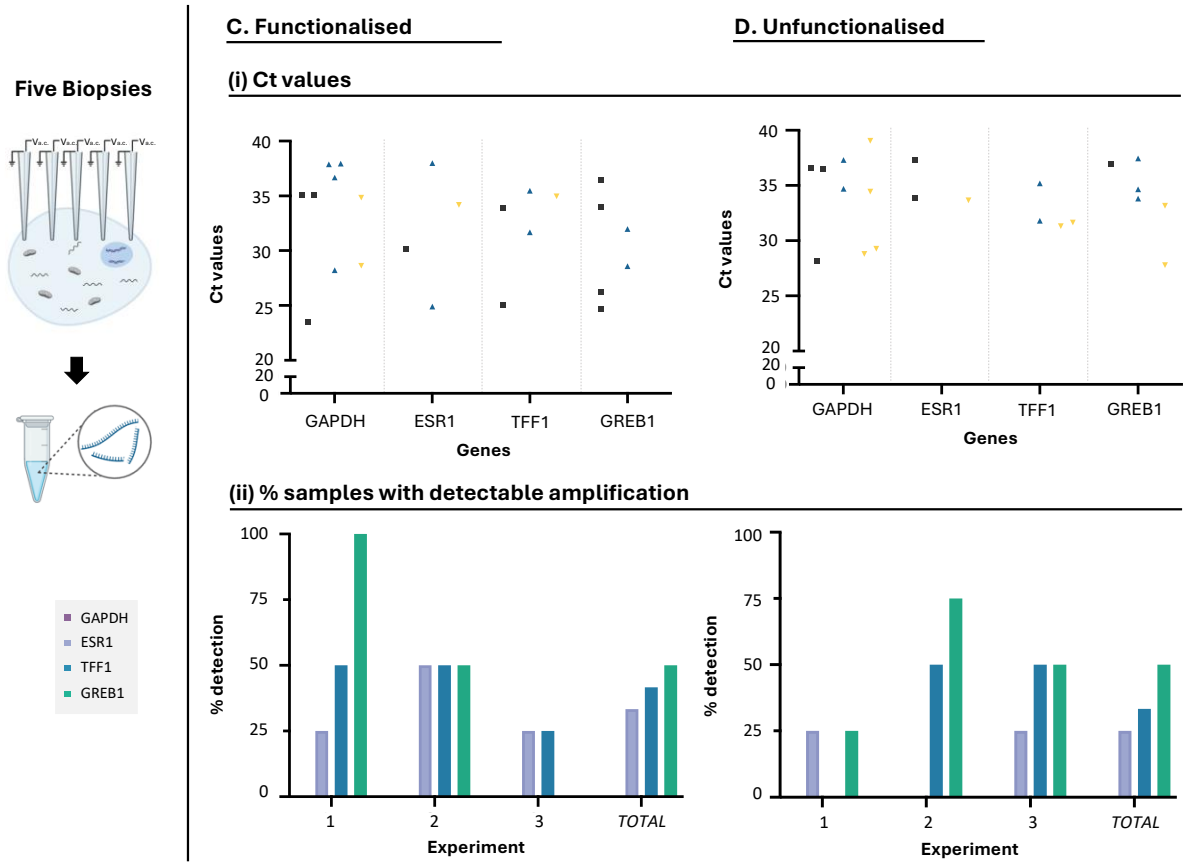


Figure 2.15 ***Ct values and percentage of pooled nanobiopsy samples, wherein three or five nanobiopsies were pooled into the same vessel prior to processing as a single reaction, with detectable Ct values.*** *(**i**) Box and whisker plots showing Ct values obtained after subjecting three (**A** and **B**) or five (**C** and **D**) pooled nanobiopsy extracts, conducted with both functionalized (A and C, respectively) and unfunctionalised nanotweezers (B and D, respectively), to RT-qPCR with the LunaScript® RT and Luna qPCR Master Mix reagent combination. Percentage of pooled samples yielding detectable Ct values are shown in **(ii)**, both in individual experiments (n = 4), and as a percentage of total pooled samples (n = 12). Each dataset corresponds to a different day of experiments, with day of experiment named arbitrarily.*

extraction. In order to further understand these variations in gene detection at a single-cell level and whether they were due to genes being localised in inaccessible regions of the cytosol or due to low gene expression, multiplexed RNAScope was carried out.

2.4.6.1 Experimental Set-up

Numerous seeding density experiments were set-up to determine the optimal conditions for long-term maintenance of cells in chamber slides (data not shown). Three main issues were identified: **(i)** cells tended to become either overconfluent for long-term experiments, even when seeded at low densities in AI conditions, **(iii)** would die if seeded at lower densities to those resulting in overconfluence, and **(iii)** as chamber slides do not provide a closed culture environment, growth media would evaporate very quickly, causing wells to dry out and the cells to subsequently die. To avoid this, cells were seeded in T75 flasks for the duration of AI treatment, detached and re-seeded into chamber slides two days prior to fixation. Cells generally tended to adhere well within chamber slides, regardless of duration of AI treatment, with minimal loss during pre-fixation washes and fixation.

While only the D0 timepoint was explored in the context of nanobiopsy extractions in this study, RNAScope was conducted for different AI durations (as shown in figure 2.8), with the rationale that results for different

durations of oestrogen-deprivation would give insight into how detection would need to be optimised when expanding this technique to working with dormant samples, especially when detecting genes that are downregulated in response to AI.

2.4.6.2 Image Analysis & Transcript Quantification

Image analysis can be conducted with a variety of softwares including FIJI, QuPath, CellProfiler and HALO (Indica labs).[118–121] For this thesis, analysis was conducted using QuPath. QuPath allows for automatic cell detection, with the accuracy of the cellular detection dependent on a number of user-defined thresholds. Wheat Germ Agglutinin conjugated with Alexa-Fluor 647 (Invitrogen) was initially used to stain the membrane of the cells (experimental outline not shown). However, this resulted in blurred images, which could not be remedied by deconvolution. This blurriness masked the transcript visualisation as punctate dotes, making transcript identification and quantification very challenging (data not shown). As such, membrane staining was omitted, and membrane boundaries were approximated using the in-built cell detection module in QuPath. Transcripts were assigned to cells using QuPath's nucleus and cell segmentation module. While somewhat stringent thresholds were applied for nucleus detection and defining membrane boundaries, these are both prone to error. Notably, as no membrane staining was used in experiments, membrane boundaries were assigned using the default setting for cell expansion (5 µm) - approximating the radius of an MCF-7 cell.[122] However, using this, the membrane is estimated to be simply an expansion of the cell's nucleus, and adopts the same shape, 5 µm away from the nucleus. This is adjusted for when neighbouring cells are within less than 5 µm distance away from each other, however, this may not be wholly accurate. Nonetheless, using the default value seemed to encapsulate most transcripts within membrane boundaries, with very few 'satellite' transcripts. The number of dots counted per cell was assessed as an indicator of total transcripts, as it was not possible to accurately determine the signal intensity of a single transcript. Despite these limitations, this method still provided a satisfactory basis for comparing gene expression across different treatment conditions and target genes.

Number of dots per cell across all genes and treatment conditions is summarised in figure 2.16 A, with specific genes and significance in variation across treatment conditions is shown in figure 2.16 B. A large amount of heterogeneity in gene expression is observed at a single cell level, even for *GAPDH*, which, as a housekeeping gene, is expected to have relatively stable expression regardless of treatment conditions. However, various studies have now revealed the unreliability of this as a housekeeping gene, especially when working with oestrogen-responsive cell lines, given its oestrogen regulation.[123] As such, further experiments (chapters **3** and **4**) were predominantly conducted with *β-ACTIN* as a housekeeping gene. RNase A treatment was also conducted on D0 samples, after which cells were hybridised only with a *GAPDH* probe, to determine specificity of probe binding, which is confirmed by the lack of transcripts in this treatment condition (figure 2.16 A). As can be seen, especially in figure 2.16 B, *ESR1, TFF1 and GREB1* expression are highly dependent on inclusion of E2 within the culture media, with large fluctuations seen upon -E2. Within as little as 48 hours of -E2, expression of these genes decreases significantly ($P \leq 0.0001$), with many cells having little to no *TFF1* and *GREB1* expression. In contrast, *ESR1* expression remains relatively stable during the first 48 hours of -E2, possibly due to epigenetic memory, but demonstrates increased variability over the course of the 28-day AI treatment. While *ESR1* expression does not immediately decline, its global expression does decrease with prolonged AI exposure. This is noteworthy given the overarching aim of this platform: to perform sequential nanobiopsies on individual cells at various stages of their adaptive response, from the onset of dormancy, through the dormant phase, and into cellular awakening.

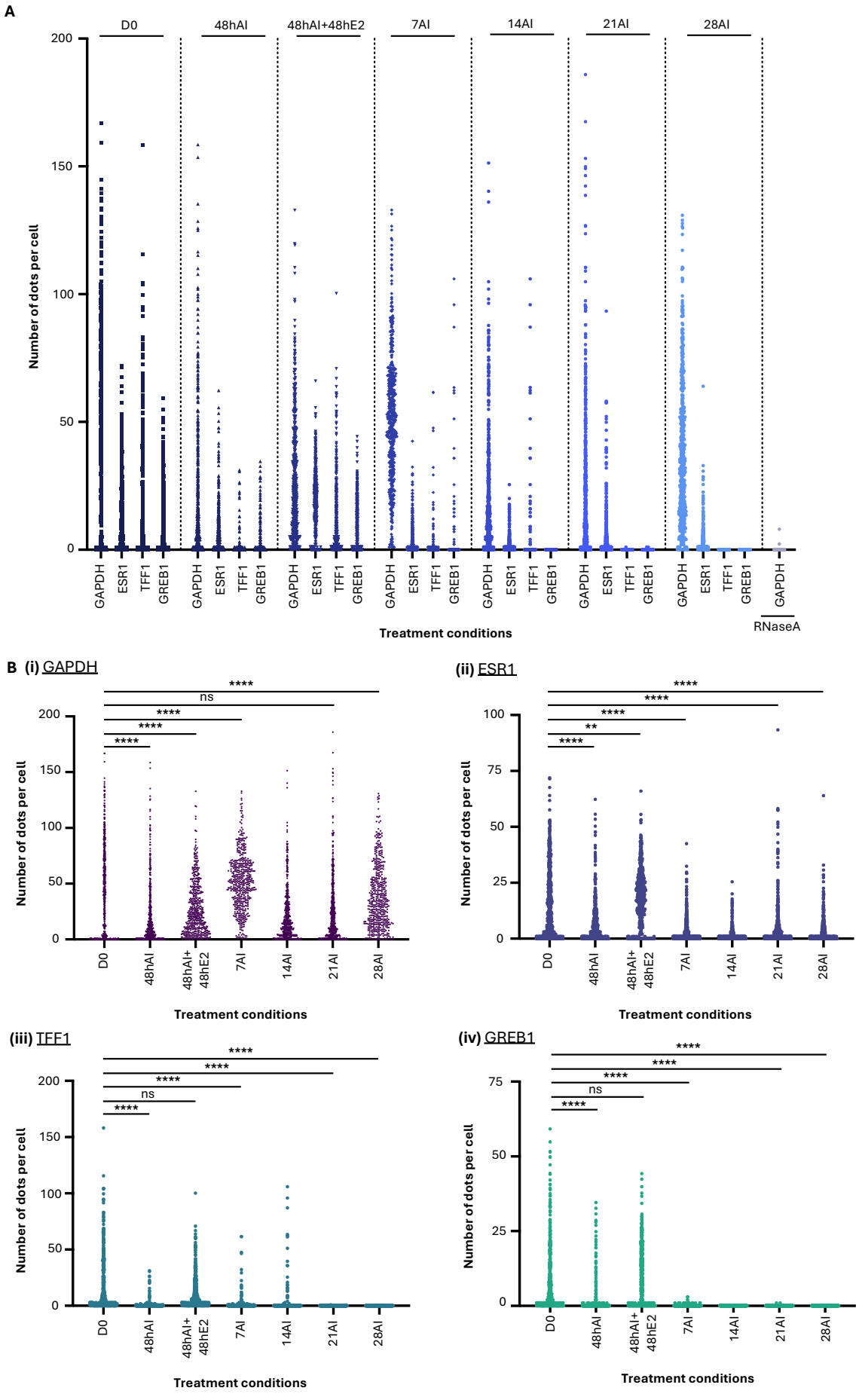

A
D0
48hAI
48hAI+48hE2
7AI
14AI
21AI
28AI
Number of dots per cell
GAPDH
ESR1
TFF1
GREB1
RNaseA
Treatment conditions
B (i) GAPDH
(ii) ESR1
(iii) TFF1
(iv) GREB1
ns
**
Treatment conditions

Figure 2.16* Number of dots counted per cell, for 500 cells, from RNAScope Assay. A** *Summary box-and-whisker plot of number of dots counted per cell using Qupath. All cells were treated in seven different ways: (1) untreated (D0), (2) 48 hours of AI (48AI), followed by 48 hours of oestrogen supplementation (3), and subjected to (4) 7 days, (5) 14 days, (6) 21 days and (7) 28 days of AI. RNAScope was conducting using the four probes of interest (same as for nanotweezer experiments) – GAPDH, ESR1, TFF1 and GREB1. One D0 GAPDH-treated well was treated with 0.1 mg/mL RNase A post-fixation to confirm selectivity of this technique.* ***B *Comparison of number of dots counted per cell per gene – **(i)** GAPDH, **(ii)** ESR1, **(iii)** TFF1 and **(iv)** GREB1. A Krukal-Wallis test was conducted in Graphpad Prism v10, and significance is shown on individual graphs - ns = no significance, $P > 0.05$; * = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$; **** = $P \leq 0.0001$.*

However, given the E2-dependence of key transcripts, which is notably depleted during dormancy, poses a significant challenge to detection sensitivity. Low transcript abundance, particularly for E2-responsive genes, may hinder reliable gene expression analysis during key transitional states, underscoring the need for enhanced sensitivity in downstream molecular assays or alternative biomarker strategies, especially when working with such small amounts of cellular material, as with NTs. Furthermore, looking at the gene localisation within the cells, *TFF1* and *ESR1* were generally found to be localised closer to the nucleus. While this is not a drawback in and of itself when working with whole single-cell analytical techniques, this does make dynamic single cell analysis quite tricky. The closer the genes are localised to the nucleus, the closer the NT would have to be approach the nucleus when conducting nanobiopsies. Given the fragility of MCF-7 cells, especially their nuclei, perturbing or rupturing the nuclear membrane can prove detrimental for cell viability. Killing a cell, in an attempt to access genes closer to the nucleus would significantly limit the dynamic nature of this technique. This is further compounded with the fact that translation can occur on free-floating ribosomes in the cytoplasm or those attached to the endoplasmic reticulum, as discussed in section **2.4.2.**[113] This gene localisation could also describe the variability in the Ct values obtained from the nanobiopsies. This is especially notable when compared to the recent success by Sahota *et al.* with extracting and sampling mRNA from hippocampal neurons using NTs (figure 2.1F).[94] While hippocampal neurons are generally more fragile than MCF-7 cells, neuronal mRNA is generally more abundantly expressed, with broader cytoplasmic distribution across the soma and dendrites. In contrast, as demonstrated by the RNAScope assays in this chapter, MCF-7 cells show more perinuclear mRNA localisation, which can also fluctuate based on cell cycle phases or treatment conditions. This spatial confinement can thus result in reduced mRNA availability within the cytoplasm, reducing nanobiopsy efficiency. This is further compounded by the stress-induced upregulation of RNases within MCF-7, which may accelerate faster mRNA degradation post-extraction. Furthermore, given the distance-dependent nature of DEP, *i.e.* the further the sample is from the electrode, the weaker the DEP force, if the NT is not positioned optimally within MCF-7 cells for trapping of nuclear-localised genes, there will be no minimal capture thereby introducing an inherent detection bias. Unless the targets of interest are labelled in some way, it is quite difficult to determine whether or not they are in fact being trapped. Overall, these results can also explain the lack of detection observed in the nanobiopsy samples - while some nanobiopsies did yield detectable amplification, those that did not could be due to the presence of significantly less transcripts within those cells.

2.5 Conclusions

Overall, it can be seen that the use of DEP enables manipulation of any particle irrespective of its net charge. Using probes with small electrode areas and nanoscale inter-electrode gaps enables pre-concentration of material at the probe tip at significantly lower operating voltages as compared to traditional EP systems, with minimal residual joule-heating effects, opening up numerous avenues for dynamically probing living cells. This chapter highlighted various steps taken to optimise the NT platform, developed by Nadapurram *et al.*, with an ultimate goal to enable the longitudinal tracking of MCF-7 cells undergoing AI.[1] Optimisation was attempted in

two ways: **(i)** by reducing the aperture size and inter-electrode distance of the NT, thereby increasing the DEP trapping force generated without increasing the applied voltage, and **(ii)** by functionalising the quartz surface of the NTs using polyT linkers, complementary to the polyA tails of mRNA, the target of interest. Together, these present methods by which greater amounts of material can be captured from cells, while minimising stress and maintaining viability.

Various issues were encountered, mainly with regard to sensitivity of detection, with many nanobiopsies either not yielding detectable amplification, or having Ct values close to or above 35 cycles. Very little advantage was conferred when conducting nanobiopsies upon functionalisation, contrary to that observed when validation the F NTs in solution. Capture efficiency could, in theory, be enhanced by functionalising NTs using random aptamers or gene-specific primers that would bind to specific RNA sequences. However, given the complex and dynamic secondary and tertiary structures that RNA can adopt, the target sequences intended for aptamer binding may be sterically inaccessible, especially under varying treatment conditions, reducing binding inefficiency. While this limitation also applies to the poly-T linkers used with the F NTs described in this chapter, poly-T sequences target the poly-A tails at the 3' terminus of mature mRNA, which tend to remain relatively unstructured and accessible, facilitating the selective enrichment of mRNA irrespective of transcript identity or secondary or tertiary structures. This increases the likelihood of successful capture and extraction, even in the presence of structural heterogeneity, and helps ensure consistent sampling across different cellular states. Additionally, the use of gene-specific primers or aptamers risks introducing detection bias. If certain transcripts become more structurally exposed or upregulated in response to treatment, they may be preferentially captured, skewing downstream transcriptomic analyses and reducing the platform's capacity to provide an unbiased snapshot of the cell's molecular state. This is especially problematic when attempting to monitor dynamic processes like treatment response, dormancy, or awakening, where changes in transcript abundance or accessibility can lead to misleading or erroneous assumptions. Therefore, despite their theoretical specificity, gene-specific aptamers would require tedious primer design, may lead to inherent capture bias, and ultimately offer no practical advantage when sampling heterogeneous or dynamic transcriptomes.

This lack of variation in detection and minimal detection can be attributed to several factors. These include inconsistencies in aperture size, which may arise from variability in nanopipette pulling parameters or in the carbon deposition process, leading to differences in electric field strength and, consequently, RNA capture efficiency. Additionally, inherent heterogeneity in gene expression between individual cells introduces biological variability that can obscure true expression patterns. There are further complicated by the susceptibility of RNA to degrade, particularly by exonucleases in culture medium and stress-induced RNases, which may be introduced at any point during sample, handling or processing. While techniques can be implemented to account or correct for these and subsequently optimise detection, these all warrant their own respective explorations and was out of the scope of this project. Nonetheless, these, which could include optimisation of the NT fabrication process, implementation of RNA stabilisation methods, live imaging of target transcripts by mRNA tagging, and co-staining for markers of stress or apoptosis, represent important avenues for future refinement of the nanobiopsy platform.

To probe whether variations in gene detection were due to transcript localisation within regions inaccessible to the NT tip, an RNAScope assay was conducted. This RNA imaging technique enabled the spatial mapping of selected transcript – namely *GAPDH, ESR1, TFF1* and *GREB1* – within the cytoplasm. Transcripts were generally found to be quite heterogeneously distributed in cytoplasm, although a large proportion were localised closer to the nucleus. This spatial variability introduces an additional layer of complexity, as the NT may not always be able to access perinuclear transcripts, without risking puncturing the nucleus, and could potentially describe the observed variations in gene detection, but may not be wholly responsible. Other confounding factors, including

differences in RNA confirmation, degradation, variability in the electric field gradient and intercellular heterogeneity, likely also play a role. Unless all transcripts of interest are fluorescently, or otherwise tagged to enable visualisation when conducting nanobiopsies, determination of whether mRNA, containing the transcript of interest, is being extracted will remain a challenge. Together, experiments conducted in this chapter show that NT trapping in cells is non-trivial and detection could be confounded by a number of factors, all of which must be explored further to enable use of this platform for longitudinal tracking of transcriptional changes. An alternative dynamic cell sampling technique is discussed in **Chapter 3** and an attempt at developing a possible solution to increase RNA concentrations, pre-RT-qPCR, with the aim of improving sensitivity and detection consistency is outlined in **Chapter 4**.

CHAPTER 3
REFINING PATCH-CLAMP TECHNIQUES FOR ENHANCED NANOBIOPSY APPLICATIONS

3.1 Graphical Summary

Table 3.1 ***Scanning ion conductance microscopy-guided nanopipette sampling modalities and recent applications in single-cell analysis.*** *Five approaches most relevant to this work are highlighted below, illustrating the progression from early designs to the experimental set-up described in this chapter. The techniques outlined rely either on extraction via electrowetting - that is by application of a voltage at a liquid-liquid interface, with the nanopipette filled with a water-immiscible organic solvent and immersed in an aqueous solution (or vice versa), enabling controlled aspiration of cytoplasmic material – or, electroporation – wherein a train of voltage pulses are applied whilst the nanopipette is positioned on the cell surface, bypassing membrane penetration and sheathing.*

Representative studies demonstrating the use of these modalities for high-resolution, minimally invasive RNA sampling are summarised, with particular emphasis on advancements incorporated into the system described in this chapter. [77,89,124–126]

Technique	Single cell surgery	Electrochemical attosyringe	Single cell-nanobiopsy; compartmental genomics	Single-cell nanobiopsy; longitudinal transcriptomics	This work; localised electroporation
Extraction technique	Electroporation	Electrowetting	Electrowetting	Electrowetting	Electroporation
Spatial Resolution	50 nm	10- 300 nm	50 – 100 nm	150 nm	160 nm
Sensitivity	N/A	aL - pL	50 fL (1 %)	≤ 200 fL	Under optimisation
Application/ Target extraction	Nanoinjections	Fluorescent injection	Cytosolic Ca^{2+} measurements, mitochondria isolation, GFP RNA sequencing	Cytoplasmic RNA for scRNA-seq	Cytoplasmic RNA for longitudinal scRNA-seq
Key Advantages	Avoids membrane shearing, high post-injection viability	Wide range of volume injections, high post-injection viability	Cell recovers within 5 s of aspiration, < 70 % viability	Longitudinal experiments (72h) with scRNA-seq	Precise and compartment-specific delivery into cell with < 100 % reproducibility
Cell Line	Human BJ fibroblasts	MCF-10A	Human BJ fibroblasts, HeLa	HeLa	MCF-7, immortal mouse melanocyte, human embryonic stem cells

3.2 Introduction

3.2.1 Patch-Clamp

Patch-clamp techniques, first developed in the 1970s by Sakmann and Neher, have revolutionised the field of electrophysiology by enabling precise electrical recordings at a single ion channel level. Their pioneering work has facilitated the direct, quantitative recording of ionic currents across cell membranes, laying the groundwork for vast advances in neuroscience, drug discovery and cardiac research, amongst others.[127–130] This technique involves positioning of a glass pipette (generally micro or nano (NP)) – the patch - on a cell's membrane. Upon contact, gentle suction results in the formation of a high-resistance (giga-ohm (GΩ)) seal, effectively electrically isolating the region of interest – the voltage-clamp. Depending on the magnitude of this suction, various configurations can be generated, resulting in isolation of small portions of membrane or access to the whole-cell. These configurations are outlined in figure 3.1 A. In brief, if the NP position is maintained after generation of a GΩ seal and the membrane remains intact, a 'cell-attached' recording can be conducted. This essentially generates a closed circuit, comprising pipette resistance (conferred by the resistance of the glass), resistance of the patch and of the seal (figure 3.1 B (ii)). Increasing suction to excise patches of the membrane results in 'inside-out' and outside-out' patches, and 'whole-cell recordings' enabling study of cells' extracellular and intracellular environments, and cytoplasmic access, respectively.[131] As access to the cell increases, such as during 'whole-cell' recordings, additional components are introduced into the electrical circuit, including a notable capacitive element (figure 3.1 B (iii)). In this configuration, the cell membrane, which has the ability to store charge thus acting as a capacitor, temporarily becomes part of the electrical circuit.[131] The effect of this capacitive behaviour can be seen in current traces as a time-delay in charging and discharging, causing a distinct spike in the current trace (section **3.4.2** and figure 3.5 B). Despite the sensitivity of this method, patch-clamp carries considerable limitations, chief among them being the reliance on a stable GΩ seal between the NP and the cell membrane. Failure to achieve or maintain this seal, whether due to irregularities in membrane surface, suboptimal pipette geometry, or contamination, can introduce current leakage. Even minor leaks at the NP-membrane interface can have major implications, significantly

distorting measurements, which if large enough, may result in inaccurate heterogeneity in measurements.[130] Moreover, formation of the seal itself introduces mechanical stress to the membrane, which may compromise cellular integrity, especially when working with fragile or adherent cells.

One significant advantage of the voltage-clamp technique, particularly compared to nanotweezers, is the ability to provide real-time electric feedback on the position of the NP relative to the cell, made possible by monitoring ionic currents flowing through the NP, which acts as a sensitive readout of spatial positioning. As the NP transitions through different positions - submerged in the bath, resting on the cell membrane, or penetrating the membrane with access to the cytoplasm - distinct variations in current, especially noise, can be detected. When the NP approaches the cell, the flow of ions is increasingly restricted, leading to noisier current recordings. This dynamic feedback allows for fine-tuned, incremental adjustments of NP position with a level of precision that is challenging to achieve with purely mechanical systems like NTs, notably reducing potential damage as a result of over-indentation. By enabling controlled membrane interaction, the voltage-clamp approach supports more consistent sampling conditions, reducing variability between experiments. This precision is particularly valuable in applications requiring high reproducibility, such as single-cell sampling or electrophysiological measurements, where even slight deviations in technique can lead to significant data variability.

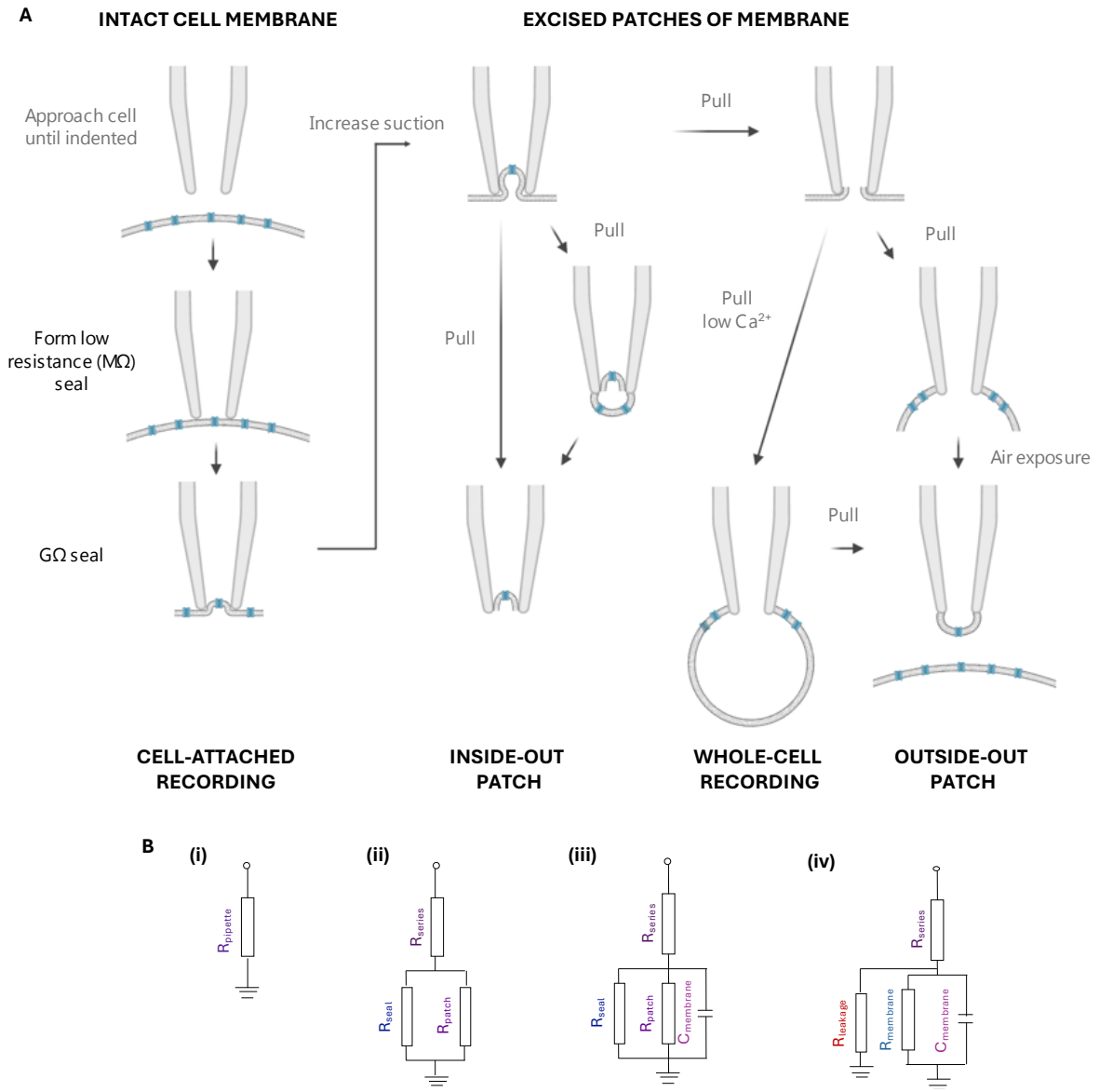


Figure 3.1 ***Overview of patch-clamp configurations.*** ***A*** *Four recording configurations are outlined (pink) – cell attached, outside-out patch, whole-cell and inside-out patch. All configurations stem from the pipette being in contact with the cell, with variations in applied suction and strength of pulls (pipette withdrawals) defining the final recording modality. Figure created with Biorender. Figure reproduced with permission from Springer Nature.*[132] *Copyright (1981) Springer Nature.* ***B*** *Circuit diagrams for different patch-clamp configurations are shown –* ***(i)*** *NP in bath,* ***(ii)*** *NP on cell (i.e. cell-attached recording),* ***(iii)*** *whole-cell recording and* ***(iv)*** *a whole-cell recording with large current leakage.*

3.2.2 Scanning Ion Conductance Microscopy (SICM)

An alternative technique, which bypasses the need for physical contact with cells, while enabling precise mapping of cellular topology and non-conductive surfaces, is scanning ion conductance microscopy (SICM).[133] Developed in 1989 by Hansma *et al.*, SICM involves scanning a fine probe over a sample, while maintaining a constant distance between the probe and the surface.[134] This control is maintained by a feedback system, similar to patch-clamp. This monitors the ion flow between two electrodes – one in a NP and another in the bath containing the sample – and moves the NP up or down to maintain a constant conductance. The NP is mounted onto a piezoelectric actuator, which enables conversion of electrical signals into displacements, and subsequently controls this fine movement in the x-, y- and z- directions.

SICM can be conducted in one of three modalities, varying in terms of feedback modulation: **(i)** non-modulated mode (figure 3.2 A), **(ii)** modulated mode, wherein distance or amplitude can be regulated, or **(iii)** hopping mode (figure 3.2 B).[133] When working in a non-modulated mode, a direct current (DC) is applied across the electrodes. On approaching the sample, current flow between the NP and the sample is hindered, resulting in an exponentially decaying current trace. The NP continues to approach the sample until the current reaches a preset set-point, at which point the Z-displacement is recorded. The NP is then moved in the x-y direction and this process is repeated at all x-y coordinates to yield a topographical image of the sample. However, this is not always seamless, as the DC current may drift, resulting in inconsistencies in the tip-sample distance or the tip crashing into the sample.[135] This can be addressed in modulated mode by the addition of an alternating current (AC) component. This focuses on the in-phase component of the AC current to control the NP-sample distance as opposed to simply displacing the piezo, additionally increasing the acquisition speed.[133] Further control can be induced in hopping mode - the mode of choice when working with samples with high surface roughness.[136] In this modality, an initial, low-resolution pre-scan is conducted to obtain a rough estimate of surface topography before a higher resolution scan is conducted (figure 3.2 C). After every measurement, the NP is retracted from the surface and approached again until the set point current is achieved, thus minimising the risk of damaging the NP and sample. Further modifications to this include development of an angular approach SICM to enable improved spatial control, incorporation with chemical moieties for extracellular pH (pHe) measurements (figure 3.2 D & E) and combining with electrowetting for single-cell nanobiopsies for transcriptional studies in cancer cells.[82,125,137]

This chapter outlines several proof-of-concept experiments utilising a combination of an angular-approach SICM and patch-clamp to probe cellular contents. The ultimate aim is to sample from MCF-7 cells to facilitate longitudinal tracking of transcriptional changes in response to therapeutic interventions. The technique allows for effective spatial control of the probe with respect to the cell, as current feedback enables monitoring of the probe's position. However, it is worth noting that this approach sacrifices molecular selectivity for broader sampling capability, which may in fact be advantageous when unbiased analysis of cytoplasmic constituents, such as transcriptomic or proteomic content, is required.

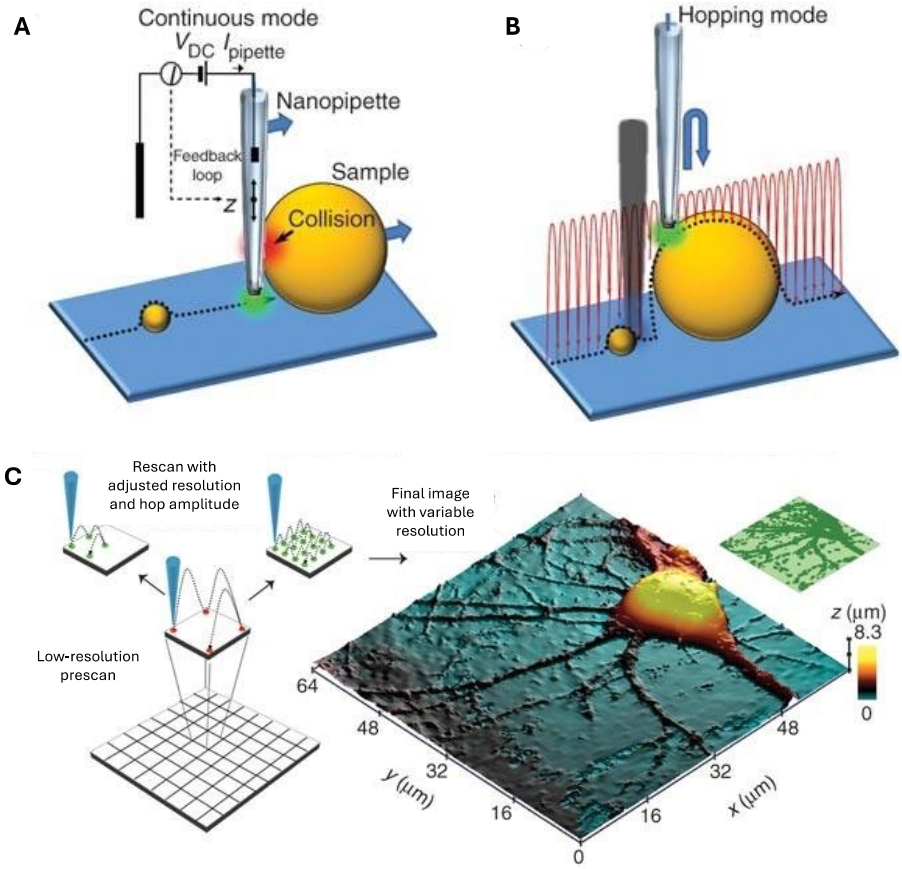


Figure 3.2 ***Scanning Ion Conductance Microscopy (SICM) Imaging modalities. A*** *An example of non-modulated SICM, wherein the nanopipette is seen to crash into the sample.* ***B*** *This is corrected for in hopping mode, wherein the NP approaches the sample from above and stops once the set point current is reached, avoiding any collisions. C This technique was used to generate a 3-dimensional topographic map of a fixed hippocampal neuron using an adaptive scanning algorithm. Figure reproduced with permission from Springer Nature.*[136] *Copyright (2009) Springer Nature.*

3.3 Materials & Methods

3.3.1 Cell Culture & Maintenance

MCF-7 cells were cultured and maintained as described in section **2.3.1.1,** until ready for use. 24 hours prior to conducting nanobiopsy experiments, cells were seeded into 35 mm TC-treated culture dishes (Corning®) at a density of 10 x 10^3 cells/mL and allowed to adhere.

3.3.2 Cell lysis and λ-DNA digestion

3.3.2.1 Cell lysis

40 x 10^5 cells were seeded in a T75 flask one day prior to extraction. Once adhered, cells were washed briefly twice with 1x PBS and lysed by addition of 400 mL lysis buffer containing 1% Triton X-100 (Sigma-Aldrich) in nuclease-free water (NEB) and 40 units Superase-In RNase inhibitor (Invitrogen). Lysate was transferred to a 1.5 mL tube (Eppendorf) and stored at +4 ˚C until ready for use on the same day.

3.3.2.2 λ-DNA digestion

1 µL HindIII (NEB) was added to a PCR tube placed on ice, containing 1 µg λ-DNA and 5 µL NEB buffer (NEB). The total volume was made up to 42 µL with nuclease-free water (NEB). λ-DNA was digested by incubation at 37 ˚C for 15 minutes, followed by enzyme inactivation at 80 ˚C for 20 minutes. HindIII-digested λ-DNA was stored at -20 ˚C until ready for use.

3.3.3 Nanobiopsies

All experiments described within this chapter were conducted using HPICM. Ion currents ere measured using an Axopatch 200B amplifier (Clampfit, Molecular Devices) with a gain of 1 mV/pA and with a 5 kHz low-pass filter setting. All data was acquired in gap-free mode, digitised and noise removed using the Digidata® 1440A low-noise data acquisition system (Molecular Devices). Current traces were analysed using pClamp10 (Molecular Devices).

3.3.3.1 Microscope set-up

An Ag/AgCl wire, connected to an ion current amplifier headstage, acting as the working electrode (WE) was thread through the back-end of the NP. The NP was secured into a microelectrode holder with a side pressure port (ESP-F10P, Harvard Apparatus), connected to a 2 mL syringe. The microelectrode holder was subsequently connected to a 3-dimensional piezo actuator (S-316.10 tip-tilt piezo scanner, enabling displacement in the x and y directions, and a P-753.3CD for movement in the z-direction, both Physik Instruments) and mounted onto a motorised micromanipulator (Patchstar, Scientifica). Samples were approached at an angle of approximately 33 °, as this was determined to be the optimal working angle when working with a 40x water objective by Shevchuk *et al.*[137] The sample was subsequently placed on the motorised stage of an upright microscope and a second Ag/AgCl wire was placed in the sample chamber (bath), as a reference electrode. The microscope set-up was placed on a vibration isolation table (PTM51509, Thorlabs) to minimise vibrational noise, and enclosed within a Faraday cage, to block electromagnetic interference and electrical noise.[138]

3.3.3.2 Nanopipette generation and characterisation

NPs were fabricated from single-barrelled borosilicate glass capillaries, with an inner diameter of 0.5 mm, outer diameter 1 mm and length 7.5 cm (BF100-50-75, Sutter Instruments), by CO_2 laser-assisted pulling (P-2000, Sutter Instruments), as described in section **2.3.1.1**. The protocol used to generate the NPs is outlined in table 3.2. However, glass capillaries were not plasma cleaned prior to pulling. Once fabricated, NPs were back-filled with either 100 µM fluorescein (100 µM FITC, 5 mM EGTA, 5 mM HEPES) diluted in 140 mM KCl, or 1x PBS (for DEP validation experiments). An Ag/AgCl was thread through the back-end of the NP and secured onto a micromanipulator (described in section **3.3.3.1**). As an initial validation step before performing any nanoinjections, the current through the nanopipettes (NPs) was measured at a potential of +200 mV. Only NPs with a current between 2 to 6 nA were deemed suitable for nanoinjections and used in further experiments. Pipettes exhibiting currents larger than 7 nA were discarded, as these readings often indicated issues such as compromised pipette tips. NPs with currents around 18 nA were considered to have broken tips, making them unsuitable for use.

Table 3.2 Two-line programme used to generate single-barrelled nanopipettes using the P-2000 laser puller.

	Heat	**Fil**	**Vel**	**Del**	**Pul**
Line 1	280	3	20	150	0
Line 2	300	4	15	120	100

Additionally, for all NPs used for nanobiopsies, aperture sizes were characterised electrochemically. NPs were placed in the bath solution and resulting current, upon application of voltages ranging from -0.5 V to + 0.5 V were measured. The current-voltage trace protocol is outlined in table 3.3.

Table 3.3 ***Current-voltage trace protocol.*** *Current changes in response to application of voltages from -0.5 to +0.5 V were measured.*

Type	Ramp	Ramp	Ramp
Sample Rate	Fast	Fast	Fast
First level (V)	-0.5	0.5	0
First duration (samples)	20000	40000	20000
First duration (ms)	2000	4000	2000

NPs exhibit Ohmic behaviour in certain voltage ranges, with ionic current, I, being proportional to the voltage bias applied through the NP, V. Given the inverse relationship between resistance, R, and conductance, G, NP conductance can be calculated from the linear region of the current-voltage (I-V) trace as follows:

$$G = \frac{I}{V} \qquad \textbf{(3.1)}$$

The value of G was subsequently used to calculate the aperture diameter, d using the following equation:[139]

$$d = \frac{4Gl}{\pi(g)(db)} \qquad \textbf{(3.2)}$$

where:

g: conductance of electrolyte, in this case MCF-7 cytoplasm = 12.99 S/m [140]

l: rough length of conical section of NP, 1.5 mm [141]

db: length of capillary before convergence, estimated to be 300 µm [141]

3.3.3.3 Patch-clamp experiments

Once the NP had been secured into the micromanipulator, it was positioned at the surface of the bath solution. Cells were approached with the NP in feedback mode, with a hopping amplitude of 1 µm. Unless otherwise specified, a potential of + 200 mV was applied through the NP during approach. The distance of the NP from the cell was monitored both visually (in brightfield mode) and by monitoring the current feedback - the distinct hopping signal was slowly lost and the signal became noisier as the NP approached the cell. Once the NP was positioned on the cell membrane, the patch-clamp control was switched to 'manual' mode, which retracted the NP from the cell by 1 µm. Using a step size of 100 nm, the NP was manually adjusted back to the original position (using 10 steps of 100 nm). The cell membrane was then penetrated in one of two ways, either (i) by indenting at least a further 2 µm before increasing indentation until the NP pierced the membrane, or (ii) by indenting the cell by 2 µm and lightly tapping the microscope stage gently with a pencil.

Two variations in terms of the stage at which potential applied was switched from +200 mV to -200 mV, were attempted. In the first instance, potential was changed to -200 mV upon cell membrane indentation and maintained until penetration. The second technique involved changing direction of potential to -200 mV only after the NP had penetrated the cell. In both cases, the direction of bias was again reversed to +200 mV and negative pressure was applied to extract material. Nanobiopsy extracts were collected in PCR tubes containing 5 µL nuclease-free water (NEB) and 5 U Superase-In RNase inhibitor (Invitrogen) by application of positive pressure.

Samples were briefly spun down and stored at -20 °C or -80 °C, depending on whether they were processed on the same day, or not, respectively.

3.3.3.4 Electroporation

3.3.3.4(a) in cells

Cells were approached in 'feedback mode' as outlined in section **3.3.3.3**. However, for electroporation experiments, once put into 'manual mode', the cell membrane was immediately indented by 2 µm. A series of 0 to + 10 V pulses were applied to the cell, without penetration, at a frequency of 50 Hz. A pulse of the same magnitude but opposite sign was applied to extract material from the cells. Samples were collected in PCR tubes as outlined in section **3.3.3.3** and were immediately processed by RT-qPCR.

3.3.3.4(b) in lysate

Experiments in this section were conducted using a gain of 0.5 mV/pA and with a 2 kHz low-pass filter setting. 10 µL Hind-III digested λ-DNA (section **3.3.2.2**) was mixed with 90 µL cell lysate (section **3.3.2.1**) to generate the lysate mix for use as the bath solution. NPs were backfilled with 1X PBS, positioned in the lysate, and a series of pulses – either 0 to +10 V or 0 to – 10 V - was applied to the sample to extract material. Samples were collected as in section **3.3.3.3** and immediately processed using PCR (section **3.3.4.2**). The protocol used for electroporation is outlined in table 3.4.

Table 3.4 Protocol used for application of short pulses to cells (electroporation) or to extract sample from the HindIII-digested λ-DNA MCF-7 lysate mix.

Type	Step	Step	Step
Sample Rate	Fast	Fast	Fast
First level (V)	0	10	0
First duration (ms)	7	3	7

3.3.4 RT-qPCR

3.3.4.1 Nanobiopsy extracts

Samples were reverse transcribed using LunaScript® RT Mix as described in section **2.3.4.1 (a)**. Resulting cDNA was taken directly (without dilution) and subjected to qPCR amplification using Luna® qPCR Master Mix, as outlined in section **2.3.4.2 (b)**, testing for four housekeeping genes: *GAPDH, β-ACTIN, EGFR and PUM1* (sequences shown in table 3.5).

Table 3.5 Forward and reverse primer sequences used for the four housekeeping genes used in all qPCR experiments conducted on nanobiopsy extracts in this chapter: GAPDH, β-ACTIN, EGFR and PUM1.

Gene	**Forward Primer Sequence**	**Reverse Primer Sequence**
GAPDH	5'-GGTCACCAGGGCTGCTTTTA-3'	5'-TTCCCGTTCTCAGCCTTGAC-3'
β-ACTIN	5'-CACTGGCATCGTGATGGACT-3'	5'- TTCTGCATCCTGTCGGCAAT -3'
EGFR	5'- AACTGTGAGGTGGTCCTTGG-3'	5'-GTTGAGGGCAATGAGGACAT-3'
PUM1	5'-GTCCTGTCATTGGCACTACAGAT-3'	5'-TTCCCGAACCATCTCATTCT-3'

Primers were used at a concentration of 2.5 µM. All nanobiopsy samples were processed alongside **(a)** a media control, wherein a nanoextraction was conducted from the culture medium, **(b)** a fluorescein only sample and **(c)** a 0.1 ng/µL RNA control.

3.3.4.2 Detecting HindIII-digested λ-DNA

Given that the target of interest from these validation experiments is one of the eight fragments resulting from HindIII-digestion of λ-DNA, extracts were qPCR amplified (without reverse transcription). A primer pair was designed for a 143 bp amplicon in the terminal λ-DNA fragment. This was initially validated on serially diluted λ-DNA and HindIII-digested λ-DNA at concentrations ranging from 0.05 µg/µL to 5 ag/µL. Primers were used at a concentration of 2.5 µM, sequences of which are outlined in table 3.6.

Table 3.6 Forward and reverse primer sequence of primer combination used to detect HindIII-digested λ-DNA, mixed into MCF-7 cell lysate.

Target	Forward Primer Sequence	Reverse Primer Sequence
λ-DNA	5'- TGGTTGCGTAACAAAGTGCG -3'	5'- CCATGCGCTTGCTCTTCATC -3'

3.4 Results & Discussion

Contributions:

- *All optimisation, patch-clamp and validation tests done as part of this chapter were done in collaboration with Fabio Marcuccio, Korchev Lab, Department of Metabolism, Digestion and Reproduction, Imperial College London.*

3.4.1 Patch-Clamp Nanoinjections & Nanobiopsies

3.4.1.1 Membrane indentation to access cytoplasm

As all cells behave differently and current traces will inevitably vary between cell types, previously validated characteristic traces may not be accurate determinants of electrode positioning with respect to the cell. As the electrode approaches and indents the cell membrane, the current trace becomes increasingly noisy due to the blockage of ion flow at the cell surface. Upon entering the cell, we expect even more significant feedback noise, as the cytoplasm is densely packed with molecules and organelles, which further hinder ion movement, creating a "busy" signal. This overlapping noise signal during both surface contact and intracellular penetration complicates signal differentiation, requiring careful interpretation of current feedback to confirm accurate cell entry. Therefore, for preliminary experiments, the NP was backfilled with 100 µM fluorescein. This enabled a visual read-out, as accumulation of fluorescence within both the nucleus and cytoplasm cell, serving as confirmation that the cell had indeed been penetrated. This dual cytoplasmic and nuclear fluorescence signal is facilitated by the small molecular size of fluorescein (332 Da), which permits rapid diffusion across cellular compartments, including the nuclear envelope. A voltage of + 200 mV was applied across the electrodes such that a current trace could be generated, and electrode positioning could be monitored. NPs were positioned at the surface of the cell and gradually indented (100 nm at a time, figure 3.3 A) until the current trace was dominated by noise (approximately 2 µm; figure 3.3 B). The potential was changed to -200 mV to enable ion and fluorescein flow out of the NP and the membrane was further indented until the NP penetrated the cell (whole-cell configuration; figure 3.1); however, this method proved inconsistent. In some instances, even with penetration beyond 2 µm, the NP failed to pierce the

membrane, suggesting that cells may not consistently respond to the same degree of indentation due to inherent cellular heterogeneity. Accordingly, variations were made to this technique and will be discussed later (section **3.4.1.2**). Cytoplasmic access was determined by almost instantaneous increase in fluorescence accumulation within the cell (figure 3.4 A). The potential was changed to + 200 mV and negative pressure was applied to the system to extract a portion of cytoplasmic volume (unquantified within the scope of this thesis) from the cell until fluorescence levels in the cells were similar to background, or until no further change was observed. The fluorescent signal in the cell, from injection until this point was analysed by Fabio Marcuccio, using a custom Python script (not included; figure 3.4 A).

NP extracts were collected, and samples were reverse transcribed, before a qPCR, testing for *GAPDH, β-ACTIN, EGFR and PUM1*, was conducted. Two media extractions, wherein the NP was placed in the culture vessel but outside of a cell and a fluorescein only control were also processed,

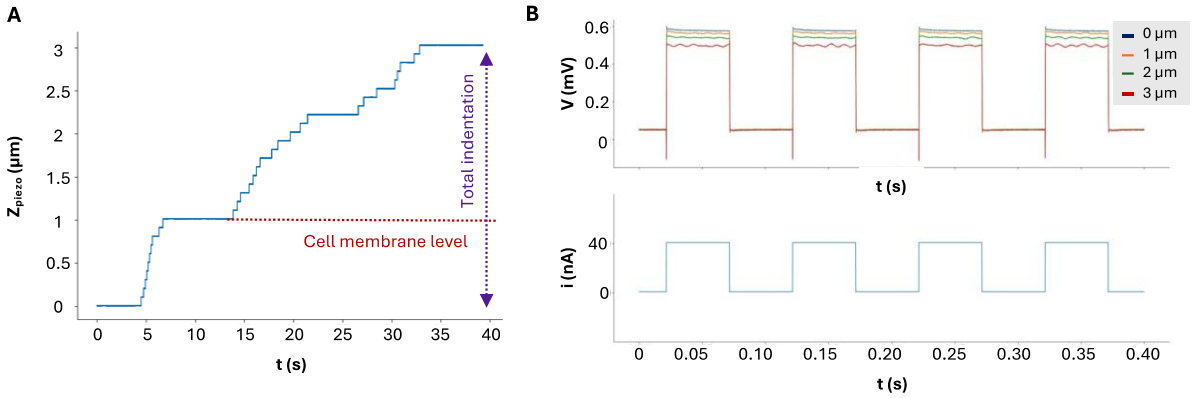


***Figure 3.3** Spatial control can be attained in patch-clamp experiments by monitoring Z-displacement height (Z_{piezo}) and corresponding current signals. **A** Displacement in the Z-direction upon entering manual mode, with 0 µm representing position upon retraction during switch from hopping and manual configuration. Cell membrane level and total indentation from retracted position of NP are shown in red and purple, respectively. **B** As cell membrane indentation is increased, current traces become increasingly noisy.*

but no Ct values were obtained for these, confirming either lack or extremely low abundance of transcripts. Results were analysed as in section **2.4.3**, considering both Ct values and melt curves of all samples and controls. From the three initial nanobiopsy extracts, only two yielded detectable amplification, both only for *GAPDH* and at significantly higher Ct values than in the RNA control. This could be due to insufficient material being extracted during the nanobiopsy process, rendering qPCR sensitivity inadequate.

Even when cellular penetration is achieved and cytoplasmic material is aspiration, quantification and validation of extracted volume remains challenging. The extraction process is inherently stochastic, subject to variations in applied pressure, NP aperture geometry, seal formation, and fluid dynamics at the nanoscale. As such we cannot rule out that variations in gene expression may also be due to variable extraction volumes. First, accurately assessing changes in cell volume as a measure of extracted volume is difficult, as cells are inherently three-dimensional structures, while our imaging setup is constrained to two-dimensional measurements. Second, the use of fluorescein-filled nanopipettes introduces a dilution factor, with any extracted cellular material immediately mixing with the internal pipette solution. Even if undiluted extracts could be quantified, the sensitivity required to accurately measure such minute volumes of biological material is currently beyond reliable analytical capability. While the use of an unfilled NP might avoid dilution, this introduces the risk of sample leakage upon withdrawal of the NP from the cell and would increase the mechanical stress on the cell given the need to penetrate the cell twice within a short time span. Capillary action and nonspecific adhesion of cytoplasmic material to the inner walls of the nanopipette present additional challenges for efficient sample recovery, often resulting in failure to expel the

collected material into PCR tubes for downstream analysis. Furthermore, the observed decrease in fluorescence might not necessarily be attributed to aspiration of cytoplasmic volume into the NP, but in fact due to leakage of cellular contents, as forming a perfect GΩ seal between the pipette and the cell membrane can be challenging. This is complicated by the presence of membrane structures, such as protruding proteins, which render the cell membrane uneven and result in the generation of a variety of patch structures, each with their own distinct characteristics.[142] Once a patch has been formed, the excised membrane is further influenced by capillary action, which can cause the membrane to migrate up the NP, complicating seal formation and resulting in leakage of aspirated material from the NP. These challenges are compounded by differences in pHe, cytoskeletal properties, and cellular heterogeneity.[142,143] Even when experiments are conducted under identical conditions, the variability in NP aperture sizes and differences in cell membrane properties can affect the formation of a stable GΩ seal, leading to inconsistent outcomes. While SICM offers the benefit of monitoring current signal and determining the position of the WE relative to the cell, distinguishing between different current signals can be challenging. When the NP is on the membrane surface, the signal becomes noisy due to restricted ion flow. Similarly, once the NP penetrates the cell, the high concentration of intracellular molecules also produces noise and a very similar characteristic current trace to when on the membrane. This makes it difficult to clearly distinguish between these two phases, and as a result, could be protruding too far into the membrane during the approach, further complicating the ability to accurately differentiate between being on the cell membrane and being fully inside the cell. Over-indentation of the membrane may pose issues with generating the GΩ seal, resulting in observed cytoplasmic leakage.

Moreover, as seen in figure 3.4 A (ii), only minimal change in fluorescence signal was observed upon application of negative pressure, suggesting that only a small quantity of material may have been extracted, or that a large amount of leakage occurred. A different complication was observed in figure 3.4 A (iii), where a membrane patch formed at the nanopipette tip before extraction. This required the application of a high positive pressure to remove the patch, before proceeding with the extraction. These issues underscore the challenges in maintaining consistent extraction conditions, with both negative and positive pressure playing critical roles in the success of cytoplasmic retrieval during the biopsy process. Notably, with this technique, no material is lost back into the media after extraction into the NP, as a membrane patch is formed at the end of the tip, preventing leakage. This is a distinct advantage over NTs, wherein material is lost exponentially back into solution as DEP is turned off. This highlights the need for application of positive pressure to expel material from NP, as membrane patch needs to be broken.

3.4.1.2 Mechanical disruption to facilitate cytoplasmic access

To address the difficulty in penetrating the membrane despite over 2 µm of indentation in some cases, a mechanical disruption technique was introduced. This involved lightly tapping on the microscope stage with a pencil after the cell had been indented by 2 µm. While the force of this could not be controlled for in the current set-up, subsequently introducing an additional source of variability, this reliably resulted in NP cell penetration. All other experimental parameters were kept consistent. Despite this, of the nine nanobiopsies conducted with this technique, seven yielded detectable Ct values (figure 3.4 B), significantly greater than that obtained without mechanical disruption. Of these, six yielded amplification for *GAPDH* only. However, these were all at much higher Ct values as compared to the positive control, indicating that the amount of RNA extracted was considerably lower than that of the control, as expected. Interestingly, for biopsy 6, we observe detection of *β-ACTIN*, albeit at a relatively high Ct value of 36. *GAPDH* expression is also at a slightly lower Ct value as compared to all other samples

(approximately 26). At this stage, as the technique had not been optimised and there were various confounding factors to account for, including the lack of control over

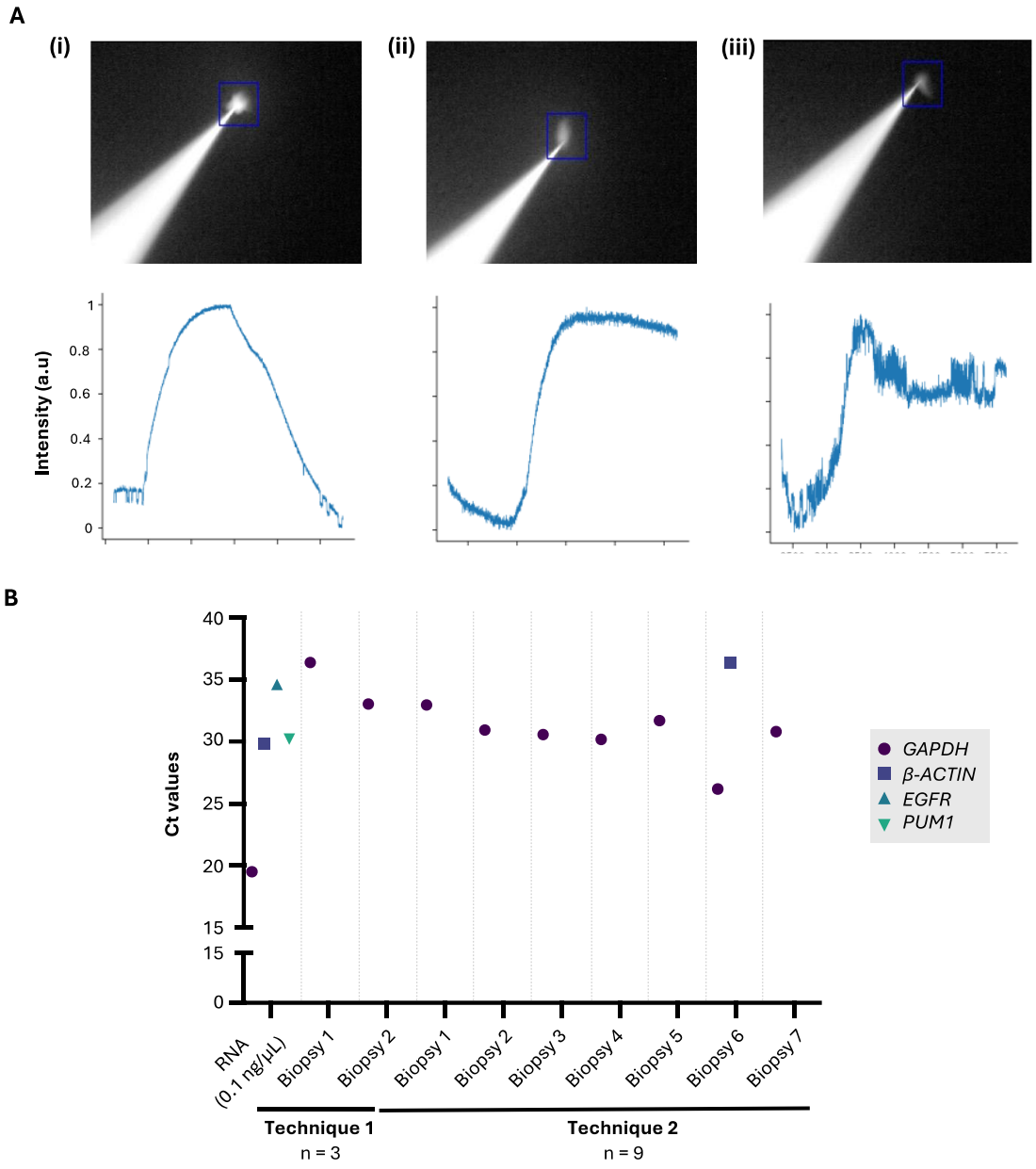


Figure 3.4 ***Fluorescence traces and Ct values obtained from nanobiopsies conducted in a whole-cell voltage clamp configuration. A*** *Cells were injected with 100 µM fluorescein using a NP by application of a negative current. Fluorescence signals were measured from nanoinjection and during extraction by application of negative pressure until all fluorescence had disappeared from the cell.* ***B*** *Samples were extracted and subjected to RT-qPCR, testing for GAPDH, β-ACTIN, EGFR and PUM1.*

the magnitude of the negative pressure applied to extract material from cell and the force of the mechanical disruption, this could not be confidently associated to cellular heterogeneity, and is most likely due to more material being extracted from cell based on the depth of NP penetration.

While access to the cytoplasm was facilitated using this technique, this approach faced several challenges. Until now, the potential was changed from + 200 mV to - 200 mV once the NP was positioned at the surface of the cell membrane, to enable fluorescein flow out of the nanopipette – visualised as a haze around the NP tip. However, in the case of biopsy 1, this haze almost immediately disappeared. Nonetheless, the cell was indented

and penetrated, yet no fluorescence accumulation was visualised, suggesting either immediate clogging of the NP aperture or membrane patch formation sealing the tip – a well-documented complication in patch-clamp methodologies. The nanobiopsy was therefore repeated using a second NP, this time shifting the potential to -200 mV only once the stage had been tapped and the NP was in the cell. Notably, given that the cell had been subjected to increased stress within such a short duration of time (two penetrations within approximately two minutes), the cell started blebbing. This phenomenon was also induced upon introduction of a negative potential in the cell, possibly as this promotes the entry of positive ions, especially Ca^{2+}, high concentrations of which has previously shown to trigger apoptosis.[144,145] Another issue encountered was related to the incomplete removal of fluorescence from the cell. In biopsy 3, fluorescein remained visible inside the cell after the initial extraction using a combination of +200 mV potential and negative pressure. Further extraction was therefore attempted using 0 V potential to ensure that any removal would be driven solely by positive pressure. However, the possible equilibrium between diffusion and electrophoresis resulted in only partial fluorescence removal, suggesting that a combination of forces may be necessary for efficient extraction.

Combined, these issues highlight the technical complexity of this technique need for optimisation of this technique, both in terms of controlling cytoplasmic access, avoiding membrane-patch formation prior to accessing the cytoplasm and increasing the sensitivity of detection to enable longitudinal probing of the same cell. While mechanical disruption *via* the manual tapping of the microscope stage provides a makeshift solution for achieving membrane penetration, the lack of control introduces additional non-trivial variables which were unaccounted for in the experiments conducted as part of this thesis, ultimately affecting cellular integrity and compromising the reliability of downstream data.

3.4.2 Monitoring cytoplasmic access using capacitance

Closer analysis of the current traces reveals the presence of a capacitive component as the NP penetrates the membrane (figure 3.5 B, in cell, vs figure 3.5 A, in bath). This is a characteristic feature exhibited upon application of a square wave in a whole-cell voltage clamp configuration, wherein the cell membrane temporarily behaves as a capacitor.[146] NPs were again filled with 100 µM fluorescein and were positioned on the cell membrane as previously described. Cells were indented by 2 µm and the microscope stage was lightly tapped to generate a whole-cell configuration and permit cytoplasmic access, visualised by an accumulation of fluorescence within the cell. While the technique could function without the use of fluorescein, it was used here as an additional measure to ensure successful cellular entry and monitor the process visually. No nanoextractions were performed during this validation step, as the primary focus was on assessing the reproducibility of visualising the capacitive component in the current trace. As the NP is withdrawn from the cell, the magnitude of this capacitive component is seen to decrease significantly, and we see an increase in the discharging component (as a spike at the terminal end of each square trace, figure 3.5 C).

The ideal scenario to enable dynamic measurements would be creating a leak-proof GΩ seal between the NP and the cell prevent leakage of ionic current, but achieving this during whole-cell recordings is incredibly difficult. Since the nanopipette must penetrate the cell membrane to access the cytoplasm, the perfect seal is often compromised. Minor leakage can be tolerated, allowing the cell to behave like a capacitor, maintaining its electrical properties. However, if the leakage becomes significant, it dominates the electrical resistance, overshadowing the cell's intrinsic resistance. This disrupts the recording, as the characteristic capacitance peaks – which indicate whether a whole-cell configuration has been generated - are no longer observed. Capacitance therefore acts as a reliable

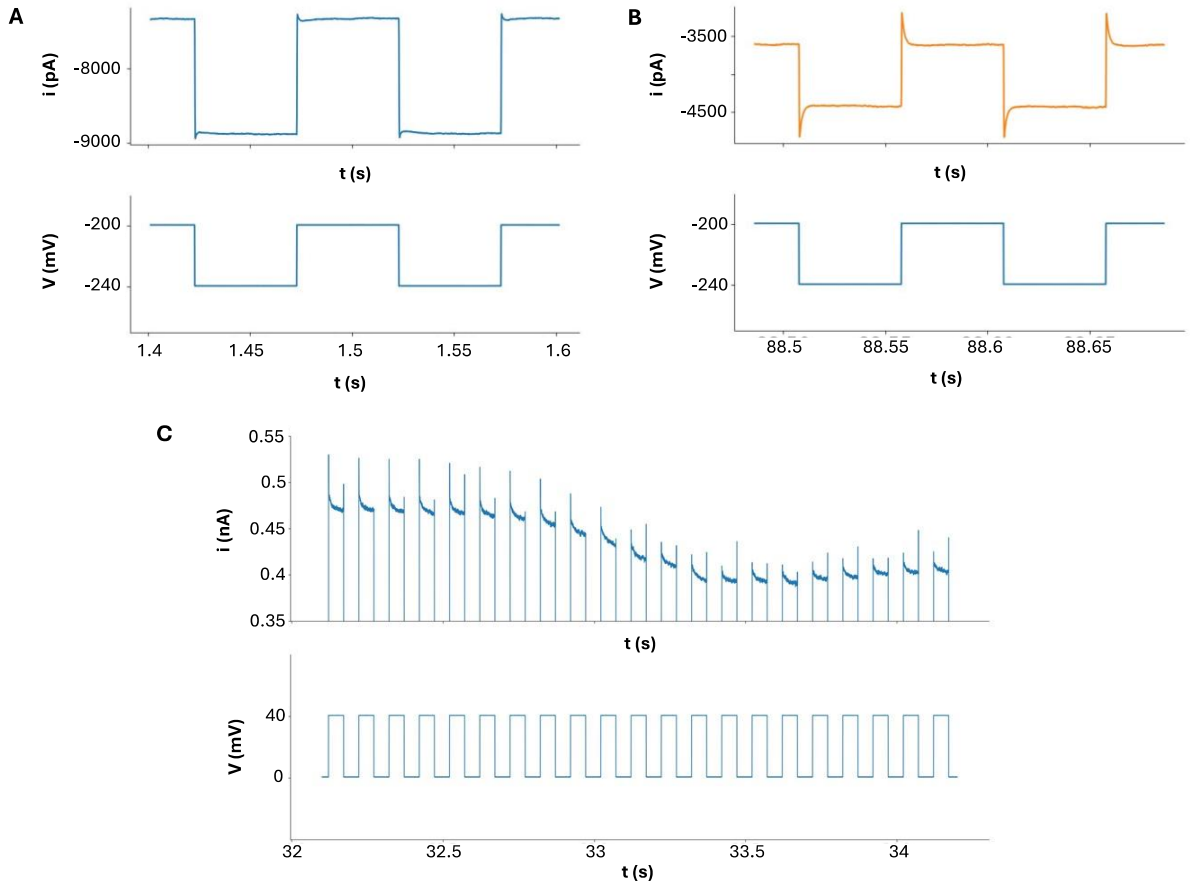


Figure 3.5 ***Ion current traces when the NP is*** *A* ***positioned in the bath,*** *B* ***has generated a whole-cell voltage clamp configuration and*** *C* ***when withdrawing the NP from the cell.*** *A distinct spike can be seen at the start and end of the square current traces in* ***B****, when a whole-cell voltage clamp configuration has been attained, enabling reliable distinction between when a NP is either in the bath or on the cell, vs when it is in the cell.* ***C*** *This capacitive component is seen to diminish as NP is withdrawn from cell membrane.*

indicator of whether a cell has been penetrated and a patch-clamp configuration has been generated. One notable limitation of this is the reliance on a physical tap on the microscope stage, once the NP is positioned on the cell membrane, to induce cell penetration. At present, as there is no way of regulating the force applied, reproducing this effect in an alternate setting and controlling for the degree of indentation is challenging, leading to potential variability and inconsistency in results.

3.4.3 Accessing cytoplasm by electroporation

To avoid this mechanical membrane disruption, an alternative technique, involving electroporation was attempted. The NP was positioned on the cell membrane in 'feedback' mode as previously described, switched to 'manual' mode and indented 3 µm in one go. Transient pores were generated by applying 100, 0 to -10 V pulses at a frequency of 50 Hz. The current traces were verified for the presence of a capacitive component both before and after injection to determine whether the NP had in fact penetrated the membrane. However, no capacitive component was identified, confirming that the whole process was conducted without membrane penetration and generation of a whole-cell voltage clamp. However, this also highlights a notable limitation with this set-up: the lack of reliable feedback to determine cytoplasmic access with this set-up. Without the ability to generate or detect whole-cell signatures, the technique relies heavily on indirect measures, such as fluorescence clearance, to infer successful extraction, limiting its utility in more controlled or quantitative settings.

To further probe whether the reduction in fluorescence observed in previous experiments (section **3.4.1**) was due to the actual removal of fluorescein from the cell, or simply a result of photobleaching, two cells were

directly compared. Fluorescein was injected into both cells *via* electroporation. Material was extracted from cell 1 by applying electroporation pulses of the opposite magnitude (0 to +10 V, figure 3.6 A), while cell 2 was left untouched and its fluorescence signal was tracked over time. Mean fluorescence signals of both cells were analysed using FIJI, and plotted in figure 3.6 B from frame 0 (being immediately upon application of the electroporation pulses) until no more change in fluorescence was observed. Given that mean fluorescence for nanoextractions settled to baseline levels within 25 frames as compared to 75 frames when allowed to photobleach, it can be confirmed that the decrease in fluorescence signal observed in both previous experiments and during nanoextractions using electroporation is in fact due to the removal of material from the cell. A third cell was injected with fluorescein and monitored periodically in both fluorescence and brightfield modes (data not shown). The fluorescent dye seemed to clear out of the cell within 20 minutes. However, between 10 and 20 minutes post-injection, the cell exhibited noticeable blebbing, as the cell presumably recovered from injection and attempted to expel the injected dye molecules.[147]

Electroporation could therefore prove to be a gentler method for cellular manipulation and extraction of cytoplasmic material as compared to creating a membrane patch using mechanical penetration. By applying a high voltage (approximately 10 V), electroporation introduces temporary pores in the membrane, allowing for the injection or extraction of material without causing long-term damage to the cell. This avoids the disruption caused by forcing a nanopipette through the membrane, which could induce undesired stress on the cell, affect cell viability and subsequently affect transcriptomic signatures, but does not necessarily preclude any impact from the high voltages applied to electroporate the membrane. Extractions by reversing the magnitude of the pulses could also act as a reliable alternative to extractions by application of a negative pressure, enabling increased control and reproducibility, while avoiding penetrating the cell and rupturing the membrane and avoiding issues such as variable aspiration forces, pipette clogging or capillary adhesion. However, given the ongoing issues with the lack of direct quantification of extracted volume makes it difficult to standardise this approach across various treatment conditions. Given the transient nature of the pores generated during both injection and extraction of material, the cell may recover from the sampling of its contents more quickly, making this approach more favourable when conducting long-term dynamic studies. Nonetheless, the applied voltages (±10 V) are relatively high by biological standards, and long-term effects of application of such voltages on transcriptomic fidelity and membrane composition remain poorly understood. Further work therefore requires the integration of electrical and optical monitoring of cytoplasmic access and the optimisation of voltage thresholds required for the exertion of minimal stress.

3.4.4 Validating extractions using DEP

To validate the feasibility of DEP-assisted extraction, a proof-of-concept experiment was conducted, wherein a known concentration of digested λ-DNA was mixed into an MCF-7 cell lysate

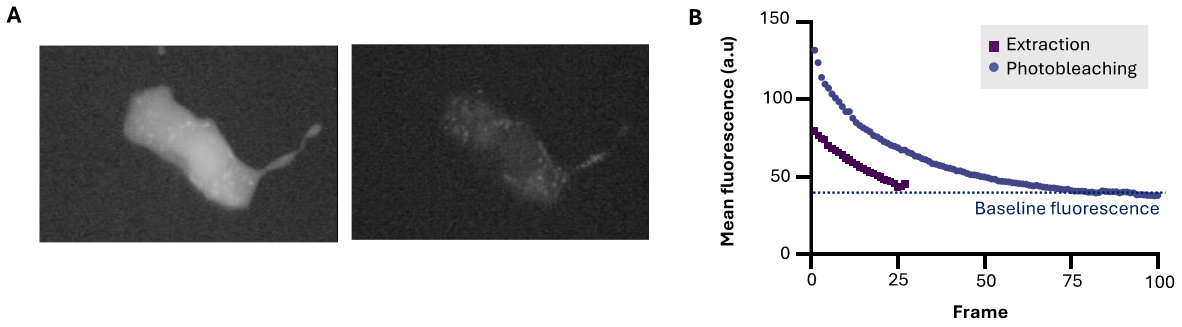


***Figure 3.6** Fluorescein injections and extractions using electroporation. **A** Fluorescein was injected into cells using a train of pulses from 0 to -10 V and extracted using pulses of the opposite magnitude (+ 10 V). **B** Number of frames required for fluorescence to settle at background levels when extracted using electroporation (purple) and by photobleaching (blue) are compared.*

solution. Initially, λ-DNA was cleaved using HindIII, for which it has 7 recognition sites. A primer set was designed specific to a 143 bp amplicon in the terminal fragment (highlighted in yellow in figure 3.7 A) and validated on serially diluted λ-DNA and HindIII-digested λ-DNA. Good linear correlation and primer efficiencies were observed for both samples – R^2 of 0.99 and amplification efficiencies of 99.8% for both samples (figure 3.7 B). Prior to conducting any collections, IV curves were recorded, with the NP placed in PBS, from which aperture size was calculated using equation **3.1** (figure 3.7 D). The NP was then positioned within the solution containing λ-DNA in cell lysate, and potential pulses, of the same magnitude and frequency as those used in electroporation experiments, was applied. In this case, both – 10 V and + 10 V were applied, as the charge of DNA in a non-uniform oscillating electric field varies depending on numerous factors, including electric field strength, gradient and frequency, as well as the properties of the medium it is dispersed in.[148] When conducted in solution, electroporation does not have the capacity to generate transient pores due to the lack of membranes and therefore acts solely as a series of short pulses that attract and aspirate material from solution. The number of pulses applied to solution was varied, to determine whether this would increase extraction from solution (and eventually cells). A control sample, wherein a NP was dipped in solution and immediately removed, was also obtained. Minimal potential (-20 mV) was applied across the electrode such that a current trace could be visualised and entry into solution could be determined. Samples were collected and subjected to qPCR amplification to detect the terminal Hind-III digested λ-DNA fragment. Application of 100 pulses at - 10 V significantly improved extraction efficiency compared to immersion alone. Furthermore, Ct values obtained (shown in figure 3.7 C as a log(DNA amount)) were comparable across all replicates for the immersion-only controls; however, the extraction replicates exhibited considerable variability, likely influenced by differences in aperture sizes, underscoring the sensitivity of this technique to even seemingly minor physical variations (figure 3.7 D). Increasing the number of negative pulses from 100 to 500 resulted in Ct values either close to, or higher than those from the immersion only samples (i.e. less material was extracted), indicating that application of too many pulses may not be beneficial. Notably, as the number of pulses applied through the NP was increased, apertures were more frequently clogged, possibly due to DNA crosslinking across apertures. This could be due to the fact that while a 10 V potential is applied to the electrode, F_{DEP} is proportional to V^2 (equation **2.4**). Therefore, significantly greater potentials are generated at the tip compared to those applied, which could be sufficient to induce cross-linking between DNA molecules. Effects of this relationship are also observed in figure 3.7 E, wherein we observe a weak linear correlation between the NP aperture size and the amount of DNA extracted. This demonstrates that variability can occur at multiple stages of the extraction protocol, all of which can contribute to the inconsistencies observed across different experiments. Additionally, the net charge and conformational of DNA present in the cell lysate remains unknown, raising questions about the effectiveness of applying a positive potential during the extraction process. While these findings show

that it is in fact possible to extract very dilute samples from mixtures and selectively detect them, they also highlight the delicate balance required in optimising pulse parameters for effective extraction. Therefore, prior to

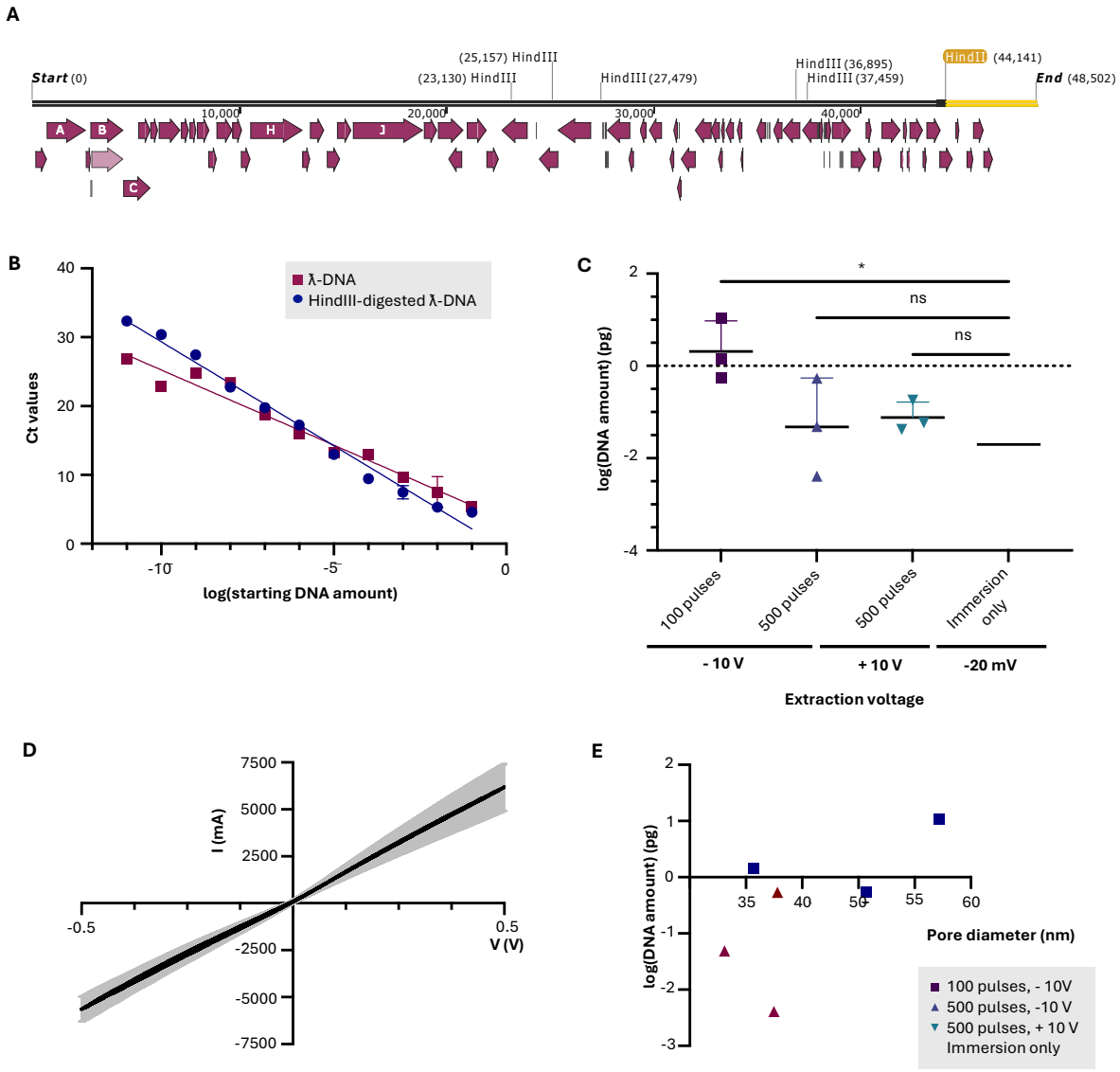


Figure 3.7 ***Validating extraction of specific λ-DNA sequences, added to MCF-7 cell lysate using DEP. A*** *λ DNA vector with HindIII restrictions sites. Terminal fragment (of interest) is shown in yellow. Vector map obtained from Snapgene.* ***B*** *Testing efficiency of primers, designed against a specific amplicon of the terminal cleaved fragment, on λ-DNA and HindIII-digested λ-DNA.* ***C*** *-10 V and + 10 V pulses were used to extract material from the digested λ-DNA-cell lysate mix. qPCR was conducted on extracts, for the detection of the terminal digested λ-DNA fragment. log(DNA amount) in picograms (pg) are shown for all experimental conditions.* ***D*** *Current-voltage traces were also obtained for all NPs prior to extraction, from which aperture sizes were calculated. The mean trace is shown in black, with shaded regions in grey to either side of the mean trace representing the standard deviation.* ***E*** *These were plotted against the log(DNA amount) in pg to determine the relationship between aperture size and amount of DNA captured.*

transitioning back to intracellular exploration, further investigation into the effects of pipette size and adjustments to pulse frequency are therefore warranted to determine optimal parameters for extraction from solution.

3.5 Conclusions

Using an angular SICM approach facilitates NP manoeuvring and subsequently minimises tip breakage that may occur as a result of approaching a cell too quickly. This is especially advantageous when working with long-term, longitudinal experiments, as cell penetration with compromised tip integrity, often resulting in oversized apertures, can induce unnecessarily large potentials and possibly result in the extraction of too much material for the cell to recover from, resulting in cell death. However, one notable limitation persists with this technique – incomplete ejection of extracted cellular material, seen through the retention of a fluorescence signal within cells and very high Ct values upon processing extracts using RT-qPCR. Precautions taken in this regard in the experiments conducted in this chapter include flushing the NP contents into lo-bind PCR tubes using positive pressure and use of fluorescence, monitored using fluorescence microscopy. Additional modifications that could facilitate the difficulties experienced with incomplete ejection of material from the NPs include surface modification of the nanopipette walls with hydrophilic coatings to minimise any hydrophobic interactions and development of pressure-controlled ejection systems.

As seen through blebbing, injection of cells with negative potentials promotes entry of positive ions into the cell, especially Ca^{2+}, which has negative effects on cell survival. Validation of this technique must therefore address effects of using Ca^{2+}-free media, as Ca^{2+} is essential for cellular adherence and junction formation.[149] However, this chapter demonstrates that electroporation serves as a reliable alternative to generation of a patch-clamp configuration, bypassing the need for membrane penetration and subsequently decreasing the influx of positive ions into cells. Furthermore, as nanobiopsies are conducted into NPs pre-filled with electrolyte, material extracted from cells is diluted immediately into the contents of the NP. Therefore, NP development could include inclusion of an immiscible oil-phase within the NP., enabling compartmentalisation of extracted material and minimising dilution. Crucially, this technique requires validation of detection, which has been partially optimized through λ-DNA sampling, though there is still a need to increase sensitivity. While cytoplasmic extracts from these samples have not yet been quantified due to the same challenges discussed in **Chapter 2**, future experiments should focus on maximizing cytoplasmic extraction or even removing whole single cells from culture to improve quantification and detection accuracy. This should be coupled with both recovery and viability tests, to determine how well cells are able to recover from stress induced by extracting large volumes of cytoplasm and whether this affects the transcriptional signature. Ultimately, this technique would be used to track single cells and how their transcriptomes adapt to therapy, with samples theoretically being sequences immediately from extraction.

The ability to monitor transcriptional changes over time is crucial in understanding cellular responses to therapy, especially in cancer research. Continuous tracking of these changes enables insight into the effectiveness of treatment strategies and further the development of new therapeutic interventions.

CHAPTER 4
ENHANCING RNA DETECTION THROUGH NUCLEIC ACID SEQUENCE-BASED AMPLIFICATION

4.1 Graphical Summary

Table 4.1 ***Comparative summary of four nucleic acid amplification techniques – PCR, NASBA, LAMP and RCA – detailing their reaction mechanisms, temperature and enzyme requirements, primer complexity, speed, sensitivity, specificity, strengths and limitations.*** *As discussed throughout this thesis, in the context of working with extremely low RNA input, such as those obtained from nanobiopsies, amplification efficiency, specificity and preservation of RNA integrity are paramount. PCR, while a gold-standard for nucleic acid amplification, relies on thermal cycling, which increases equipment complexity and may accelerate RNA degradation during high-temperature denaturation steps. In contrast, isothermal amplification, by maintaining a constant, moderate temperature, significantly reduce thermal stress on RNA. NASBA (Nucleic Acid Sequence-Based Amplification) was prioritised in this work due to its low incubation temperature (~41 °C), relatively straight-forward primer design requirements, direct RNA-to-RNA amplification capability, and compatibility with RT-qPCR for quantitative analysis – all characteristics which make NASBA well-suited to low input reactions where the ultimate goal is downstream RNA sequencing. Alternative approaches, such as LAMP (loop-mediated isothermal amplification) and RCA (rolling circle amplification) have limitations in this regard. While LAMP offers enhanced reaction speeds and high sensitivities, it produces complex stem-loop DNA structures, complicating subsequent sequencing and quantification. RCA on the other hand requires pre-circularised templates, is significantly slower and can be prone to incomplete amplification, providing more opportunity for RNA degradation. Section 4.2 provides an in-depth overview of both LAMP and NASBA, outlining their respective principles and advantages, and providing additional rationale for selecting NASBA as the preferred method for the experiments conducted in this chapter.*[150–152]

Feature	PCR	NASBA	LAMP	RCA
Template	(c)DNA	RNA	RNA/DNA	Circular DNA/RNA
Amplification	Exponential	Exponential	Exponential	Linear (or branched)
Temperature	Thermal cycling 95 – 60 - 72 °C	Isothermal 41 °C	Isothermal 60 - 65 °C	Isothermal 30 – 37 °C
Enzymes	DNA polymerase	RT, RNase H, T7 polymerase	RT, strand-displacing DNA polymerase	Strand-displacing polymerase
Primer Requirements	2 primers	2 - 3 primers	4 - 6 primers	1 – 2 primers
Detection	qPCR, gel, melt curve	qPCR, gel, fluorescence	Fluorescence, turbidity, gel	Fluorescence, gel, nanopore
Sensitivity	High	High (RNA only)	Very high (10– 100 copies)	Moderate - high
Specificity	High	High	High	Moderate
Speed	1 - 2 hours	90 – 120 min	30 – 60 minutes	1 - 3 hours
Strengths	Widely used, current gold standard	Direct RNA amplification	Fast, robust, compatible with field-use	Single-molecule, biosensing
Limitations	Requires thermal cycler, less ideal for low-input RNA	Enzyme-sensitive, low reproducibility	Primer complexity, non-specific amplification	Requires circular template, slow

4.2 Introduction

As seen from experiments discussed in chapters 2 and 3, the minute amounts of material extracted from cells remain below the reliable detection threshold of standard PCR amplification techniques, and often cannot be detected at all. Variations in the amount of material extracted, due to gene localisation (section **2.4.6.2**), heterogeneity in gene expression in single cancer cells and non-uniformity in nanotweezer aperture sizes, present additional obstacles to the already challenging gene detection from nanobiopsy extracts.

While improvements were observed in terms of mRNA extraction yields by functionalising nanotweezer surfaces using broad-spectrum poly-T linkers in solution, these effects could not be replicated in cells, suggesting intracellular conditions may impede effective mRNA capture. Additionally, increasing V_{pp} and nanotweezer and nanopipette aperture size as alternative methods to increase mRNA capture risk inducing joule heating effects and bubble formation, and increased membrane stress upon penetration, respectively, none of which are desired when attempting to conduct long-term dynamic studies. Combined, these severely limit the ability of such nanobiopsy techniques to be used to sequentially probe single cells and map their dynamic epigenetic profiles. Therefore, to overcome this bottleneck, an RNA amplification technique, to be conducted downstream of mRNA trapping, was attempted, with the rationale that increasing starting RNA template should enable increased cDNA generation and subsequent qPCR detection. While conceptually feasible, this approach introduces its own set of challenges, including amplification bias, template degradation during pre-processing, and the risk of false positives, all of which demand rigorous optimisation and validation before such protocols can be trusted for dynamic single-cell analysis.

4.2.1 Nucleic Acid Amplification Techniques (NAATs)

Nucleic acid amplification techniques (NAATs) are molecular methods which enable the amplification and detection of specific sequences of nucleic acids, including DNA and RNA at very low concentrations, from biological samples. The sensitivity of these techniques makes them very attractive for use in infectious disease testing and diagnostics, genetic analyses and vaccine potency testing.[153–155] Among the most widely used NAATs is the polymerase chain reaction (PCR) which involves cycling through multiple controlled heating and cooling cycles to amplify DNA exponentially. This can be combined with a reverse transcription step (RT-PCR), wherein RNA is initially reverse transcribed (RTed) to generate its complementary DNA (cDNA), before amplification. Amplified cDNA can also be continuously quantified at the end of each amplification cycle (qPCR), using either a non-specific DNA-binding dye, such as SYBR Green, or a sequence-specific fluorescent probe e.g. a molecular beacon (MB). The number of cycles after which this fluorescence signal surpasses a pre-defined threshold is termed the cycle threshold (Ct) and is inversely proportional to the amount of starting DNA. In cases where extremely small RNA amounts are RTed, cDNA generated, even after numerous amplification cycles, may fall below the detection limits of commercially available qPCR kits or fluorescence detection modules, limiting its application and introducing the risk of false negatives This also works under the assumption that reverse transcriptases (RTs) work with 100% efficiency, that is, one RNA molecule is converted to one cDNA molecule, which may not always be the case, due to secondary RNA structures, primer-template mismatches and enzyme variability. As such, RT-PCR methods are inherently prone to bias, particularly when working with degraded RNA or when working with sub-cellular RNA concentrations.

Further advancements in the field of NAATs has resulted in the generation of techniques that are able to amplify material at a fixed temperature, i.e. isothermally. These include loop-mediated isothermal amplification (LAMP) and nucleic-acid sequence based amplification (NASBA), amongst others.[156] These are especially beneficial when working with RNA templates given they are conducted at a stable, moderate temperature (typically between 37 to 65 °C), thereby minimising any potential RNA degradation stemming from drastic temperature changes, as required for qPCR assays. A summary of how these techniques work to amplify nucleic acids is shown in figure 4.1.

4.2.2.1 Loop-mediated isothermal amplification (LAMP)

LAMP is a rapid and highly specific NAAT that amplifies DNA or RNA isothermally (generally between 60-65 °C), eliminating the need for thermal cycling. Using a set of four to six distinct primers – two inner primers

for target recognition, two outer primers that aid initiation of the reaction and two optional loop primers that can accelerate the reaction by binding to loop structures generated during amplification - LAMP generates intricate stem-loop structures, with the aid of a strand-displacing DNA polymerase. Use of such a polymerase allows for exponential target amplification without the need for a denaturing step, enabling the generation of detectable amounts of DNA from starting concentrations as low as 10^7 copies/ml within 30 minutes.[151] Although primer design for LAMP can be complex, this NAAT technique has demonstrated exceptional specificity and sensitivity across a large dynamic range, and even supports the use of multiple primer sets for multiplexed detection.

The LAMP protocol occurs in three main stages (figure 4.1 A).[151] **(i)** Generation of stem-loop DNA structures, sequences of which are derived from the inner primers; **(ii)** initiation of strand-displacement by hybridisation of an inner primer to the looped structures to generate a second stem-loop twice the size of the original structure; and **(iii)** formation of cauliflower-like, stem-loop products, composed of alternating inverted repeats of the target sequence, which indicate successful amplification. Firstly, the inner primers, which contain complementary sequences at their 5' and 3' ends, bind to the target DNA. Once bound, a single-stranded DNA (ssDNA) is displaced by DNA polymerase, which can then act as a template for further priming. A second inner primer and an outer primer can then bind to this newly released ssDNA, generating a stem-loop structure. The reaction then enters its cyclic amplification phase, wherein binding to a ssDNA alternates between the two inner primers, resulting in the generation of up to 10^9 copies of initial target within an hour. Given that strand-displacement DNA synthesis is one of the key rate-limiting steps, amplification success is highly dependent on the length of the DNA target, with optimal amplicon length ranging between 130 to 200 base pairs (bp). This is critical when working with fragmented or degraded RNA, such as that obtained from nanobiopsies. Once amplified, success of DNA amplification can be determined either using agarose gel electrophoresis, by turbidity analysis, which measures the accumulation of magnesium pyrophosphate, or fluorometrically, by inclusion of DNA-intercalating dyes such as SYBR-Green.

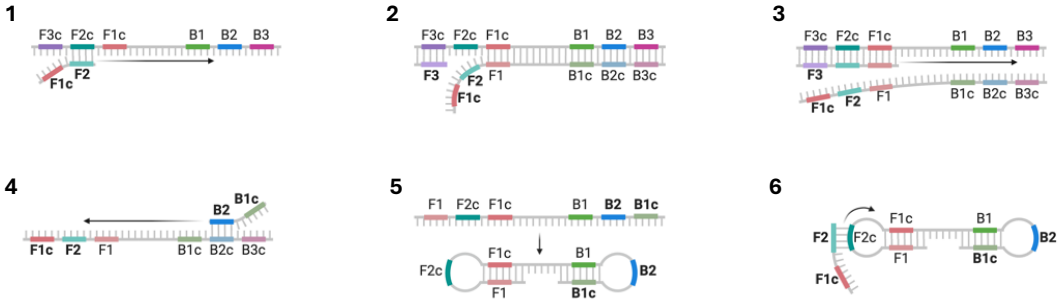
While LAMP offers considerable advantages in speed, sensitivity and equipment simplicity, its use is limited by interference of non-specific amplification products, technical challenge of primer design and its variable performance with structurally complex or low-input RNA targets. As such, its application to nanobiopsy-extracted RNA remains promising but demands further refinement and was not explored in the context of this work.

4.2.2.2 Nucleic Acid Sequence-Based Amplification (NASBA)

Nucleic Acid Sequence Based Amplification (NASBA) is an isothermal technique that can amplify RNA, from concentration as low as 10^6 copies/mL, by a factor of 10^8 using three different enzymes: avian myeloblastosis virus RT (AMV-RT), RNase H and T7 DNA-dependent RNA polymerase (DdRp).[157,158] Cycling through these enzymes at a fixed temperature of 41 °C, this technique amplifies RNA in two phases - a linear phase and an exponential phase (figure 4.1 B). Importantly, given the use of T7 DdRp, which requires the presence of a promoter sequence to initiate transcription, the forward primer must incorporate a T7 promoter sequence at its 5' end (5'-TAATACGACTCACTATAG-3').[159] While this introduces an additional layer of complexity to primer design, it is significantly more trivial than that of LAMP primer design. Furthermore, primers must be designed such that the maximum amplicon length does not exceed 250 nucleotides (nt). While this provides slight additional flexibility over LAMP, it is still limiting in applications requiring longer or full-length transcript amplification and detection. Mechanistically, input RNA is converted to cDNA by AMV-RT to generate an RNA-cDNA hybrid, from which the RNA is destroyed by RNase H. The reverse primer then binds to the cDNA, and dsDNA, containing a T7 promoter sequence, is produced by AMV-RT. The T7 promoter sequence is subsequently recognised by the T7 DdRp, initiating

transcription and resulting in the production of new RNA complementary to the target RNA. Once transcription has been initiated, the amplification phase begins. As in the linear phase, the reverse primer binds to the newly synthesised RNA, which is then extended by AMV-RT to generate an RNA-cDNA hybrid. RNA from the hybrid is destroyed by RNase H and the forward primer binds to the remaining cDNA strand. AMV-RT then extends the cDNA

A *Loop-mediated isothermal amplification (LAMP)*



B *Nucleic acid sequence-based amplification (NASBA)*

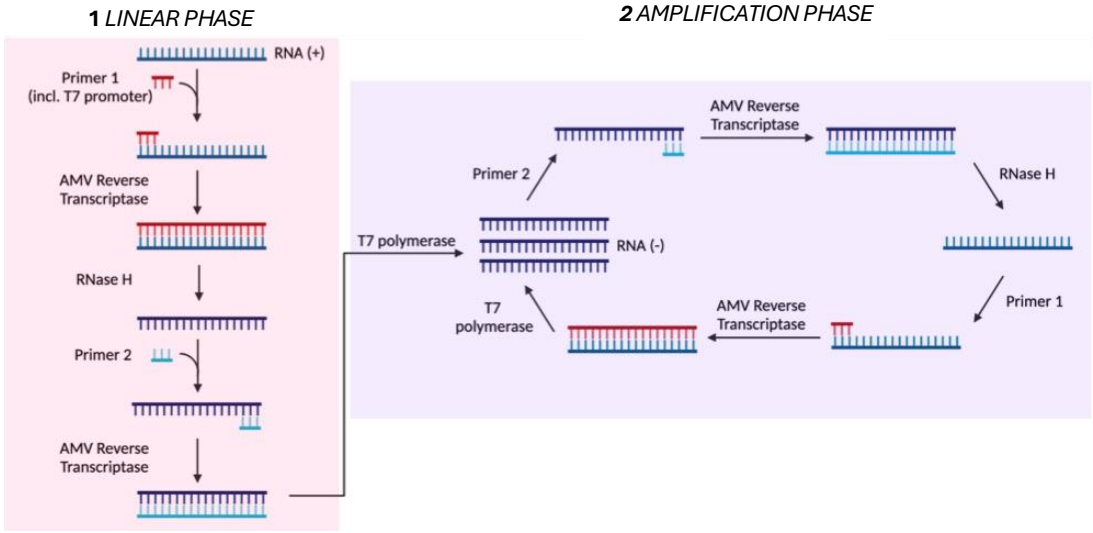


Figure 4.1 ***Overview of LAMP and NASBA techniques. A*** *LAMP: 1. The forward inner primer (F1c-F2) binds to the complementary binding region on the target sequence, F2c, initiating polymerisation, 2. enabling F3 to bind to its complementary sequence, F3c. 3. F3 displaces the new F1c-F2 strand, initiating another polymerisation cycle. 4. A second primer (B2-B1c) binds to the displaced strand, and steps 1-3 are repeated (not shown). 5. Given the complementarity between regions of close proximity on the resulting B2-B1c strand, the amplicon adopts a dumbbell shape. 6. Another F1c-F2 primer binds to the dumbbell structure and the process repeats to generate millions of copies of the starting sequence as stem-loop structures. Figure produced using the BioRender Loop-Mediated Isothermal Amplification (LAMP) template.* ***B*** *NASBA occurs in two phases: 1. a linear phase and 2. an amplification phase. 1. A primer, containing a T7 promoter sequence, binds to the RNA target. An RNA-DNA hybrid is created as RNA is converted to cDNA by AMV Reverse Transcriptase (RT). The RNA strand is degraded by RNase H, and primer 2 binds to the cDNA. The primer is extended and double-stranded (ds) DNA is produced by AMV-RT. 2. T7 polymerase recognises the T7 promoter sequence on the dsDNA, resulting in the initiation of transcription. Primer 2 binds to the newly synthesised RNA, an RNA-cDNA hybrid is created by AMV-RT and the RNA strand is degraded by RNase H. The process repeats to generate millions of copies of linear RNA.*

further using the forward primer as a template to generate more dsDNA containing the T7 promoter sequence. This is again recognised by the T7 DdRp, resulting in the production of more RNA and initiating another amplification cycle. Notably, as NASBA generates linear, single-stranded RNA products and avoids the complex secondary structures seen in LAMP products, downstream detection, quantification and further processing is facilitated,

rendering NASBA a more suitable technique for application to nanobiopsies. Amplification can be monitored in real-time using MBs. MBs are single-stranded nucleic acid sequences, which adopt a stem-loop structure. A sequence complementary to the target is sandwiched by complementary linkers (which bind to form the stem of the stem-loop) with both termini labelled – a fluorophore at the 5' end and a universal quencher at the 3' end.[160] In the presence of the complementary target, the MB undergoes a conformational change, causing hybridisation and fluorescence unquenching. As more target is generated throughout the cyclic amplification stage, more MB binds and more fluorophores are unquenched, resulting in a fluorescence signal that is directly proportional to the target concentration.

A compelling demonstration of NASBA's capabilities came about at the height of the COVID-19 pandemic, when diagnostic testing demand was at its peak with Wu *et al.* publishing a study exploring the use of NASBA for detecting COVID-19 RNA from saliva samples.[153] Herein, they compared the performance of a commercially available NASBA reagent kit (NWK-1, AMS-BIO) with an in-house kit. By varying the buffer components and pH, enzyme concentrations and primer concentrations, they were able to significantly greater RNA amplification using their in-house kit (up to 1700 times greater when comparing fold change of template amplification versus no template). All other parameters were kept same, with the same primer combinations, incubation times and temperature, and inclusion of an enzyme denaturing step for both techniques. This contributed to the development of a barcoded NASBA technique that allowed for fluorescence-based readout within one to two hours, with a limit of detection of 46.6 copies of the S-gene of SARS-CoV-2 per 20 µL reaction. Notably, this could be achieved directly from unprocessed saliva samples, eliminating the need for RNA pre-extraction - a significant advantage when working with COVID-19 samples, and low abundance transcripts, as sample loss during processing is minimised. Furthermore, incorporation of unique, sample-specific barcodes facilitates downstream sequencing, enhancing accuracy and scalability. Combined, this gave rise to a novel variation of NASBA - isothermal NASBA (nucleic acid sequence–based amplification) sequencing–based high throughput test (INSIGHT) – enabling optimised disease detection in both point-of-care and centralised settings.

Given the success in amplification exhibited by this study, this protocol was implemented for the amplification of MCF-7 RNA. While the maximum amplicon length of 250 nt does act as a limitation, this is slightly greater than the upper limit for LAMP. While detection methods differ for both LAMP and NASBA, given that the aim of this technique is to apply it to nanobiopsy extracts, which are currently analysed using RT-qPCR, NASBA-amplified samples would ideally need to be processed in the same way during optimisation. As NASBA generates linear amplicons as compared to the stem-loop structures generated using LAMP, qPCR amplification and subsequent detection of NASBA-amplified samples is likely to be more consistent and reliable.

4.3 Materials & Methods

4.3.1 RNA Extraction

MCF-7 cells were cultured as outlined previously in section **2.3.1.1** until ready for RNA extraction. RNA was extracted using the Chloroform/TRIzol method and purified using the RNEasy Mini Kit (Qiagen), as in section **2.3.1.3**.

4.3.2 Primer Design

Using the respective RNA FASTA accession numbers, three sets of specific primers were designed for each of the genes of interest – *ESR1*, *TFF1*, *GREB1* – and three housekeeping genes – *GAPDH*, β-*ACTIN* and *28S*, using

the NCBI Primer designing tool.[161] The NCBI primer designing tool designs primers using Primer3 and BLAST, while enabling fine tailoring of primer design parameters including PCR product size, primer melting temperature and exon/intron selection. Designed primers are carefully screened using BLAST to verify primers are unique and will not form primer pairs, in all possible binding combinations. Primer sequences were designed such that the resulting amplicons were **(i)** intron-spanning, **(ii)** between 100-600 nt in length (across exons) to minimise the risk of secondary structure formation, which can inhibit cDNA synthesis and qPCR primer binding, and **(iii)** contained the qPCR amplicon sequences.[162] Given that the NASBA cocktail contains a T7 RNA polymerase, all forward primers were designed with a T7 promoter sequence at the 5' terminus (5'-TAATACGACTCACTATAG -3'; T7) to allow for enzyme recognition. Primer sequences and corresponding amplicon lengths are summarised in table 4.1. Primer combinations were varied throughout this work, and will be referred to as a set of numbers, with the first number corresponding to the set from which the forward primer was taken and the second number corresponding to the set from which the reverse primer was taken, e.g. *GAPDH* 1.2, with forward primer taken from set 1 and reverse primer taken from set 2. Only one reverse primer sequence covering the qPCR amplicon region was generated for *28S* and was thus used as the reverse primer pair for all primer combinations. All primers were purchased from IDT-DNA Technologies.

4.3.3 Molecular Beacon Design

Similarly, a MB, complementary to a region within the *TFF1* gene, including a region contained within the NASBA amplicon, was designed according to the primer design rules outlined in table 2B in reference.[159] The MB was designed to have a GC content of 50-60 % and a loop sequence complementary to the target RNA flanked by five-base GC-rich sequences complementary to each other. The MB was modified at its 5' end with a 6-Carboxyfluorescein (6-FAM) molecule and a 4-((4-(dimethylamino)phenyl)azo)benzoic acid (DABCYL), a fluorescence quencher, at its 3' end. The MB was purchased in its HPLC-purified, dry form from Sigma-Aldrich.

5'-(6-FAM)- CCAAC AGGTGATCTGCG GTTGG-DABCYL-3'

Table 4.2 ***Primer sequences for all genes subjected to RNA Amplification.*** *All forward primers contained a T7 promoter sequence at the 5' terminus (sequence shown in section* ***4.2.2****), highlighted as T7 in the sequences.*

Gene	Set	Forward Primer Sequence	Reverse Primer Sequence	Amplicon Length
GAPDH	1	5'-(T7)GGGGAAGTCAGGTGGAGC-3'	5'-GAGGGATCTCGCTCCTGGAA-3'	292
	2	5' – (T7)GAAGTCAGGTGGAGCGAGG - 3'	5'–TCGCTCCTGGAAGATGGTGA-3'	282
	3	5'-(T7)TAGCTGGCCCGATTTCTCCT - 3'	5'-GACTCCACGACGTACTCAGC-3'	315
β-ACTIN	1	5'-(T7)GGCTGTGCTATCCCTGTACG-3'	5'-TTCTGCATCCTGTCGGCAAT-3'	533
	2	5'-(T7)ATCCAGGCTGTGCTATCCCT-3'	5'-CGCTCAGGAGGAGCAATGAT-3'	598
	3	5'-(T7)CACTGGCATCGTGATGGACT-3'	5'-TTCATTGTGCTGGGTGCCA-3'	533
28S	1	5'-(T7)TTCATGGACGACACGAGCC-3'	5'-AGCGAGCAGCCAAGCTC-3'	126
	2	5'-(T7)GTGCGCGTGGAATTCATGG-3'	-	137
	3	5'-(T7)CGCGTGGAATTCATGGACG-3'	-	134
ESR1	1	5'-(T7)CACTCAACAGCGTGTCTCCG-3'	5'-GCACAGTAGCGAGTCCTC-3'	187
	2	5'-(T7)GTGCGACGTTGTGTGTCATC-3'	5'-GGGCTCTGAGCACTCATCTC-3'	183
	3	5'-(T7)GAAAGGTGCGACGTTGTGTG-3'	5'-CGTTCCAGGCCTCCCTTTAG-3'	238
TFF1	1	5'-(T7)GGGGTCGCCTTTGGAGC - 3'	5'-GTCTAAAATTCACACTCCTCTTCTG-3'	202
	2	5'-(T7)GCCTTTGGAGCAGAGAGGAG-3'	5'-TGGAGGGACGTCGATGGTAT-3'	257
	3	5'-(T7)TGGAGGGACGTCGATGGTAT-3'	5'-GGGACGTCGATGGTATTAGGA-3'	269

GREB1	1	5'- (T7)GTGGGAGCTAGCTTTTTG - 3'	5'-TCTTCAGAAAACTCGCCGCT-3'	212
	2	5'-(T7)AGGTGCCCTACTACCTGGAG-3'	5'-ACACTGCACAGTAGCGAGTC-3'	209
	3	5'-(T7)TGATGCTACTGCACCCGC-3'	5'-CACTGCACAGTAGCGAGTCT-3'	247

4.3.4 RNA Amplification

RNA was serially diluted to yield 1 µL aliquots containing 10, 1 and 0.1 pg of starting RNA. Throughout the experiment, samples, and thus starting RNA amounts, were diluted 100 times. RNA was amplified using two different kits: a commercial kit (AMS-BIO) and a reagent mix adapted from the INSIGHT protocol by Wu *et al.* [153] All incubations were conducted on a ProFlex™ PCR System (Applied Biosystems), unless a MB was used, in which case a QuantStudio 3 Real-Time PCR System (Applied Biosystems) was used for real-time fluorescence detection.

4.3.4.1 AMS-BIO Kit

In the first instance, experiments were conducted using a pre-mixed NASBA wet kit (NWK-1, AMS-BIO) with three different combinations of MB – 1 µL 2 µM MB (1x MB), 2 µL 2µM MB (2x MB) or without MB (xMB) - using primer sets 1.1 for *GAPDH*, *ESR1*, *TFF1* and *GREB1*. Each gene and MB concentration was processed as a separate reaction.

1 µL diluted RNA, 2 µL 5 µM primer mix and either (i) 1 µL MB and 1 µL water (1xMB), (ii) 2 µL MB (2x MB) or (iii) 2 µL water (x MB) were added to a mix containing 6.7 µL reaction buffer (NECB-24, containing TrisHCl (pH 8 @ 25 °C), $MgCl_2$, KCl, DTT, DMSO)) and 3.3 µL nucleotide mix (NECN-24, containing dATP, dCTP, dGTP, dTTP, ATP, CTP, GTP, UTP, ITP) that had already been incubated at 41 °C for 5 minutes. Primers were annealed by heating at 65 °C for 2 minutes and cooling at 41 °C for 10 minutes. 5 µL enzyme mix (NECN 1-24, containing AMV RT, RNase H, T7 RNAP, BSA (0.48 mg/mL), high MW sugar matrix.) was then added to each RNA-primer-buffer mix and samples incubated for a further 45 minutes at 41 °C. Experiments were conducted both with and without a final enzyme denaturation step (ED; 95 °C, 2 minutes), to determine any cross-reactivity between NASBA enzymes and RT used for cDNA synthesis. Samples were made up to a total volume of 100 µL and processed *via* RT-qPCR immediately.

4.3.4.2 INSIGHT Reagent Mix

Given the success in amplification exhibited by Wu *et al.* using their in-house NASBA kit, this protocol was implemented for the amplification of MCF-7 RNA.[153] Reagent compositions, including variations in concentrations where used, are outlined in table 4.3. 1 µL RNA, 1 µL 0.5 µM primer mix and 4 µL nucleotide mix were added to a PCR tube containing 4 µL water and 5 µL buffer with DMSO. The sample was incubated at 65 °C for 2 minutes and 41 °C for 10 minutes before adding 5 µL enzyme mix and incubating for a further 90 minutes at 41 °C. As in section **4.3.4.1**, the experiment was initially conducted both with and without an enzyme denaturation step. Experiments conducted with the INSIGHT reagent mix also resulted in variable, unreliable amplification (as discussed in section **4.4.3.1**),

Table 4.3 Composition of reagent and enzyme mixes used for RNA Amplification using the INSIGHT method. *Variations in enzyme concentrations are specified.*

		Concentration			
Component	**Reagent**	**Stock**	**Final**	**Volume**	**Variations**
Buffer with DMSO	Tris HCl (pH 8.4, Invitrogen)	1M	40 mM	120 µL	
	$MgCl_2$ (Invitrogen)	1M	13.2 mM	39.6 µL	

	KCl (Invitrogen)	2M	75 mM	112.5 µL	
	DTT (Invitrogen)	1M	10 mM	30 µL	
	DMSO (Sigma-Aldrich)	100%	11%	450 µL	
	Nuclease-free water (NEB)	247.9 µL		247.9 µL	
	Total volume	*1000 µL*		1000 µL	
Buffer without DMSO	Tris HCl (pH 8.4, Invitrogen)	1M	40 mM	120 µL	
	$MgCl_2$ (Invitrogen)	1M	13.2 mM	39.6 µL	
	KCl (Invitrogen)	2M	75 mM	112.5 µL	
	DTT (Invitrogen)	1M	10 mM	30 µL	
	Nuclease-free water (NEB)	697.9 µL		697.9 µL	
	Total volume	*1000 µL*			
Nucleotide Mix	dNTP mix (NEB)	10 mM each	1 mM each	2 µL	
	NTP (NEB)	25 mM each	4 mM each	3 µL each	
Enzyme Mix	Diluted RNase H	5000 U/mL	3.75 U/mL	0.17 µL	
	Protoscript II (NEB)	200 000 U/mL	2500 U/mL	0.28 µL	1250 U/mL; 5000 U/mL
	T7 polymerase (NEB)	50 000 U/mL	6250 U/mL	2.75 µL	3125 U/mL
	BSA (NEB)	20 mg/mL	0.12 mg/mL	0.13 µL	
	Buffer without DMSO	1.78 µL		1.78 µL	
	Nuclease-free water (NEB)	0.4 µL		0.4 µL	
	Total volume	*5.5 µL*		5.5 µL	
Diluted RNase H	RNase H (NEB)	5000 U/mL	500 U/mL	5 µL	250 U/mL
	BSA (NEB)	0.48 mg/mL	20 mg/mL	1.2 µL	
	Buffer without DMSO	-	-	16.67 µL	
	Nuclease-free water (NEB)	-	-	27.13 µL	
	Total volume	-	-	50 µL	

and as such, enzyme and primer concentrations, and incubation durations were varied.

Alternative concentrations tested include: primers (2.5 µM, 5 µM), protoscript II RT (1250 U/mL), T7 polymerase (2250 U/mL), and samples were incubated for 30, 45, 60, 90 and 120 minutes. All optimisation experiments conducted with these varied concentrations were done only with optimised *β-ACTIN* primers, as determined in section **4.4.2.2**. A summary of all variations in enzyme concentrations, relative to the original INSIGHT reagent mix, is shown in figure 4.2. Experiments 1-8 were conducted with a 10 µM primer concentration, experiments 9-20 with a 5 µM concentration, experiments 21-24 with a 2 µM concentration and experiments 26-28 using a 1 µM concentration.

4.3.4.3 cDNA Synthesis and qPCR

4.3.4.3 (a) Primer Validation

Primers were validated in all possible combinations of forward and reverse primers, to determine optimal amplification efficiencies. RNA was serially diluted to cover a range of concentrations between 1 µg and 0.1 pg. All dilutions were RTed using LunaScript® RT as described in section **2.3.4.2 (a)** and qPCR amplified using Luna® Universal qPCR Master Mix as in section **2.3.4.1 (b)**. Primer sequences used are outlined in table 4.2. Correlation coefficients (R^2) were calculated for all primer combinations, using the RSQ function in excel.

4.3.4.3 (b) Processing NASBA samples

Once the RNA samples had been amplified, all samples were subjected to cDNA synthesis and qPCR using LunaScript® RT and Luna® Universal qPCR Master Mix as outlined in sections **2.3.4.2 (a)** and **(b)** to determine success of amplification, if at all. All samples were processed alongside a cDNA standard curve, with RNA amounts ranging from 1 µg to 10 pg. qPCR primers used for *GAPDH*,

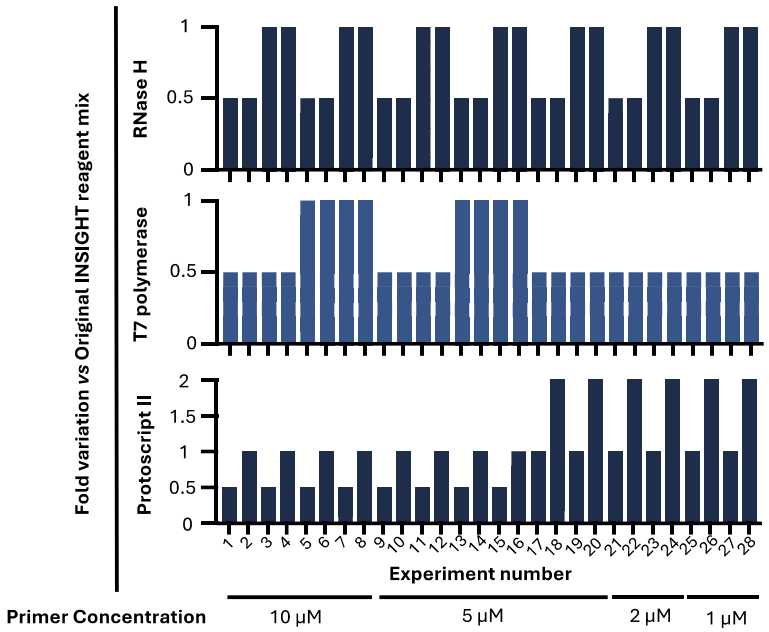


Figure 4.2 ***Fold variation in enzyme concentrations as compared to composition of original reagent mix, as optimized by Wu et al.[153]*** *28 different combinations of enzyme mixes were used to NASBA amplify β-ACTIN from 10 pg, 1 pg and 0.1 pg of starting RNA. Primer concentrations were also varied across reactions, ranging from 1 µM to 10 µM.*

Table 4.4 ***Forward and reverse qPCR primer sequences for the six genes amplified in this chapter.*** *Expected amplicon sizes are specified.*

Gene	Forward Primer Sequence	Reverse Primer Sequence	Amplicon Length
GAPDH	5'-GGTCACCAGGGCTGCTTTTA-3'	5'-TTCCCGTTCTCAGCCTTGAC-3'	237
β-ACTIN	5'-GGACTTCGAGCAAGAGATGG-3'	5'-AGCACTGTGTTGGCGTACAG-3'	378
28S	5'-CGATCCATCATCCGCAATG-3'	5'-AGCCAAGCTCAGCGCAAC-3'	100
ESR1	5'- CAGGTGCCCTACTACCTGGA-3'	5'- ACTGGCCAATCTTTCTCTGC-3'	123
TFF1	5'- CACCATGGAGAACAAGGTGA-3'	5'- TGACACCAGGAAAACCACAA-3'	134
GREB1	5'- ATCAGCTGCTCGGACTTGCTG-3'	5'- TGAGCTCCGGTCCTGACAGATG-3'	134

ESR1, *TFF1* and *GREB1* are as outlined in table 2.4, but are summarised again, with qPCR primer sequences for β-*ACTIN* and *28S*, with amplicon size in table 4.4.

4.3.4.4 Validating Amplification

In order to troubleshoot the poor amplification obtained when varying enzyme combinations and incubation times, DNA fragments sizes and concentrations were determined using a Tapestation 2200 (Agilent), an automated electrophoresis platform. Amplified RNA was RTed, and resulting cDNA was run on a D1000 high-sensitivity tape. 2 µL cDNA was mixed with 2 µL D1000 sample buffer, mixed briefly by pipetting and vortexed before loading into the instrument. Concentrations of all peaks were determined and peaks were classified as either primer carryover, or amplified RNA.

4.4 Results & Discussion

4.4.1 Amplification using the AMS-BIO Kit

In the first instance, three different experiments were conducted using the AMS-BIO kit according to the manufacturer's instructions, with each experiment varying in the amount of MB used. For all experiments conducted using this reagent mix, only the first primer set (i.e. 1.1) was used for *GAPDH*, *ESR1*, *TFF1* and *GREB1* - *β-ACTIN* and *28S* were not studied in this section and all genes and experiments were processed in parallel. The MB designed for this experiment was complementary to a region of the *TFF1* gene. In the presence of the complementary sequence, the MB binds to the template RNA, causing the MB to unloop and the fluorophore to unquench. As the RNA is amplified and concentration of template RNA increases, more MB binding and fluorophore unquenching should occur, resulting in a stronger signal. Given that *TFF1* is known to be an E2-responsive gene and NASBA validation experiments were conducted on RNA extracted from E2-treated cells, regardless of amplification, we expect that there should be some binding and subsequent fluorescence detection. However, given the low input concentrations – 10 pg, 1 pg and 0.1 pg RNA - this fluorescence may be outside of the limit of detection of the qPCR machine. Despite this gene-specific complementarity, the MB was used in the reaction mixes for all genes, but no fluorescence was expected for any samples other than *TFF1*. Interestingly, the addition of the recommended concentration of MB (1x MB) resulted in no fluorescence detection, and as such, the experiment was repeated with double the suggested concentration (2x MB). Unfortunately, this did not remedy the lack of fluorescence. This could simply be due to no or insufficient RNA amplification rendering the amplified RNA below the detection threshold, fluorescence being undetectable, the MB was unable to bind to the sequence of interest potentially due to secondary structure formation or poor MB design. Nonetheless, given that no fluorescence detection does not necessarily indicate a lack of RNA amplification, the samples were taken further.

The 'amplified' samples were RTed in two ways, **(i)** taken directly from the NASBA reaction sample, and **(ii)** after an enzyme denaturation step at 95 °C for 2 minutes, to minimise interference between the NASBA enzyme mix and the RT. This was done alongside a range of serially diluted MCF-7 RNA (1 µg – 10 pg) that was RTed without any amplification, in order to determine how much RNA amplification had occurred, if any. Samples were then processed using qPCR and melt curves of all samples which yielded any Ct values corresponding to positive amplification were analysed to determine true amplification. While Ct values were obtained for a number of the remaining 'amplified' samples during qPCR amplification, individual melt curves revealed that these were in fact false positives – exhibited by a large left shift (i.e. to lower temperatures) as compared to the RNA standard samples). This could potentially be explained by the formation of secondary structures resulting from the long amplicon length as well as the presence of excess primers and MBs, which may have been amplified in lieu of the target RNA. False positives were omitted from further analysis and have not been plotted. The fold-change for samples that seemed to have undergone some amplification are summarised in figure 4.3 A. As can be seen from this figure, amplification seemed to be very variable, ranging from 5-fold to 1000-fold increase, in only 11% of

total samples, while 9% exhibited RNA degradation and 2 samples showing no change. Given this variation and limited amplification in the majority of these cases, it can be assumed that amplification was unsuccessful. Interestingly, despite yielding a measurable fluorescence signal at low RNA concentrations in previous experiments (including RT and qPCR dye optimisation, section **2.4.4**), when similar concentrations were RTed for use as controls for the RNA amplification reaction, C_t values could not be determined for 1 ng, 0.1 ng and 10 pg samples for the *ESR1* and *GREB1* genes (figure 4.3 B). Conclusions could therefore not be drawn from the 'amplified' samples, specifically for the denatured *GREB1* sample, where, if

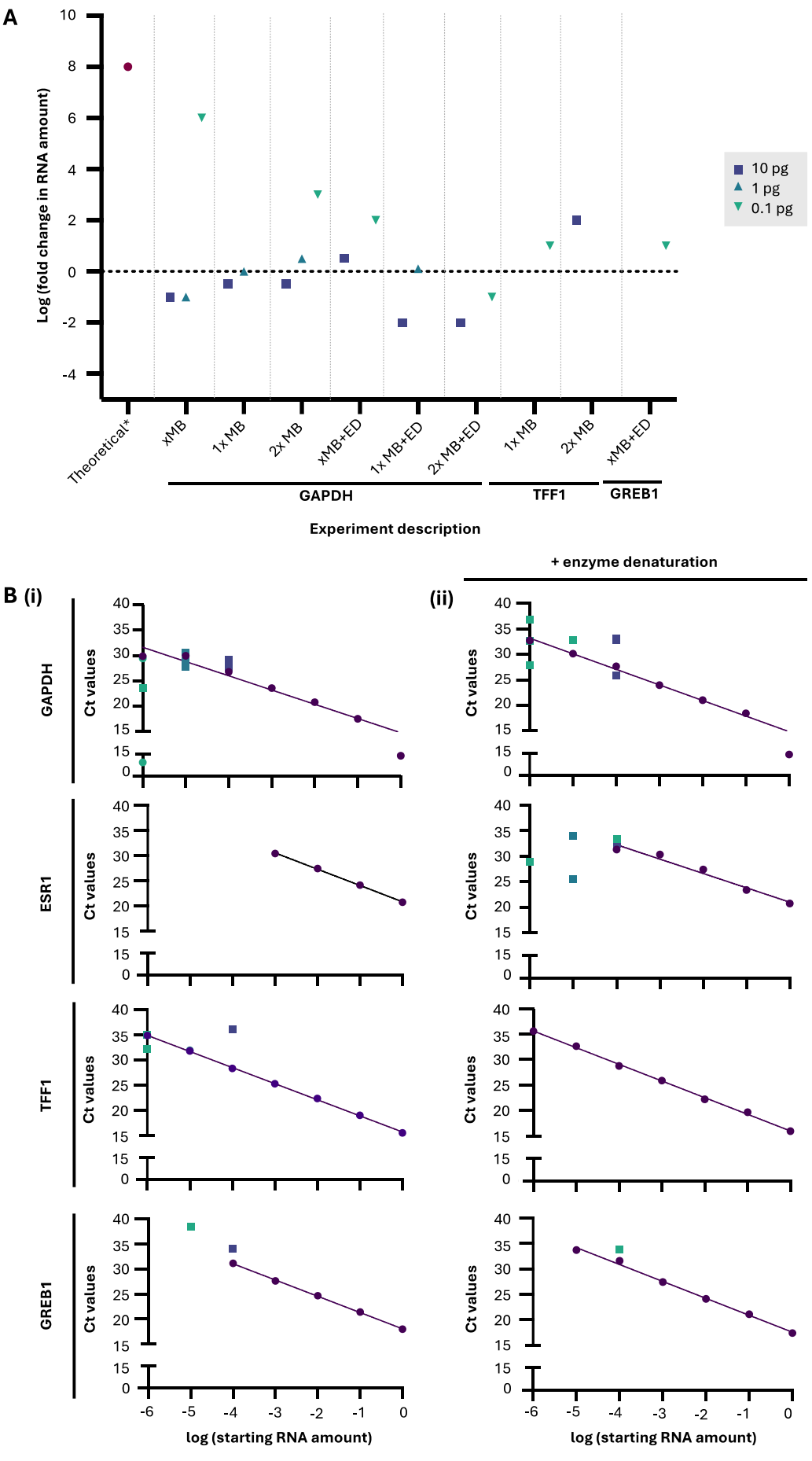

A
Log (fold change in RNA amount)
10
8
6
4
2
0
-2
-4
10 pg
1 pg
0.1 pg
Theoretical*
xMB
1x MB
2x MB
xMB+ED
1x MB+ED
2x MB+ED
1x MB
2x MB
xMB+ED
GAPDH
TFF1
GREB1
Experiment description
+ enzyme denaturation
B (i)
(ii)
GAPDH
ESR1
TFF1
GREB1
Ct values
log (starting RNA amount)

Figure 4.3 ***Scatter plots showing RNA amplification after NASBA using the AMS-BIO Wet Kit. A*** *Log of the fold change in starting RNA amount after NASBA using the AMS-BIO Wet Kit. *Theoretical amplification is shown in red (from K. H. Abd El Galil et al., 2005), while amplification of 10 pg (purple), 1 pg (blue) and 0.1 pg (green) of RNA are shown only for genes that could be detected using qPCR.[158] Experimental conditions are specified for each sample, including amount of molecular beacon (MB) used and whether an enzyme denaturation (ED) step was carried out.* ***B*** *Amplification plots showing Ct values of samples alongside standard curves for GAPDH, ESR1, TFF1 and GREB1. Reaction were either (i) taken directly for RT-qPCR processing or (ii) after an enzyme denaturation step.*

extrapolated, would seem like there could have been some amplification. Given that none of the *ESR1* and *TFF1* 'amplified' samples, without an enzyme denaturation step, yielded detectable C_t values with the desired melt curve characteristics, this did not prove to be a limitation during analysis. This could potentially be explained by the high enzymatic complexity of NASBA reactions, which poses a risk of uncontrolled side reactions, especially when working with very low RNA input. In such cases, even slight contamination, degradation or incomplete removal of reaction components can result in qPCR mispriming.

Overall, very poor amplification was observed with regard to the expected theoretical amplification, with the two lowest RNA concentrations – 1 pg and 0.1 pg – exhibiting the most amplification. The post-NASBA ED step seemed to result predominantly in RNA degradation, or if amplification occurred, was also significantly lower than the theoretical amplification. Therefore, all further experiments were conducted without an ED step, thus prioritising RNA integrity over reducing enzymatic interference. Furthermore, inclusion of MBs provided no additional value, as no fluorescence was detected in the *TFF1* samples during RNA amplification. The reliability of MB-based assays and detection is contingent on multiple variables, including perfect MB-template alignment, enzyme activity, template integrity and detection sensitivity, all requiring further optimisation.

4.4.2 Amplification using the INSIGHT Reagent Mix

Given the limited amount of reagent provided in the AMS-BIO kit, and the lack of flexibility to modify the constituents of the reagent mix, it was decided that a custom reagent mix would be used. Coincidentally, Wu *et al.* published a study wherein they developed their own NASBA kit, enabling high-throughput sequencing of COVID-19 samples, and comparing it to the AMS-BIO kit.[153] As small variations in reagent concentrations and components resulted in significantly enhanced amplification, and given the flexibility in modifying enzyme and buffer mix components and concentrations, this reagent mix was used for all further experiments. Key differences between the AMS-BIO kit and INSIGHT mixes are summarised in table 4.5.

Table 4.5 ***Key differences in NASBA reaction mix components between AMS-BIO Kit and INSIGHT Reagent Mix.***

	Reagent	**AMS-BIO NASBA Kit**	**INSIGHT Enzyme Mix**
Variation	Reverse Transcriptase	AMV	Protoscript II
Volume	RNase H	1x	3x
	T7 polymerase	1x	10x
Concentration	Primer	5 µM	25 nM

4.4.2.1 RNA Amplification

Experiments were conducted as in section **4.4.1**, now using the INSIGHT reaction mix, with the same primers and primer concentrations, starting RNA amounts and variations in MB concentrations. 'Amplified' RNA was RTed

to generate cDNA both straight from the reaction and after an ED step, before conducting qPCR alongside a standard curve. Again, despite the MB being complementary only to a region contained within the *TFF1* amplicon, MB was added to all samples, but no fluorescence was observed for any. This suggests that the MBs either **(i)** failed to bind due to the target due to lack of complementarity or inaccessibility of the target sequence, **(ii)** failed to unloop or fluoresce under NASBA conditions, or **(iii)** were incompatible with the INSIGHT mix's buffer or ionic environment. Therefore, to avoid any potential interference and complications from adding MB to the reaction mix, MB was omitted for all further experiments conducted as part of this chapter, with RNA amplification instead monitored through RT-qPCR. Ct values and melt curves were analysed for all 'amplified samples', and curves exhibiting non-specific amplification, or the presence of primer dimers were omitted from analysis. The log of the fold change of all samples, as compared to the starting RNA amounts, is summarised in figure 4.4 A, with individual standard curves and Ct values of amplified samples shown in figure 4.4 B. Similar to amplification using the AMS-BIO kit, there was very variable amplification across samples, ranging from no amplification for all *ESR1* samples, to a maximum of 1000-fold increase in two samples – 0.1 pg *TFF1* xMB and xMB+ED. Interestingly, an inverse relationship is observed between the amount of amplification and starting RNA amount for *TFF1* xMB, *TFF1* xMB+ED and *GREB1* xMB+ED, as with the AMS-BIO kit. For all three conditions, the amount of amplification observed is greatest with the lowest RNA amount, with the 10 pg sample for *GREB1* xMB+ED showing no amplification. This is also seen in *GAPDH* xMB, where, while only 10-fold, the greatest amplification is observed in the 0.1 pg sample. This could potentially be caused by inhibitory effects at higher RNA concentrations, wherein remaining impurities from RNA isolation could interfere with template amplification, or due to low-input reactions amplifying template by chance rather than specificity. 10 pg samples for *GAPDH* xMB+ED, *TFF1* 2xMB and *TFF1* 2xMB+ED exhibited only degradation and no Ct values were obtained for the two lower concentrations. This could be indicative of further degradation, pushing the Ct values of the samples outside of the detection limits of the qPCR machine.

Overall, use of the INSIGHT reagent mix did not seem to provide any benefit in terms of amount of RNA amplified, percentage of samples amplified, or total amount of amplification obtained compared to theoretical amplification using this NAAT (red dot, figure 4.4 A). This lack of amplification could be explained in three ways: **(i)** the primers being used for NASBA were not efficient, **(ii)** amplicons were too long, or **(iii)** the enzyme and buffer mixes were not optimal for this reaction. In order to avoid the generation of secondary structures, primer design rules recommend designing primers such that the amplicon length is no longer than 250 nt in length.[159] Except for *GAPDH*, which has a resulting amplicon length of 292 nt using the 1.1 primer combination, amplicons of all other genes using the 1.1 primer combination fall below this limit. As such, formation of secondary structures is unlikely to contribute to inefficient amplification in this case, and further optimisation was conducted on primers and reagent combinations. Furthermore, no significant difference or correlation was observed between samples processed with and without an ED step prior to RT-qPCR. Given the large proportion of degraded samples and samples with no Ct values post-ED using both the AMS-BIO and INSIGHT reaction mixes, the ED step was omitted from all further reactions. Results described in this section therefore prompted primer re-design and re-optimisation of reaction conditions, efforts of which are discussed later in this chapter.

4.4.2.2 Primer Optimisation

As minimal amplification was observed, primers were validated on cDNA, to ensure that they were not only binding to the template of interest but also providing efficient amplification. qPCR was conducted using all independent forward and reverse primers were tested in all possible combinations at four different concentrations (2.5 µM, 0.25 µM, 25 nM and 2.5 nM), on a single cDNA replicate (n = 1). Considering that a 0.5 µM primer

concentration yielded better amplification using INSIGHT kit as compared to the AMS-BIO kit and the extremely small amounts of starting RNA, it was thought that using a lower primer concentration could result in more effective, on-target amplification and subsequent detection.[153] The gene pool was also expanded to include *β-ACTIN* and *28S* as additional housekeeping genes, as potential housekeeping genes, due to the variable *GAPDH* expression as seen with NASBA using the AMS-BIO kit (section **4.3.1**) and in RNAScope (section **2.4.6.1**). However, these were only tested at 2.5 µM concentrations, in triplicate (n = 3). Primers were named in a way such that the first number in the 'primer pair' nomenclature corresponded to the identity of the forward

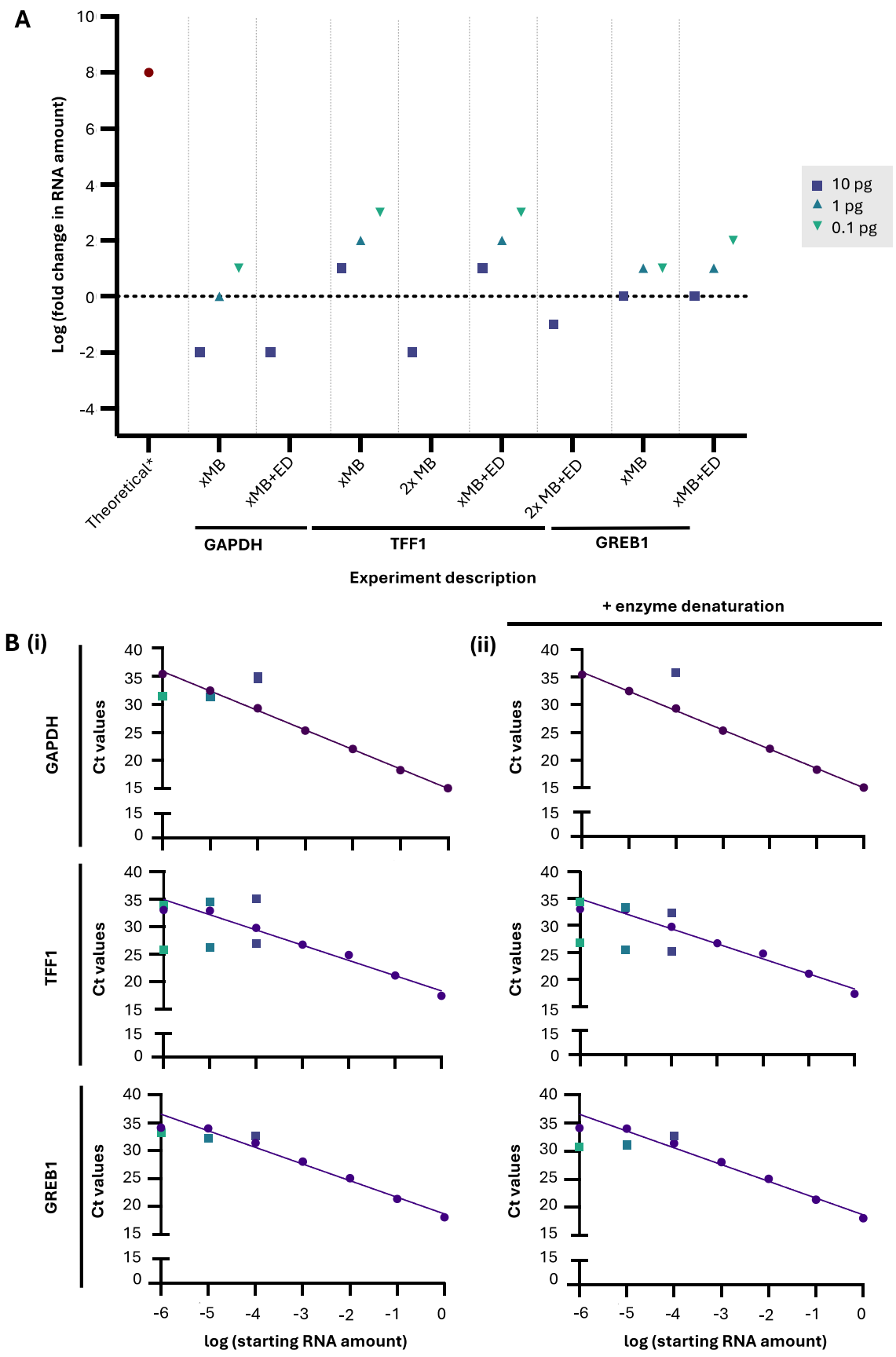

A
Log (fold change in RNA amount)
10
8
6
4
2
0
-2
-4
10 pg
1 pg
0.1 pg
Theoretical*
xMB
xMB+ED
xMB
2x MB
xMB+ED
2x MB+ED
xMB
xMB+ED
GAPDH
TFF1
GREB1
Experiment description
B (i)
+ enzyme denaturation
(ii)
GAPDH
TFF1
GREB1
Ct values
40
35
30
25
20
15
15
0
-6
-5
-4
-3
-2
-1
0
log (starting RNA amount)

***Figure 4.4** **Scatter plots showing RNA amplification after NASBA using the INSIGHT reagent mix. A** Log of the fold change as compared to starting RNA amount after conducting NASBA using the INSIGHT reagent mix. *Theoretical amplification is shown in red (from K. H. Abd El Galil et al., 2005), and amplification of 10 pg (purple), 1 pg (blue) and 0.1 pg (green) of RNA are shown only for genes that could be detected using qPCR.[158] Experimental conditions are specified for each sample, including amount of molecular beacon (MB) used and whether an enzyme denaturation (ED) step was carried out. **B** Amplification plots showing Ct values of samples alongside standard curves for GAPDH, ESR1, TFF1 and GREB1. Reactions were either (i) taken directly for RT-qPCR processing or (ii) after an enzyme denaturation step.*

primer, while the second number in the 'primer pair' nomenclature referred to the identity of the reverse primer. Primers were tested on starting cDNA amounts of 1 µg, 10 ng, 1 ng and 0.1 ng. These concentrations were chosen instead of 10 pg, 1 pg and 0.1 pg (as used for previous NASBA experiments), as, theoretically, if RNA amplification were to occur on these amounts, given this NAAT yields up to 10^8-fold amplification, we would be detecting within the cDNA range chosen. Furthermore, given the ongoing difficulties in detecting low RNA concentrations, there would be no benefit in testing for the lower concentrations, especially when determining primer efficiencies.

Of all the primer combinations and concentrations tested, very sparse data (few to no data points) was obtained for primer concentrations 0.25 µM, 25 nM and 2.5 nM. Only *GREB1* 3.3 yielded detectable fluorescence at a 25 nM concentration and no data was obtained for any of the primers when used at a 2.5 nM concentration (figure 4.5). No trend was observed in terms of RNA concentration and probability of obtaining detectable amplification, and in certain cases, no Ct values were obtained for the 1 µg sample either, suggestive of inefficient primer annealing or incompatible primer pairs. Unsurprisingly, a Ct value was only obtained for the highest cDNA concentration using the *ESR1* 1.1 primer combination, and that too at a high threshold of 35 cycles. This highlights the primer inefficiency and explains why there was no *ESR1* detection using this primer combination using both the AMS-BIO and INSIGHT reagent combinations.

From here, where Ct values were obtained, R^2 values and amplification efficiencies (E) were determined for the individual conditions, using the RSQ function in excel and **equation 4.1**, respectively (summarised in table 4.6):[163]

$$\mathrm{E} = 10^{\frac{-1}{\text{gradient of slope}}} - 1 \qquad \textbf{(4.1)}$$

Overall, most calculated R^2 values were above 0.94, suggesting good correlation, and almost linear relationship, between the log of the dilution factor and the Ct values obtained. However, certain samples exhibited significantly lower correlations, ranging between 0.64 for 2.5 µM *GAPDH* 2.2 and 0.66 for 0.25 µM *ESR1* 1.2 to 0.89 for 0.25 µM *GREB1* 3.2. Interestingly, despite having low correlation efficiency, amplification efficiencies for 0.25 µM *ESR1* 1.2 and *GREB1* 3.2 were 99.06 and 98.94,

***Table 4.6** **R^2 and amplification efficiencies for all primer combinations tested in this chapter, for GAPDH, ESR1, TFF1 and GREB1.** Primers were tested at concentrations of 2.5 µM, 0.25 µM, 25 nM and 2.5 nM on a single replicate (n = 1) over a range of cDNA concentrations. RNA was serially over a concentration range of 1 µg to 0.1 pg, from which cDNA was synthesised. For most cases, reliable data was obtained only for a primer concentration of 2.5 µM. Data is shown only where obtained. Where n = 1, only two data points were obtained. No data was obtained for blank cells.*

Gene			Primer Combination								
			1.1	**1.2**	**1.3**	**2.1**	**2.2**	**2.3**	**3.1**	**3.2**	**3.3**
GAPDH	2.5 µM	R^2	0.957	1	0.889	0.994	0.968			0.893	
		Efficiency	98.958	98.293	99.837	96.785	97.546			81.027	
	0.25 µM	R^2					0.6439				
		Efficiency					98.943				
ESR1	2.5 µM	R^2		0.945	0.973	0.964	1	1		0.964	1
		Efficiency		76.991	92.664	90.482	97.625	90.846		58.827	93.785
	0.25 µM	R^2		0.664	1		1			1	1
		Efficiency		99.06	55.521		61.366			49.167	59.164

TF E1	2.5 µM	R^2	0.998	0.941	0.975	0.939	0.597				
		Efficiency	99.957	65.526	65.888	80.817	58.704				
GREB1	2.5 µM	R^2	0.997				0.999		0.999		0.994
		Efficiency	44.157				55.016		67.909		55.781
	0.25 µM	R^2							0.997	0.887	0.992
		Efficiency							36.698	51.945	54.827
	25 nM	R^2									0.982
		Efficiency									50.111

respectively. This suggests that while there may not necessarily be a linear relationship between the dilution factor and Ct values (possibly due to poor dilution or inconsistent RNA input), primers seem to anneal and amplify the target relatively efficiently. While in certain cases, a strong correlation appears to exist ($R^2 = 1$; 0.25 µM: *ESR1* 1.3, *ESR1* 3.2, *ESR1* 3.3), these are not truly representative of a good fit given that **(i)** only 2 C_t values were obtained for the various RNA concentrations that were tested and **(ii)** only one replicate was conducted ($n = 1$). The primer combinations resulting in the best correlation from these $n = 1$ experiments for each gene, excluding those with $R^2 = 1$, were then taken further, regardless of the calculated primer efficiencies. This was especially important for the *GREB1* primers, as, despite yielding high R^2 values, the amplification efficiencies ranged from 44-67% - relatively far off from the desired 90-110% characteristic of efficient amplification.[164] However, as minimal amplification was obtained using all primer concentrations of 0.25 µM and 25 nM, all further experiments were conducted using a 2.5 µM primer concentration. Furthermore, as only three combinations of *28S* primer could be generated, two of which could not yield Ct values for all cDNA concentrations tested, all further experiments were conducted using β-*ACTIN* as the housekeeping gene. R^2 values for *GAPDH* ranged between 0.88 and 0.99. However, apart from primer set 2.2 at a concentration of 2.5 µM, even the highest cDNA concentration yielded Ct values around 35, which

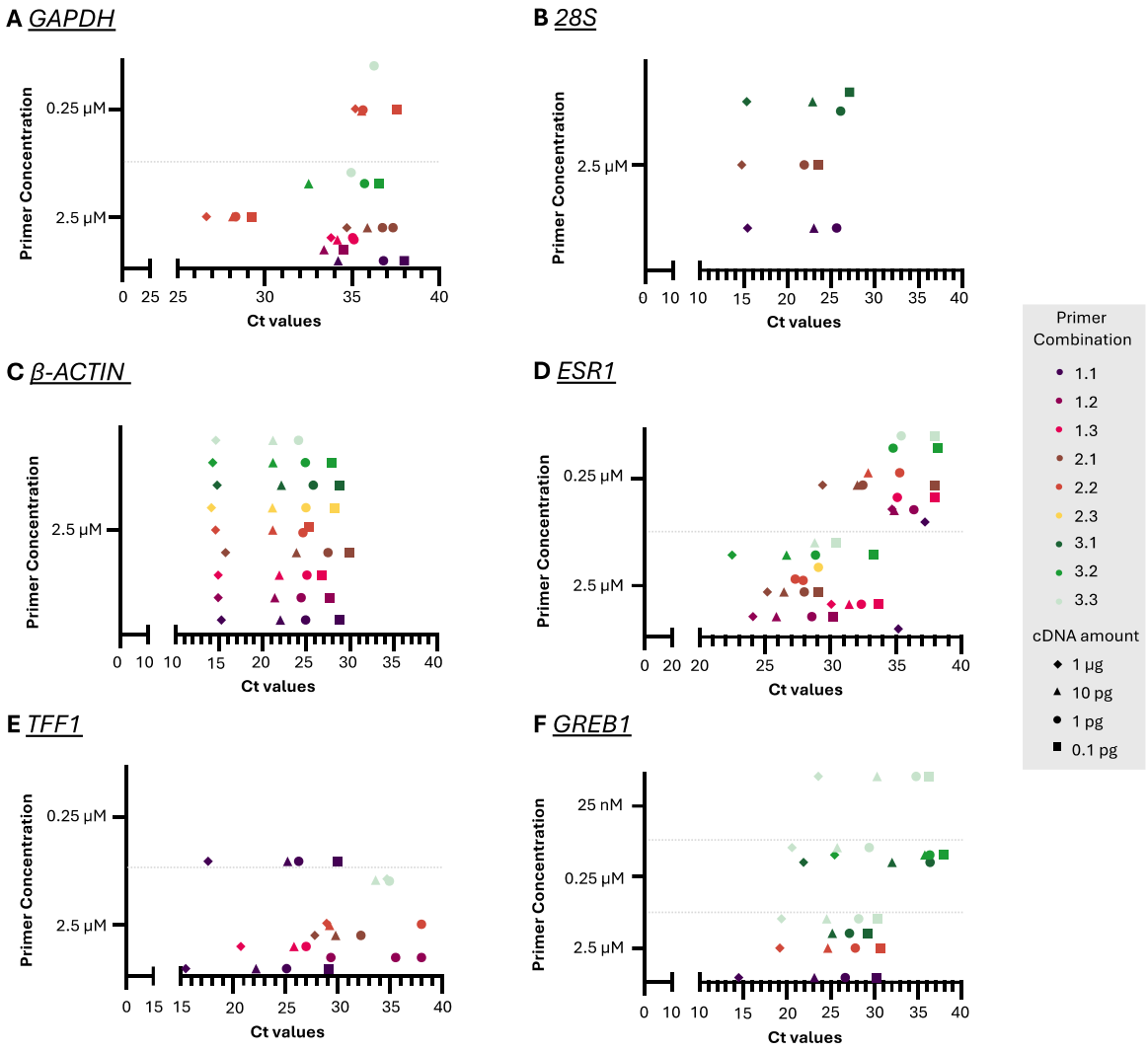


***Figure 4.5** Primer validation experiments, with all combinations of forward and reverse primers for A **GAPDH**, B **28S**, C **B-ACTIN**, D **ESR1**, E **TFF1** and F **GREB1**. Primers were tested at concentrations of 2.5 µM, 0.25 µM, 25 nM and 2.5 nM on four different cDNA concentrations (1 µg (diamond), 10 pg (triangle), 1 pg (circle) and 0.1 pg (square)). Ct values are shown only for samples that yielded amplification. Three housekeeper genes (A GAPDH, B 28S and C B-ACTIN) were tested, alongside the three genes of interest (D ESR1, E TFF1 and F GREB1).*

raises concerns on the reliability of detection when attempting to quantify amplification in much more dilute samples.

Experiments were subsequently repeated using the primers with the best correlation, this time on a larger range of serially diluted cDNA concentrations: 1 µg – 0.1 pg, with three replicates of each concentration (n = 3). R^2 and E values were calculated for all primer experiments, with results summarised in table 4.7. All R^2 values for *β-ACTIN* were found to be in very close proximity to perfect linear correlation (0.81 < R^2 < 0.999) and amplification efficiencies were all above 99%. The optimal primer combination for all genes was therefore chosen based on the total number of cDNA concentrations that yielded detectable Ct values, and the combination of highest R^2 and E values. The

***Table 4.7** R^2 and amplification efficiencies for primer combinations with highest R^2, as determined from table 4.5. Primers were tested only at a concentration of 2.5 µM on the same cDNA concentration range as in table 4.5, on three replicates (n = 3). Data is shown only where obtained. Primer combinations with cells with an 'x' in were not tested in triplicate. No data was obtained for blank cells.*

	Primer Combination								
Gene	**1.1**	**1.2**	**1.3**	**2.1**	**2.2**	**2.3**	**3.1**	**3.2**	**3.3**

β-ACTIN	R^2	0.982	0.949	0.956	0.813	0.949	0.999	0.924	0.934	0.999
	Efficiency	99.937	99.898	99.591	99.667	99.874	99.970	99.932	99.902	99.855
28S	R^2	0.9876	x	x	0.969	x	x	0.9971	x	x
	Efficiency	99.963	x	x	99.487	x	x	99.925	x	x
ESR1	R^2	x		0.969		x	0.909	x		x
	Efficiency	x		99.916		x	99.529	x		x
TFF1	R^2	0.9532	x	0.910	x	x	x	x	x	x
	Efficiency	98.048	x	98.959	x	x	x	x	x	x
GREB1	R^2	0.971	x	x	x	0.888	x	x	x	x
	Efficiency	99.586	x	x	x	98.768	x	x	x	x

final optimised primer panel comprised β-*ACTIN* 1.1, *ESR1* 2.3, *TFF1* 1.3 and *GREB1* 1.1.

4.4.2.3 Reagent Mix Optimisation

NASBA was repeated as in section **4.4.2.1**, on RNA concentrations of 10 pg, 1 pg and 0.1 pg, using the INSIGHT Reagent mix, with the optimised primers at a 2.5 µM concentration. However, no Ct values were obtained for any of the samples for all genes, indicating either failure in amplification or detection. As amplicons resulting from the new set of primers are all approximately 200 nt in length, the generation of secondary structures during amplification can be ruled out as a cause of poor qPCR detection. As such, further optimisation, involving variation of the enzyme concentrations within the reagent mix, was conducted using *β-ACTIN* 1.1. RNase H and T7 polymerase were tested at half their recommended concentrations, and Protoscript II at half and double the recommended concentration. Experiment outlines are shown in figure 4.2. Additionally, primer concentrations were varied and tested at 1 µM, 2 µM, 5 µM and 10 µM, to ensure that experiments were not limited by too low primer concentrations, despite having significantly lower starting RNA concentrations.

Overall, Ct values were obtained only for 60% of experimental conditions tested, with no observed correlation between primer or enzyme concentrations (figure 4.6 A). Where Ct values were obtained for 10 pg samples, concentrations were either unchanged (experiment 12) or RNA was degraded by up to 10,000 times (experiments 13, 15 and 21), consistent with all other experiments using 10 pg RNA conducted in this chapter. This could suggest that an RNA concentration of 10 pg is too large to be efficiently amplified. With greater RNA concentrations, there may potentially be increased interactions with degrading enzymes within the enzyme mix (RNase H), which could lead to over-degradation of RNA from the RNA-DNA hybrid. Alternatively, if the concentration of RNA is high

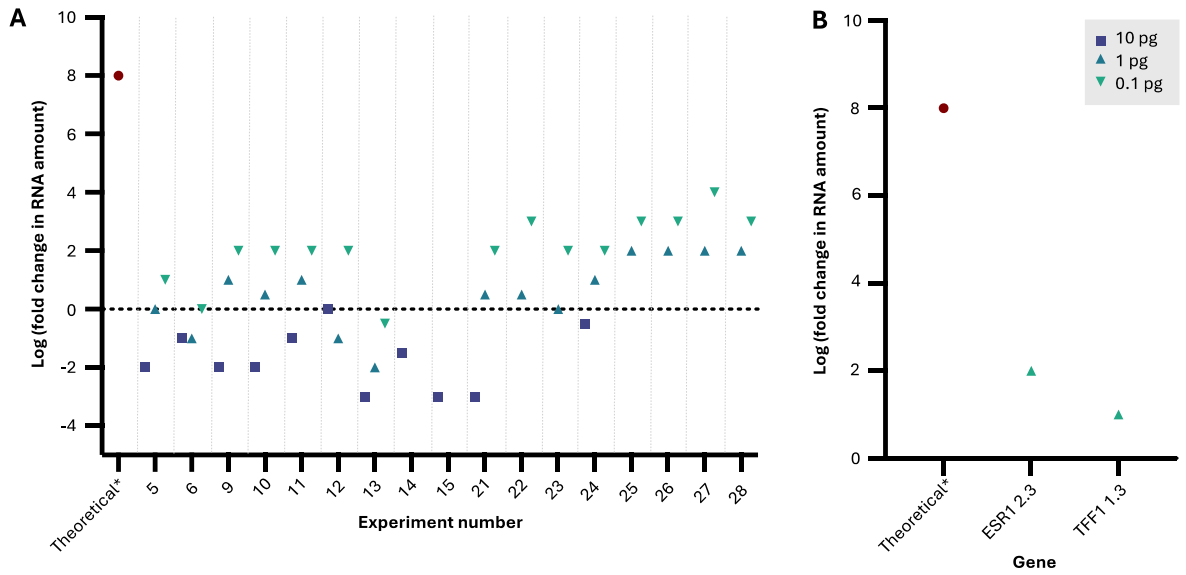


Figure 4.6 ***Scatter dot plots showing the fold change in RNA amounts with variations in INSIGHT enzyme concentrations. A*** *Log of the fold change as compared to starting RNA amount after conducting NASBA using variations of the INSIGHT reagent mix. Enzyme mix optimization was conducted using β-ACTIN 1.1 primer combination. Components of different experimental condtions and primer concentrations used for each experiment are outlined in figure 4.2.* ***B*** *Log of the fold change in RNA amounts for all genes of interest using experimental condition 27 and optimised primer combinations – β-ACTIN 1.1, ESR1 2.3, TFF1 1.3 and GREB1 1.1, using 0.1 pg RNA. Ct values were obtained only for ESR1 and TFF1. *Theoretical amplification is shown in red (from K. H. Abd El Galil et al., 2005), and amplification of 10 pg (purple), 1 pg (blue) and 0.1 pg (green) of RNA are shown only for experimental conditions and samples that could be detected using qPCR.*[158]*

enough and the reaction conditions permit, molecules may aggregate and generate secondary structures or become more susceptible to degradation, both of which decrease the amount of RNA accessible for primer annealing and subsequently compromise amplification efficiency.[165] Greater amplification was obtained for 1 pg samples, with most exhibiting up to 10^2-fold amplification. The concentration of one sample remained unchanged while three samples degraded, although to a lesser extent than with the 10 pg samples. The most amplification for the 0.1 pg samples was observed for experiment 27, with 10^4-fold amplification elicited by halving the T7 polymerase concentration and using a primer concentration of 1 µM. Amplification for the 1 pg sample using this experimental condition was also consistent, highlighting that a balance between enzyme and primer concentrations is indeed required for optimal performance at low inputs. However, again, none of the experimental conditions tested yielded amplification close to the theoretical 10^8-fold amplification that this technique can achieve, underscoring a large and persistent gap between expected and actual outcomes.

The gene pool was nonethless expanded to include the genes of interest, using the optimised primer sets – *ESR1* 2.3, *TFF1* 1.3 and *GREB1* 1.1 – using only 0.1 pg samples (figure 4.6 B). However, despite keeping the reaction conditions identical, no amplification was obtained for β-*ACTIN* 1.1. *TFF1* 1.3 only exhibited 10-fold amplification and *ESR1* 2.3 exhibited 100-fold amplification. Given this unreliability in amplification and subsequent detection, one last effort was made to understand whether reaction conditions could be improved, this time by varying NASBA incubation times.

4.4.2.4 Varying Incubation Times

In a final attempt to understand the poor amplification obtained using the INSIGHT mix on RNA extracted from MCF-7 cells, NASBA was again carried out using the original enzyme mix components, with 2.5 µM *β-ACTIN*

1.1, this time with varying incubation times. Reactions were set up as described in section **4.3.4.2**, using 0.1 pg RNA and incubated for 30, 45, 60, 90 or 120 minutes, in duplicate, alongside a water control for each timepoint. Amplified RNA was RTed and run on a D1000 screentape. Peaks were analysed and assigned one of two identities: **(i)** primer carryover (peaks under 100 bp), or **(ii)** an amplicon peak (between 500-600 bp – note that the length of the amplicon generated using *β-ACTIN* 1.1 is 533 bp) (figure 4.7). Ignoring the peaks associated with primer carryover, we see that only two samples generated amplicon peaks, both with incubation times of 60 minutes or greater. Overall, given that 0.5 µM primer is used, compared to only 0.1 pg RNA, such a high primer carryover percentage/concentration is expected. Incubating beyond 90 minutes resulted in more off-target amplification or primer dimer formation, with no peaks visible in the expected amplicon range of 500-600 nt, possibly due to RNA degradation associated with increased incubation at elevated temperatures, which RNA is extremely sensitive to. In general, the amplicons only constitute a very small proportion of overall concentration of the samples. This raises further doubt on the reliability of qPCR detection to determine NASBA efficiency and success. While it is possible to differentiate between true and off-target amplification, and primer dimer formation from qPCR melt-curves, peaks and Ct values could be biased toward off-target amplification, hiding valuable information. This high variability, low yield and inconsistent detection of the *β-ACTIN* 1.1 amplicon at different incubation times highlights a fundamental hurdle in NASBA reaction parameters and the detection method used. This not only casts doubt on the reliability of the fold change data obtained for experiments reported within this chapter, but also highlights the need for alternative detection strategies, including gel electrophoresis or sequencing-based to accurately confirm NASBA success when working with low RNA input. However, these will also be riddled with their own set of challenges, which warrant extensive validation.

4.5 Conclusion

This chapter aimed to explore use of NASBA-based RNA amplification using low-input RNA

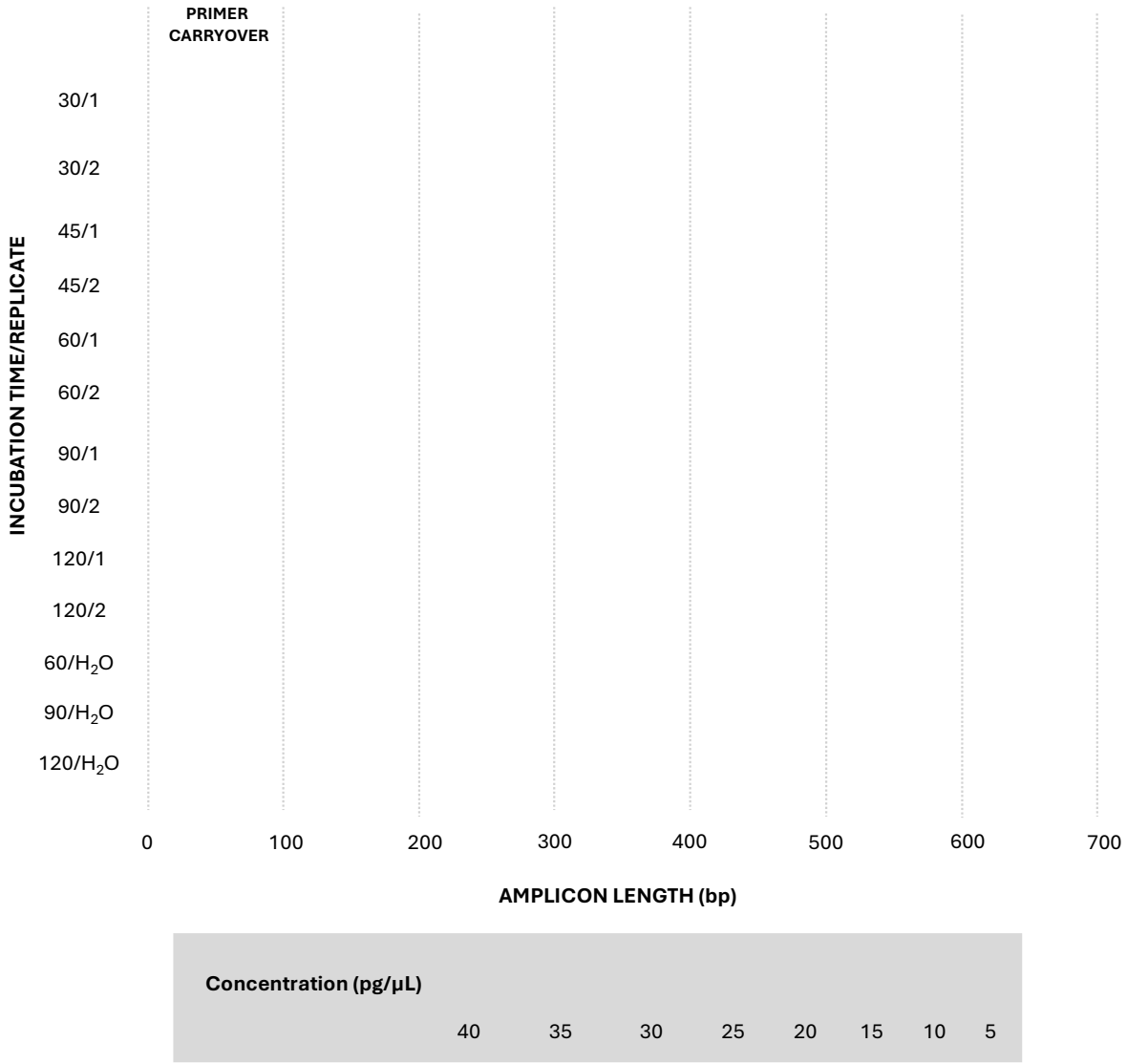


Figure 4.7 ***Dot plots showing concentration and amplicon length of cDNA after varying NASBA incubation times. RNA was subjected to NASBA using the INSIGHT reagent mix, with 2.5 µM β-ACTIN 1.1 primer.*** *Samples were isothermally incubated for 30, 45, 60, 90 and 120 minutes before being reverse transcribed. Amplicon sizes and concentrations were quantified using a D1000 high sensitivity screentape on a Tapestation 2200 (Agilent). Each circle corresponds to a peak centred around the amplicon length, with size corresponding to concentration (to scale). All resulting amplicons for all samples are shown Peaks within the desired amplicon length range are shown in red.*

extracted from MCF-7 cells, with the long-term aim of coupling with nanobiopsy extracts to facilitate gene expression analysis and improve sub-cellular gene detection. Despite validating primers, testing primers at various concentrations and varying enzyme concentrations and incubation times, RNA amplification was found to be extremely variable and unreliable, using two different established reagent mixes, with no clear indication of what could be resulting in the observed inconsistencies. Given the difficulty in obtaining reliable amplification, despite various attempts at optimising reaction conditions, the decision was made to not progress with this project. Furthermore, as qPCR amplification of nanobiopsy extracts already proved to be problematic (as outlined in chapter 2), even when using reagents of variable sensitivity, this was not attempted on nanotweezer extracts. The inability to quantify starting concentrations of nanobiopsy extracts would severely bias quantification of amplification, as no comparison could be made to determine fold amplification and subsequently quantify gene expression within cells. In order to develop this technique further, for ultimate use on nanobiopsy extracts, additional

troubleshooting is necessary at several stages of the experiment. This includes, but is not limited to, optimising reaction conditions further, such as enzyme concentrations, buffer constituents and incubation times. Further work is also required to develop an efficient primer panel and determining optimal working concentrations and enabling amplification of RNA concentrations greater than 1 pg, and less than 0.1 pg, as nanobiopsy extracts could contain variable amounts of RNA.

One limiting factor of this technique, given the ultimate aim to apply to RNA collected from nanobiopsies, is the need for primers to be gene-specific, as amplicons larger than 250 nt cannot be efficiently amplified due to the risk of secondary structure formation. Therefore, when expanding the gene panel for analysis, potentially hundreds of primers need to be optimised individually and together, to enable multiplexed RNA amplification prior to further processing, either by RT-qPCR or sequencing. Alternative RNA amplification techniques, such as LAMP, could also be utilised in conjunction with nanobiopsies, but is again limited by the need for gene-specific primers. Nonetheless, LAMP is able to tolerate light sequence variation and does have improved multiplexed functionality. However, as LAMP produces stem-loop structures, quantifying the end-product, especially when multiplexed, would be quite challenging, requiring purification and linearisation prior to further processing. Experiments conducted as part of this chapter highlights the need for fine-tuning of reagent combinations and reaction conditions for improved detection and reproducibility, especially when working on such small scales.

A key takeaway from experiments conducted within this chapter is that previously successful experimental conditions are not always reliable and do not necessarily translate to success in near identical experiments, especially when working with low input concentrations. NASBA efficiency appears to be highly sensitive to even minor fluctuations in reagent quality, enzyme activity, incubation parameters and RNA integrity, all of which are incredibly difficult to control for. The effect of these is especially heightened when working at a picogram or sub-picogram scale, where even slight variations in pipetting or temperature can have disproportionately large effects on reaction success and outcome. Furthermore, even when target-specific amplicons were detected (*β-ACTIN,* albeit at increased incubation times), they only constituted a small fraction of the total 'amplified' material, undermining the reliability of RT-qPCR Ct values as the sole indicator of amplification success. As such, progress in this regard requires integration with alternative detection techniques, such as gel or capillary electrophoresis.

In summary, this chapter highlights that robust, reproducible detection of low-abundance RNA from nanobiopsy samples remains a substantial challenge. Overcoming these obstacles will require significant innovation across enzymatic chemistry, nanoscale engineering and amplification technologies. Without such advancements, current methodologies, though conceptually promising, are unlikely to deliver the level of sensitivity and reproducibility necessary for high-resolution single-cell transcriptomics or dynamic therapeutic monitoring in contexts such as breast cancer.

CHAPTER 5
SINGLE-CELL INSIGHTS INTO CHROMATIN ARCHITECTURE AND THE TRANSCRIPTIONAL MODULATION OF 'TRANSIENT AWAKENING'

5.1 Introduction

5.1.1 Preliminary findings from TRADITIOM

Previous studies conducted in our lab by Rosano *et al*. have revealed that adaptation of MCF-7 cells to AI therapy involves a series of cycling events, wherein cells within dormant colonies of most replicate flasks transiently re-enter the cell cycle, before ultimately regressing and generating new dormant persistor populations.[2] Characterised by a temporary increase in colony size, this phenomenon, termed 'failed awakening' (FA), suggests that while awakening from dormancy involves re-entry into the cell cycles, this in itself is not sufficient for lead global resistance, with cells requiring further transcriptional reprogramming (figure 5.1 A). Further exploration into this phenomenon was conducted by use of the fluorescent ubiquitination-based cell cycle indicator (FUCCI) system. Geminin, a gene known to play an essential role in the cell cycle, with marked up-regulation as cells enter proliferation and expression increasing as the cell progresses through S/G2/M phases, was tagged with mCherry (mCh), while the nucleus was tagged with enhanced localised GFP, enabling tracking of the cells through the cell cycle, and cell detection and quantification, respectively. (figure 5.1 C (i)).[2,166,167] By monitoring both cell number and mCh expression upon -E2, an spike in mCh was observed, consistent with initial exponential cell growth during the latency period (figure 5.1 C (ii)), confirming that mCh, and thus geminin expression, are in fact reliable indicators of proliferation in this context. Cell counts markedly decrease as cells enter dormancy while mCh expression remain stable until approximately day 50, whereupon mCh signals begin to oscillate. Given the lack of variation in total cell count, this suggested that even during dormancy, cells may transiently re-enter the cell cycle, but stall in G2, avoiding the resultant increase in population size. This observation was distinct to the original proliferation dynamics observed (figure 5.1 A), wherein cell populations were found to expand and regress, without resulting in *bona fide* awakening – FA. As such, this cycling pattern, wherein cells transitioned to the S/G2/M phases without necessarily entering mitosis and resulting in population growth, was termed 'sporadic cycling'. Furthermore, single lineage tracking experiments, enabled by barcoding cells in the founder populations, revealed that while lineages enter dormancy with similar barcode dynamics, all awakening populations comprised different winner barcodes, with each following independent trajectories. Notably, this stochasticity in awakening, heterogeneity in the journey to the development of resistance to therapy and divergence during adaptation could not be associated to the presence of genetic mutations (both pre-existent and *de novo*), suggesting that the process of adaptation to therapy is predominantly driven by the acquisition of epigenetic cell state transitions. While this study gave rise to a number of valuable findings regarding non-genetic contributions to the development of resistance to AI therapy, it was limited by the inability to specifically probe the transcriptome of the cells that were found to transiently cycle, and truly

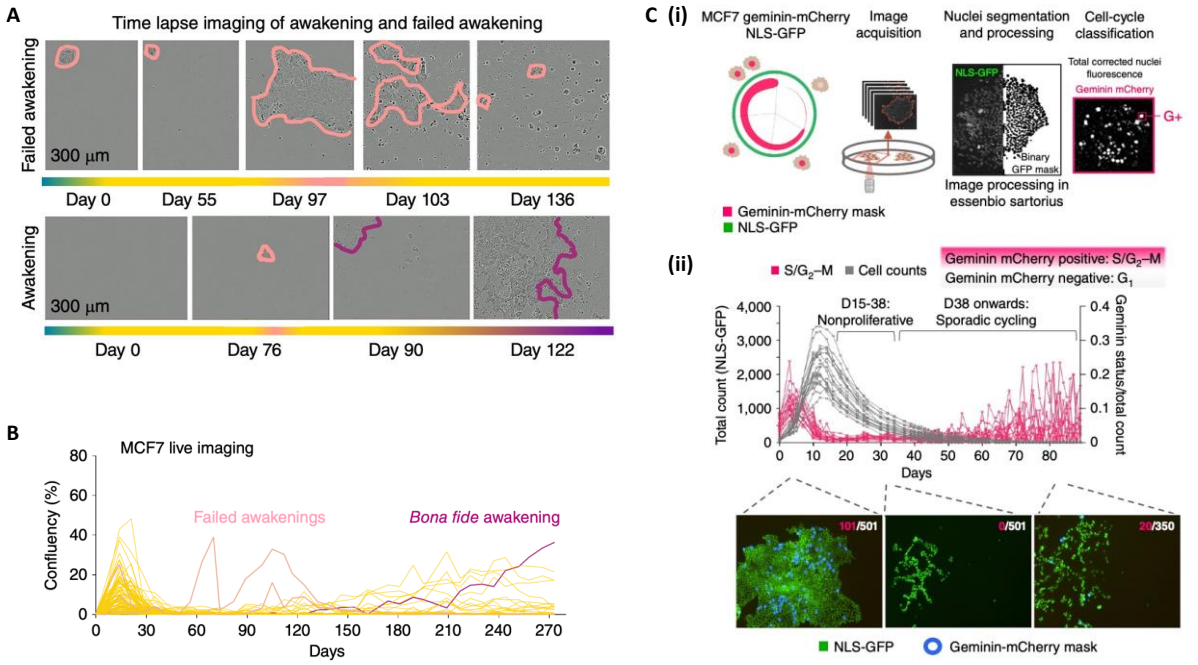


***Figure 5.1** Overview of Failed Awakening (FA) dynamics observed in Rosano et al.[2] **A** Time lapse imaging of MCF-7 cells revealed that certain colonies are able to transiently re-enter the cell cycle and recommence proliferation, before undergoing apoptosis, sometimes regenerating more dormant colonies – FA (pink). This is markedly distinct from global awakening (magenta), wherein colonies expand and continue to proliferate upon developing resistance to therapy. **B** Cell counts from each of the 108 scanning windows were plotted to show proliferation dynamics of MCF-7 cells, from commencement of AI to bona fide awakening (magenta). Colonies that undergo FA are shown in pink. **C** Further imaging experiments using an MCF-7 geminin-mCherry NLS-GFP cell line were conducted to further probe these dormancy-awakening dynamics. **(i)** Overview of image analysis; nuclei were segmented according to GFP expression before counting number of cells expressing mCherry, indicative of cells being in the S/G_2-M phases of the cell cycle. **(ii)** Cells were oestrogen deprived (-E2) and imaged daily for 3 months. A third of total replicates were characterized by the presence of dormant persistor colonies from entry into dormancy until the end of the imaging window, while the remainder of replicates exhibited signs of sporadic cycling from day 38 onwards. Geminin status is normalized to the total count (NLS-GFP positive; grey) and plotted in pink. Figure and legends adapted from D. Rosano et al., 2024.[2]*

differentiate between the transcriptome of **(i)** dormant, **(ii)** sporadic cycling and **(iii)** awakening cells.

5.1.2 Developing a multi-omic platform: combining scRNA and scATAC sequencing

To overcome these limitations, experiments in this chapter aimed to develop a single-cell multi-omic approach, combining scRNA-seq and scATAC-seq from the same cell, to enable a direct coupling of transcriptional state and regulatory landscape, with the goal of identifying how specific chromatin accessibility patterns give rise to, or fail to sustain, transcriptional programs required for reawakening. In this regard, a similar long-term, multi-modal study, using a FUCCI-like reporter cell line engineered by H. Dewhurst (Magnani Lab), was set-up. Despite not being wholly dynamic in nature, the ability to differentiate between cell cycle phases at key timepoints of the dormant life cycle opens up a wide range of analytical opportunities.

As outlined in section **1.3**, single-cell studies are at the forefront of research due to their ability to capture the complexity of biological systems with unprecedented resolution. This is critical in systems marked by stochasticity and plasticity, such as cancer, where a small subset of cells may drive treatment resistance, relapse, or metastatic progression. The transition from bulk to single-cell analytical techniques has also been accompanied by a natural progression into the field of multi-omics, combining these highly sensitive individual techniques to yield an even more powerful exploratory tool. Single-cell multi-omic techniques facilitate the simultaneous profiling of multiple

sub-cellular facets, providing a richer, integrative view of cell identity. For example, by coupling scRNA-seq and scATAC-seq, transcriptional outputs can be linked to their governing regulatory architecture – particularly valuable when studying FA, wherein the mechanisms underlying state transitions may not be solely genetic and may rely heavily on epigenetic reprogramming.

Sections **5.3.3.2** and **5.3.3.3** outline the technicalities of the scRNA-seq and scATAC-seq protocols chosen for this experiment - full length single-cell RNA-sequencing (FLASH-seq) and symmetrical strand single-cell combinatorial indexing (sci) ATAC sequencing (S3-ATAC-seq), respectively. FLASH-seq was selected as the scRNA-seq platform of choice due to its superior sensitivity, full-length transcript coverage, and efficiency with low-input material, all of which are properties that are particularly critical when working with rare, transient cell states.[169] Compared to droplet-based techniques such as 10x Chromium, which require 3'-capture, FLASH-seq's full-length capabilities facilitate detection of subtle transcriptomic reprogramming. Its high sensitivity also makes is uniquely suited to settings such as ours, wherein only partial RNA content is recovered, specifically cytoplasmic RNA, with the nuclear fraction reserved for ATAC-seq. This partial recovery, paired with the inherently low RNA content of dormant cells, necessitates a method capable of amplifying sparse transcripts without losing biological signal. Moreover, FLASH-seq has demonstrated success in cancer systems where rare resistant subpopulations are to be resolved from a heterogeneous background. Alternative full-length scRNA-seq techniques include vast transcriptome analysis of single cells by dA-tailing sequencing (VASA-seq). By enriching for both polyA and non-polyA RNA, via dA tailing, VASA-seq offers the advantage of enriching for non-coding RNA at significantly higher levels compared to current state-of the art methods, such as 10x Chromium and Smart-Seq3, thus providing a more inclusive view of a cell's transcriptome.[170] However, the resulting data complexity, heightened sensitivity to RNA degradation, and technical variability make it unideal in cases where material is limited, partially degraded, or derived from cytoplasmic fractions. S3-ATAC-seq was selected as the scATAC-seq method of choice, over other conventional techniques such as single-nucleus ATAC, droplet-based scATAC-seq and 10x Genomics' scATAC-seq due to its superior molecular capture rate, higher coverage and reduced rate of barcode collisions, all critical for accurately elucidating the subtle changes in chromatin accessibility that may underlie the early steps of dormancy escape.[171] Notably, compared to single-nucleus ATAC, droplet-based scATAC-seq and 10x Genomics' scATAC-seq, S3-ATAC-seq exhibited 13.7x, 6.22x and 6.02x-fold improvement in unique, paired, nuclear reads per cell. Combined, the technical superiority and enhanced signal resolution of S3-ATAC-seq and FLASH-seq informed the decision to incorporate these method into the multi-omic pipeline described in this chapter, enabling the high-resolution identification of reawakening trajectories from rare persistor cell states, with the hope that this knowledge could unlock new therapeutic strategies, either to eliminate reawakening-prone cells or prolong dormancy indefinitely, thereby reducing recurrence and improving long-term outcomes in ER+ breast cancer.

5.2 Materials & Methods

5.2.1 Cell line generation

A reporter MCF-7 cell line, employing the FUCCI system, was developed in the lab by Hannah Dewhurst. This involved tagging p21 at its endogenous locus with green fluorescent protein (GFP; heterozygous knock-in) using CRISPR-Cas9 and viral transduction of geminin-mCherry packaged HEK293-GP2 cells. This dual-fluorescence system enabled the visualization of cell cycle phases by monitoring p21 and geminin expression: GFP expression indicates that the cell is in the G1 or G2/M phase (with high GFP levels in G1 and low levels in G2/M), while mCherry expression increases as the cell progresses through S, G2, and M phases. Due to its ability to simultaneously

track geminin and p21 expression throughout the cell cycle, this cell line will be referred to as Gem21-MCF-7 in this chapter. A detailed schematic of the cell line generation is available in supplementary figure **S5.2**.

5.2.2 Experimental set-up

Gem21-MCF-7 cells were seeded at a density of 1.2 x 10^5 cells into 13 T75 flasks, in standard culture conditions as outlined in section **2.3.1.1**, two days prior to commencing treatment. After 48 hours, one flask was removed from culture and processed as outlined in section **5.2.4** or **5.2.5** (Day 0). The remaining flasks were washed twice with 1X PBS, and culture conditions were altered to mimic aromatase inhibition (-E2, oestrogen deprivation) – culture media was replaced with phenol-red free DMEM supplemented with 10% charcoal-stripped FCS (First Link) and with 2 mM L-glutamine, 100 units/mL penicillin and 100 units/mL streptomycin (Sigma), and media was no longer supplemented with E2. Cells were also supplemented with 500 µg/mL geneticin (Gibco) to maintain mCherry expression within the cell line. Notably, cells were not passaged throughout treatment, and underwent renewal of culture medium at regular timepoints as outlined in sections **5.3.2.1**. Cells were harvested at the following timepoints, determined by monitoring growth using Incucyte imaging, as described in section **5.2.4.1**: day 44 (entry to dormancy), day 63 (established dormancy), day 76 (FA/sporadic cycling) and day 104 (awakening).

5.2.3 Incucyte Imaging

5.2.3.1 Imaging experiments

Flasks were imaged three times weekly for the onset of -E2 treatment, initially on an Incucyte® Zoom Live-Cell Analysis System (from day 0 to 85), and then on an Incucyte S3 Live-Cell Analysis System (from day 85 onwards), due to technical difficulties. Prior to imaging, culture medium was changed to FluoroBrite™ DMEM (Gibco) supplemented with 2 mM L-glutamine, 100 units/mL penicillin and 100 units/mL streptomycin (Sigma) and changed back to standard WM conditions before returning to culture. Flasks were split into 108 scanning windows and imaged with a 10x objective in phase-contrast mode, with RFP and GFP filters.

5.2.3.2 Image Processing

Images were processed in FIJI (FIJI is just ImageJ) using two macros developed by Stephen Rothery, Facility for Imaging by Light Microscopy, Imperial College London. Initially, images of the same region of the flask were merged to combine all timepoints as a single timepoint (macro script in supplementary **5.2**). Images were then processed using a series of commands (macro script in supplementary **5.3**): frames were first smoothed by convoluting with Gaussian Blur, background signals were subtracted, brightness and contrast were adjusted and signals sharpened. Images were labelled as corresponding to either the RFP channel (mCh+) or GFP channel ($GFP_{high/low}$). A third channel was generated by merging RFP and GFP channels. From here, a binary image was created to enable quantification of the different fluorescent signals. Notably, since flasks were imaged with both the Incucyte Zoom and S3, images were analysed separately and intensity thresholds for positive cells were adjusted accordingly: RFP and GFP limits were set to 5 and 10 for Incucyte Zoom images, and 15 and 25 for Incucyte S3 images, respectively.

5.2.4 Sample processing

5.2.4.1 Fluorescence-activated cell sorting (FACS)

On collection days, cells were detached from flasks using 0.5% Trypsin-EDTA (Gibco) as outlined in section **2.3.1.1**. Cells were pelleted, washed with 1X PBS and resuspended in approximately 500 µL FACS buffer (10% FCS (First Link) in 1X PBS), before straining into a 5 mL round bottom test tube with a cell-strainer (Corning). Cells were sorted into four populations – mCherry positive (mCh+), GFP high (GFP_{high}), GFP low (GFP_{low}) and a double positive, i.e. mCherry and GFP positive (mCh+GFP+) population – on a BD FACS Aria II by the LMS NIHR Flow Cytometry Facility, Imperial College London (for experiments outlined in section **5.2.4.2**) and on a BD Symphony S6 by the Flow Cytometry Facility, Institute of Cancer Research, Chelsea (for experiments outlined in section **5.2.4.3**). Fractions were either collected in FACS buffer and kept on ice until ready for dispensing into a cellenCHIP (for experiments outlined in section **5.2.4.2**), or in DMEM (Lonza) supplemented with 10% FCS (First Link) and 2 mM L-glutamine, 100 units/mL penicillin and 100 units/mL streptomycin (Sigma) and transported to ICR Sutton on ice for multi-omic experiments (for experiments outlined in section **5.2.4.3**).

5.2.4.2 cellenCHIP 384-3' RNA-seq kit

All experiments conducted in this section were conducted on a Cellenone X1, hosted by the NIHR Imperial BRC Genomics Facility, at Imperial College London. A medium sized piezo-dispensing capillary (PDC), with a drop volume of between 300-450 µL, was used. Operating parameters used for cell dispensing into the Cellenchip are as follows: humidity = 60 %; dew point = 3.2; temperature = 4 °C; nozzle voltage = 119 V; pulse length = 49 µs and frequency = 500 Hz. All library incubation steps were conducted on a ProFlex™ PCR System (Applied Biosystems).

5.2.4.2 (a) Dispensing into cellenCHIP

When ready to dispense, cells were carefully washed and resuspended in 1X PBS at a concentration of 200 cells/µL. 96 cells from each population fraction were dispensed into separate arrays of the 384-well cellenCHIP using the pre-set cellenCHIP_FTRP program. The chip was spun down at 2000 x g for 90 seconds at 4 °C and immediately processed further to generate the 3'-RNA-seq library.

5.2.4.2 (b) Library preparation

The cellenCHIP and thermal adapter were secured into the thermal module of a Veriti 96 Thermocycler (Applied Biosystems). A cellenSEAL was placed on top of the cellenCHIP and cDNA was synthesised by incubating the chip at 40.4 °C for 1h30, with the lid set to 50 °C. The cellenCHIP was then spun down at 2000 x g for 90 seconds at 4 °C. Unless otherwise specified, all centrifugation steps from here were conducted using these parameters. The cellenSEAL was removed and the cellenCHIP, placed on top of a cellenCHIP 384 funnel, facing downwards and spun down to collect the cDNA product. Material was pooled into a single 1.5 mL reaction tube and processed as a single sample. Sample volume was made up to 50 µL using nuclease-free water (NEB), 1 volume of cellenCLEAN beads were added and mixed thoroughly by pipetting, and the sample was incubated at room temperature for 10 minutes to allow cDNA to bind. The tube was then transferred to a magnetic stand and once the supernatant ran clear, it was discarded. The cDNA-conjugated beads were washed twice, with the tube still on the magnetic rack, with 200 µL 80 % ethanol. cDNA was eluted into 16 µL nuclease-free water by incubating at room temperature for 10 minutes, before adding 20 µL cellenAMP Enzyme Mix and 5 µL cellenAMP Primer Mix to amplify the cDNA using the following programme (table **5.1**):

Table 5.1 ***CellenCHIP cDNA amplification PCR programme parameters.***

Step	Temperature	Time	
1	98 °C	5 min	
2	98 °C	30 sec	x18
3	65 °C	1 min	
4	72 °C	4 min	
5	4 °C	Hold	

The amplified product was cleaned up by adding 32 µL cellenCLEAN beads and mixing thoroughly by pipetting. The sample was incubated at room temperature for 10 minutes before transferring to a magnetic stand and discarding the supernatant once it ran clear. The cDNA-conjugated beads were again washed twice with 200 µL 80 % ethanol. Amplified cDNA was eluted by adding 17 µL nuclease-free water and incubating for 10 minutes. cDNA was quantified using a Qubit™ 1X dsDNA HS Kit (Invitrogen™) and the cDNA profile was verified by running a High Sensitivity D5000 Screentape on a Tapestation 4200 (both Agilent).

If cDNA fragments fell within a 250-5000 bp range, a library was prepared for Illumina sequencing. 4 µL 0.25 ng/µL cDNA sample and 6 µL cellenTAG buffer were added to a PCR tube containing 8 µL cellenTAG reagent and carefully mixed by pipetting. cDNA was tagmented by incubating at 55 °C for 15 minutes. Tagmentation was stopped by adding 9 µL cellenSTOP solution to the reaction and incubating at 68 °C for 10 minutes. 27 µL cellenCLEAN beads were added to the now barcoded library, mixed, and incubated at room temperature for 5 minutes, before placing the tube on a magnetic rack for a further 5 minutes. The supernatant was discarded, and the remaining beads were washed twice with 200 µL 80 % ethanol. Purified library was eluted by adding a further 21 µL nuclease-free water to the cellenCLEAN beads and incubating off the magnetic stand for 10 minutes. 25 µL cellenAMP enzyme mix and 5 µL cellenLIBRARY primers were added to the sample, and the library was amplified using the following protocol:

Table 5.2 ***CellenLIBRARY amplification PCR programme parameters.***

Step	Temperature	Time	
1	72 °C	10 min	
2	95 °C	3 min	
3	98 °C	30 sec	x12
4	64 °C	15 sec	
5	72 °C	30 sec	
6	72 °C	3 min	
7	4 °C	Hold	

The amplified library was topped up to 50 µL with nuclease-free water (NEB) and 25 µL cellenCLEAN beds were added. The tube was incubated at room temperature for 5 minutes before transferring to a magnetic rack. The supernatant was transferred to a new reaction tube once it ran clear, and a further 15 µL cellenCLEAN beads were added and incubated for 5 minutes, before discarding the supernatant and washing the beads twice with 200 µL 80% ethanol. The final size-selected library was eluted into 17 µL nuclease-free water by initially incubating the tube off the magnet for 10 minutes before transferring to the magnet for a further 5 minutes. The library was quantified using a Qubit™ 1X dsDNA HS kit (Invitrogen™) and the size profile was verified using a High Sensitivity D5000 screentape (Agilent).

Samples were sequenced by Novogene using Illumina's paired-end 150 (PE150) approach.

5.2.4.3 Multi-omic experiments

This section highlights the development of a multi-omic technique, enabling both full-length scRNA and sc-ATAC-seq from the same single cell. This was achieved using a combination of three techniques: Genome & Transcriptome-sequencing (G&T-seq), full length scRNA-sequencing with unique molecular identifiers (FLASH-seq with UMI) and S3-ATAC-seq.[171–173] All cell dispensing experiments conducted in this section were done on a Cellenone X1, hosted by the Genomics and Evolutionary Dynamics Group, Institute of Cancer Research, Sutton. Operating parameters used for Cellenone dispensing are as follows: humidity = 55 %; dew point = 15.3; temperature = 8 °C; nozzle voltage = 95 V; pulse length = 52 µs and frequency = 500 Hz. Composition of all reagent mixes and primer sequences used in this section are outlined in table 5.6. All heated incubations were conducted in a C1000 Touch™ Thermal Cycler (Bio-Rad).

5.2.4.3 (a) Nuclei extraction and tagmentation

5 x 10^4 cells per fraction were pelleted by centrifugation at 300 x g for 4 minutes at 4°C. The cell pellet was washed with 2% BSA/PBS and pelleted again before resuspending in 100 µL nuclei extraction buffer (table 5.6.1; adapted from EpiCypher® CUTANA™ CUT&RUN Protocol v1.7) and incubated on ice for 3 minutes.[174] The reaction was quenched by addition of 500 µL 2% BSA/PBS. Nuclei were pelleted and resuspended in 100 µL 2 % BSA/PBS. Success of nuclei extraction was assessed by checking cell viability using a LUNA FL™ Dual Fluorescence Cell Counter (Logos Biosystems), with 0% viability confirming successful nuclei extraction.

Nuclei were then resuspended in 50 µL bulk Tn5 mix (table 5.6.2) by gently pipetting up and down 6 times. Nuclei were tagmented by incubating at 37 °C for 30 minutes at 1000 rpm on a ThermoMixer® F1.5 (Eppendorf®). The tagmentation reaction was stopped by adding 50 µL STOP buffer (table 5.6.3) and adding 400 µL 2% BSA/PBS. Tagmented nuclei were spun down at 400 x g for 10 minutes at 4 °C, resuspended in 70 µL 2% BSA/PBS and kept on ice until ready for Cellenone dispensing.

5.2.4.3 (b) Cellenone dispensing

Cells were strained using a 20 µm pluriStrainer Mini (pluriSelect) to minimise clumping and loaded into the Cellenone, ready for dispensing. 96 cells from each fraction were dispensed into a 384-well plate, pre-filled with 1 µL ATAC lysis buffer (table 5.6.4). Plates were spun down at 1500 x g for 5 minutes at 4 °C and stored at -80 °C until ready for processing.

5.2.4.3 (c) Initial processing: combined G&T-seq & FLASH-seq approach

50 µL Streptavidin-conjugated Dynabeads (Invitrogen™) were placed on a magnetic rack until the supernatant ran clear. The Dynabeads were resuspended in 200 µL Dynabead solution A (table 5.6.5) and returned to the magnet until the supernatant ran clear again. This was repeated once more with Dynabead solution A and once with Dynabead solution B (table 5.6.6). The tube was removed from the magnet and Dynabeads were resuspended in 50 µL 2X Binding and Wash buffer (table 5.6.7). 7.6 µL 100 µM SRTR-P1-T31 oligo (table 5.6 – primer sequences) and made up to a total volume of 50 µL with nuclease-free water. The sample was incubated for 30 minutes at 20 °C and 1000 rpm on a ThermoMixer® F1.5 (Eppendorf®). Oligonucleotide-conjugated Dynabeads were then washed with 1X Binding and Wash buffer by resuspending in buffer, placing on a magnetic rack and discarding supernatant once it ran clear while the tube was still place on the magnetic rack. This wash was repeated a total of 4 times. Beads were resuspended in 384 µL Bead Mix (table 5.6.8) and stored on ice until ready for dispensing.

When ready to process, plates were thawed on ice and spun down at 1500 x g for 5 minutes at 4 °C, followed by a 5-minute incubation at 55 °C to release RNA. 1 µL Bead Mix containing the Dynabeads was then dispensed into each well. The plate was spun down and vortexed to ensure the Dynabeads were fully resuspended, before being incubated for 20 minutes at 1300 rpm. Supernatant was separated from the beads using a 384-well magnet and transferred to a new pre-i5-indexed plate for processing via S3-ATAC-seq. From here, all processing was carried out separately for FLASH-seq and S3-ATAC-seq (sections **5.2.4.3 (c) and (d)** respectively).

5.2.4.3 (c) FLASH-seq

RNA was processed according to the FLASH-seq UMI protocol v3.[173] 5 µL RT mix (table 5.6.9) was added to each well of the original plate (now containing only dry Dynabeads with captured RNA). The plate was spun down and RT-qPCR was conducted according to the following protocol (table 5.3):

***Table 5.3** FLASH-seq RT-PCR programme parameters.*

STEP		TEMPERATURE	TIME	CYCLES
RT		50 °C	60 min	1x
PCR	Initial denaturation	98 °C	3 min	1x
	Denaturation	98 °C	20 sec	20-24x
	Annealing	65 °C	20 sec	
	Elongation	72 °C	6 min	
		15 °C	Hold	

A mix of 3.6 mL SPRIselect beads (Beckman Coulter) and 4 mL nuclease-free water was allowed to equilibrate at room temperature for 15 minutes. 19 µL of this bead mix was added to each well of the 384-well plate, mixed by vortexing and left to incubate for 5 minutes at room temperature. The plate was then returned to the 384-well magnet for a further 5 minutes to allow the beads to separate. The supernatant was discarded and the 15 µL nuclease-free water was added to each well and mixed by vortexing. The plate was again left to incubate for 2 minutes at room temperature before returning to the magnet for a further 2 minutes. 14 µL supernatant (containing RNA) was transferred to a new 384-well plate for subsequent processing. From here, success of amplification was tested by pooling 1 µL of sample from each well of the 384-well plate into a single tube, and checking the cDNA profile with a High Sensitivity D5000 screentape on a Tapestation 4200 (both Agilent). As no cDNA was detected here, further library preparation was not conducted.

5.2.4.3 (d) S3-ATAC-seq

The pre-i5-indexed plate, containing supernatant, was carefully spun down and incubated for 55 °C for 15 minutes, 70 °C for 10 minutes before cooling to 10 °C. 1 µL 1% Triton X-100 and 1 µL NEBNext® Ultra™ II Q5 ® Master Mix (NEB®) was added to each well and mixed by vortexing. Gap filling and reverse strand synthesis was then initiated using the following limited extension PCR protocol (table 5.4):

***Table 5.4** S3-ATAC-seq gap filling and reverse strand synthesis PCR programme parameters.*

STEP		TEMPERATURE	TIME	CYCLES
PCR	Initial denaturation	98 °C	30 sec	1x
	Denaturation	98 °C	10 sec	10x
	Annealing	59 °C	20 sec	
	Elongation	65 °C	60 sec	
		10 °C	Hold	

5 µL ATAC-seq PCR Master Mix (table 5.6.10) was added to each well and the sample was PCR amplified according using the following protocol (table 5.5):

Table 5.5 S3-ATAC-seq PCR amplification parameters.

STEP		TEMPERATURE	TIME	CYCLES
PCR	Initial denaturation	98 °C	30 sec	1x
	Denaturation	98 °C	10 sec	14x
	Annealing	65 °C	75 sec	
	Elongation	72 °C	2 min	
		10 °C	Hold	

All wells from the 384-well plate were pooled together to yield a final volume of 3072 µL. To this, 5 volumes (15.36 mL) DNA binding buffer (Zymo Research) was added, and the whole sample was run through a column. Resulting DNA was eluted in 40 µL nuclease-free water (NEB®) and library size and quality were validated by running a High Sensitivity D5000 screentape on a Tapestation 4200 (Agilent).

Table 5.6 Summary of reagent mix and buffer compositions, and primer sequences, used for multi-omic experiments, methods of which are outlined in section 5.2.4.3.

REAGENT	FINAL CONCENTRATION/VOLUME
1. Nuclei Extraction Buffer	
HEPES-KOH pH 7.9 (ICR Media Store)	20 mM
KCl (Invitrogen™)	10 mM
10% (v/v) Triton X-100 (Sigma-Aldrich®)	0.1%
Spermidine (Sigma-Aldrich®)	0.5 mM
Glycerol (Sigma-Aldrich®)	20%
cOmplete Protease Inhibitor EDTA-Free Inhibitor (Roche)	1 tablet/10 mL
Nuclease-free water (NEB®)	Up to 10 mL
2. (a) Bulk Tn5 mix	
Tagmentation Buffer*	25 µL
PBS (ICR Media Store)	16 µL
Tween 10% (Sigma-Aldrich®)	0.5 µL
Digitonin 1% (Sigma-Aldrich®)	0.5 µL
RiboLock RNase Inhibitor (40 U/µL) (Thermo Scientific™)	1 µL
MACS® BSA (10%) (Miltenyi Biotec)	5 µL
Tn5 (indexed) (provided by Max Mossner)	2 µL
2. (b) Tagmentation Buffer*	
Tris-HCl pH 7.6 (Invitrogen™)	20 µL
$MgCl_2$ (Invitrogen™)	10 µL
DMF (Sigma-Aldrich®)	0.1%
Nuclease-free water (NEB®)	Up to 50 µL
3. STOP Buffer	
Nuclei dilution buffer (20x) (10X Genomics®)	2.5
DTT (Thermo Fisher Scientific)	0.5
Nuclease-free water (NEB®)	47
4. ATAC lysis buffer	
Tagmentation Buffer*	0.5 µL
Qiagen protease (30 AU) (Qiagen)	0.2 µL
A14T-SO primer	0.01 µL

SDS (1 %) (Sigma-Aldrich®)	0.01 %
Nuclease-free water (NEB®)	Up to 1 µL
5. Dynabead Solution A	
NaOH (Invitrogen™)	0.1 M
NaCl (Invitrogen™)	0.05 M
6. Dynabead Solution B	
NaCl (Invitrogen™)	0.1M
7. 2X Binding and Wash Buffer	
Tris-HCl (pH7.4) (Invitrogen™)	10mM
EDTA (Invitrogen™)	1mM
NaCl (Invitrogen™)	2M
8. Bead Mix	
RiboLock RNase Inhibitor (40 U/µL) (Thermo Scientific™)	1 U/µL
9. RT mix	
10% (v/v) Triton X-100 (Sigma-Aldrich®)	0.20%
dNTP mix (25 mM each) (NEB®)	6 mM
DTT (Thermo Fisher Scientific)	5 mM
RiboLock RNase Inhibitor (40 U/µL) (Thermo Scientific™)	1U/µL
Betaine (Sigma-Aldrich®)	1M
dCTP (NEB®)	9 mM
$MgCl_2$ (Invitrogen™)	9.2 mM
SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific)	2 U/µl
KAPA HiFi HotStart ReadyMix (Roche)	1X
TSO-UMI primer	1.84 µM
Tn5_ISPCR- F primer	0.5 µM
DI-PCR-P1A-R primer	0.1 µM
10. ATAC-Seq PCR Master Mix	
NEBNext® Q5U® Master Mix (NEB®)	3 µL
i7-N706 primer	0.05 µM
Nuclease-free water (NEB®)	Up to 5 µL
Primer Sequences	
STRT-P1	5'-(Biotin)AATGATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'
TSO-UMI	5'-(Biotin)AAGCAGTGGTATCAACGCAGAGTNNNNNNNNCTAACrGrGrG-3'
Tn5-ISPCR- F	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCAGTGGTATCAACGCAGAGT-3'
DI-PCR-P1A-R	5'-AATGATACGGCGACCACCGA-3'
i7-N706	5'- CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTC'-3'

5.3 Results & Discussion

Contributions:

- *The gem21-MCF-7 cell line used in this chapter was generated and validated by Hannah Dewhurst, Magnani Lab, Imperial College London.*
- *Experimental design, all imaging experiments and analysis, cellenCHIP 384-3' RNA-seq experiments and multi-omic sample processing were conducted as a joint project with Hannah Dewhurst.*
- *Multi-omic experimental design and all Cellenone dispensing in relation to this were conducted with as a joint project with Hannah Dewhurst, in collaboration with Max Mossner, Genomics and Evolutionary Dynamics Group, Institute of Cancer Research, Sutton.*

5.3.1 Imaging experiments

The gem21-MCF-7 cell line was seeded into 13 replicate flasks and treated with -E2 for various timepoints. In addition to the geminin-mCherry reporter described in section **5.1,** the gem21-MCF-7 cell line used

for these experiments also included conjugation of the cyclin-dependent kinase, p21, with a GFP label. P21 is known to be an important cell cycle regulator, arresting in G1/S and G2/M to prevent cell cycle progression, by inhibition of various cyclins – CDK4,6/cyclin-D and CDK2/cyclin-E, respectively.[176] As such, incorporation of a label to enable tracking of p21 expression alongside mCh expression allows for the following fluorescence-cell cycle classifications: mCh+ = G2/M phase, or TA; mCh+GFP+ = S phase; GFP_{high} = G1 arrest, or late G1 phase; GFP_{low} = G1 arrest, or early G1 phase. For the duration of cell culture, flasks were imaged twice weekly using the Incucyte® Live-Cell Analysis System. Media was changed to Fluorobite™ DMEM prior to imaging. With autofluorescence comparable to PBS, imaging using Fluorobite™ significantly increases the SNR, especially beneficial when measuring weak fluorescent signals during live-cell imaging.[177] Due to technical difficulties, images were captured using the Incucyte® Zoom until day 85 and with the Incucyte® S3, which has significantly superior resolution, from day 85 until final collection. This resulted in much more accurate brightfield and fluorescent detection and quantification. Given the difference in fluorescent channel sensitivity and image resolution, continuous analysis could not be conducted, as significantly more cells were detected in both phase-contrast, and the GFP and RFP channels using the Incucyte® S3. As such, images from day 0 to day 85 and day 85 until awakening are treated as different experiments. However, effects of this can only be seen in the awakening timepoint flasks, as all other flasks were collected prior to this changeover timepoint. Data for the awakening timepoint is not shown, as this timepoint was not taken forward for sequencing due to technical issues which are discussed in section **5.3.2.2**.

Each flask was segmented into 108 scanning windows, and the change in cell count and fold changes in GFP+, mCh+ and mCh+GFP+ expression were analysed. Analysis of cell growth dynamics reveals initial exponential growth until approximately day 10, consistent with E2 latency, wherein cells continue to proliferate, as if in +E2 conditions, followed by mass cell death until around day 30 (figure 5.2 A). This is also mirrored in the normalised count of mCh+ cells, with an increase in expression until day 10, suggesting that mCh+ expression is a good indicator of cycling dynamics (figure 5.2 B). Cell counts fluctuate slightly beyond this timepoint for most scanning windows, with some showing larger spikes until approximately day 40 before settling and remaining mostly stagnant for the remainder of the imaging period. Colonies in certain replicates however did seem to oscillate between slight growth and regression from day 40 onwards (figure 5.2 A; pink), indicative of sporadic cycling. A jump in the mCh+ expression at this timepoint also suggests the presence of a larger proportion of cycling cells (figure 5.2 B). Closer examination of Incucyte™ images of this specific imaged region show little fluctuation in overall percentage of the various reporter populations in individual flasks (figure 5.2 C), beyond day 16. This could suggest that a similar proportion of each cell cycle population dies off during this period, which could occur distinctly, or could be coupled with cells continuing to cycle, but at a much slower rate than would otherwise occur in normal growth conditions. As Imagelock plates were not used for cell culture and imaging, it is difficult to confidently assume that the same cell populations are imaged within each frame, as there was no way to ensure precise image registration. Therefore, comparison of these frames between different timepoints is prone to error, and we may in fact be considering adjustment in flask positioning, resulting in different imaged areas, as colony size expansion, although this is minimal in this case given the induction of dormancy.

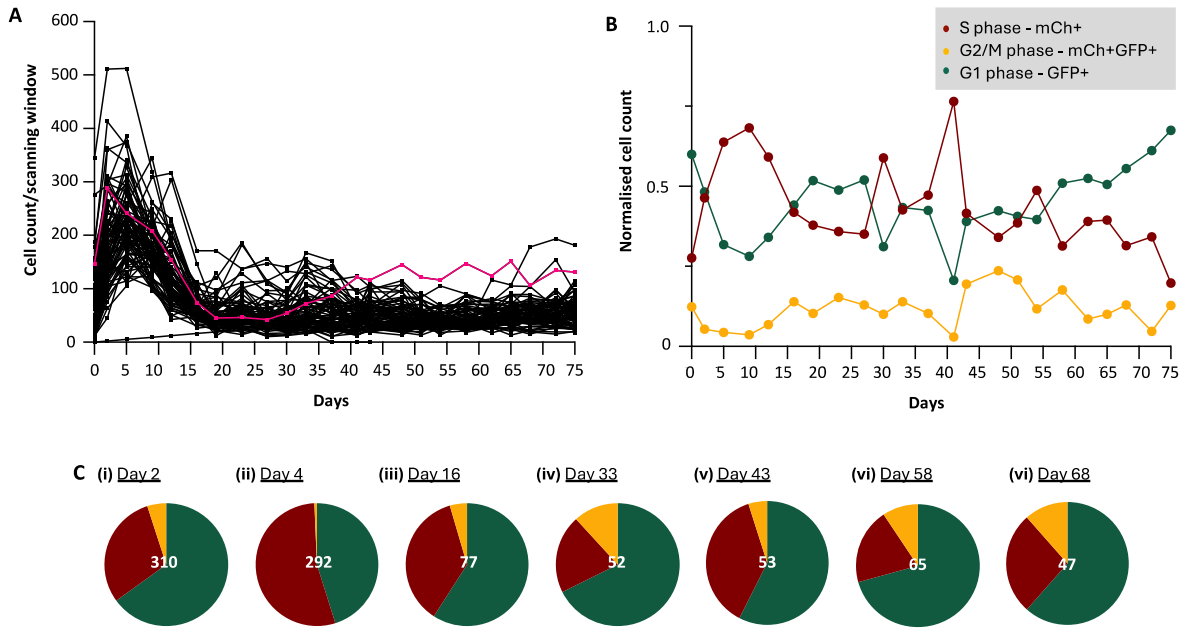


Figure 5.2 ***Live-cell imaging of gem21-MCF-7 cells undergoing aromatase inhibition therapy.*** ***A*** *Cell proliferation dynamics over the duration of imaging (75 days). Flasks were divided into 108 scanning windows, with each scanning window treated as a separate replicate in this case. Cell counts from each of these windows was plotted as a function of the imaging timepoint. Scanning window showing signs of sporadic cycling is highlighted in pink.* ***B*** *mCherry and GFP expression in cells, normalised to the total cell count, showing fluctuations in proportion of total cells expressing the different reporters, enabling determination of cell cycle phase. This data is also* ***C*** *summarized in pie charts, allowing more visual representation of proportion of cells in each cell cycle phase. Total cell count at each of these timepoints is highlighted in white at the centre of each pie chart.*

Results obtained here highlight that population size and cell cycle dynamics can in fact be monitored longitudinally, and inform us of escape from dormancy, whether that is temporary (FA or sporadic cycling) or permanent (by establishing resistance to therapy), enabling downstream processing of samples to further understand cellular processes at key timepoints. Presence of colonies that transiently cycle through different cell states, even during dormancy, confirms that cells can adopt hybrid states, between dormancy and awakening, warranting further exploration.

5.3.2 Cellenchip 384-3' RNA-seq Kit

The Cellenone is a standalone system that enables high-precision single-cell isolation and down to pico (pL) drop volume dispensing. Using a proprietary piezo-dispensing technology (sciDROP PICO), cells can be dispensed into user-defined locations based on their size, morphology or even fluorescent markers using gentle acoustic waves, ensuring minimal stress on cells.[178] Given the extremely small volumes dispensed through the PDC and high recovery, this platform holds great promise for rare cell detection and isolation, provided the sample can be accurately categorised and differentiated from other cells within the samples using the Cellenone's detection parameters.

5.3.2.1 Preliminary fluorescence detection tests

Previous experiments with the Cellenone, conducted by other members of the lab, revealed difficulty in detecting both GFP and mCherry signals – both of which were present in the reporter cell line. Therefore, preliminary experiments, using both mixed populations and FACS-sorted GFP+ and mCherry+ cells were conducted. An example of an isolation scatter graph is shown in figure 5.3 A (i), for FACS-sorted GFP+ cells. While GFP+ cells were generally successfully detected and could be sorted from a mixed cell population, many low

intensity GFP+ cells from the sorted GFP+ population could not be detected; with only 16% of total cells detected in the FITC channel, despite using the maximal FITC threshold range. Ignoring the size outliers (those in pink), 30 % of the remaining cells were successfully detected according to the thresholds, with the remaining 70 % of detected cells within the right diameter range but notably falling outside the GFP channel threshold. Additionally, cells were only detected to be GFP+ if FITC intensity exceed 30 a.u. (figure 5.3 A (ii)). Further issues were encountered when attempting to detect mCherry. Despite sorting for an mCherry+ population prior to use in the Cellenone, very little to no signal was detected in single cells, with only very feint signals observed, even in cell clumps. These issues were compounded when attempting to differentiate between GFP+ and mCherry+ cells from a mixed, unsorted population. While it was initially hoped that population sorting could be conducted on the Cellenone due to its gentle, low-stress vibrational dispensing mechanisms, the inability to reliably differentiate between cells in early and late G1 based on GFP intensity, along with the issues encountered in detecting mCherry made this unfeasible. It was therefore decided that cells would be FACS-sorted into their distinct populations prior to Cellenone dispensing, despite the potential risk of this exerting increased stress on the cells.

5.3.2.2 Dispensing cells into the 384-3' RNA-seq kit and library preparation

Gem21-MCF-7 cells were detached from flasks and FACS-sorted into their respective populations (as classified in section **5.3.1**) at four timepoints, corresponding to untreated cells (day 0), established dormancy (day 44), and two sporadic cycling timepoints (day 63 and 76). A final timepoint corresponding to awakening was also due to be collected, but was decided against due to the issues discussed later in this section. At each timepoint, 96 cells from each population fraction were dispensed into the pre-barcoded and RTready arrays of a cellenCHIP, using only the Cellenone's cell size threshold. Coupled with an optimised miniaturised library preparation kit, the cellenCHIP enables the transcriptomic analysis of individual cell. Using 3'-RNA-sequencing, this kit measures the genome-wide abundance of untranslated regions and isoforms at the 3' end of mRNA, yielding insight into the post-transcriptional fates of cells' RNA, with low sequencing depth and high-sensitivity.[179,180] While this sequencing technique provides less differential gene expression than whole-transcript RNA-sequencing, it provides more accurate gene quantification and detects differential gene expression more efficiently in transcripts shorter than 500 bp. For all cellenCHIP dispensing experiments, low temperatures and dew points were used to minimise reagent evaporation, which is significantly enhanced on a pL scale. This is especially important when working with the cellenCHIP, as evaporation of lysis buffer and RT reagents from individual wells prior to or during cell dispensing would result in significant skew. Once cells were dispensed, cellenCHIPs were processed according to optimised kit instructions, with success of cDNA amplification and subsequent library preparation (only if cDNA amplification was successful) determined by running samples on an Agilent D5000 screentape. Unfortunately, given the novelty of this kit and its component software, various issues were experienced

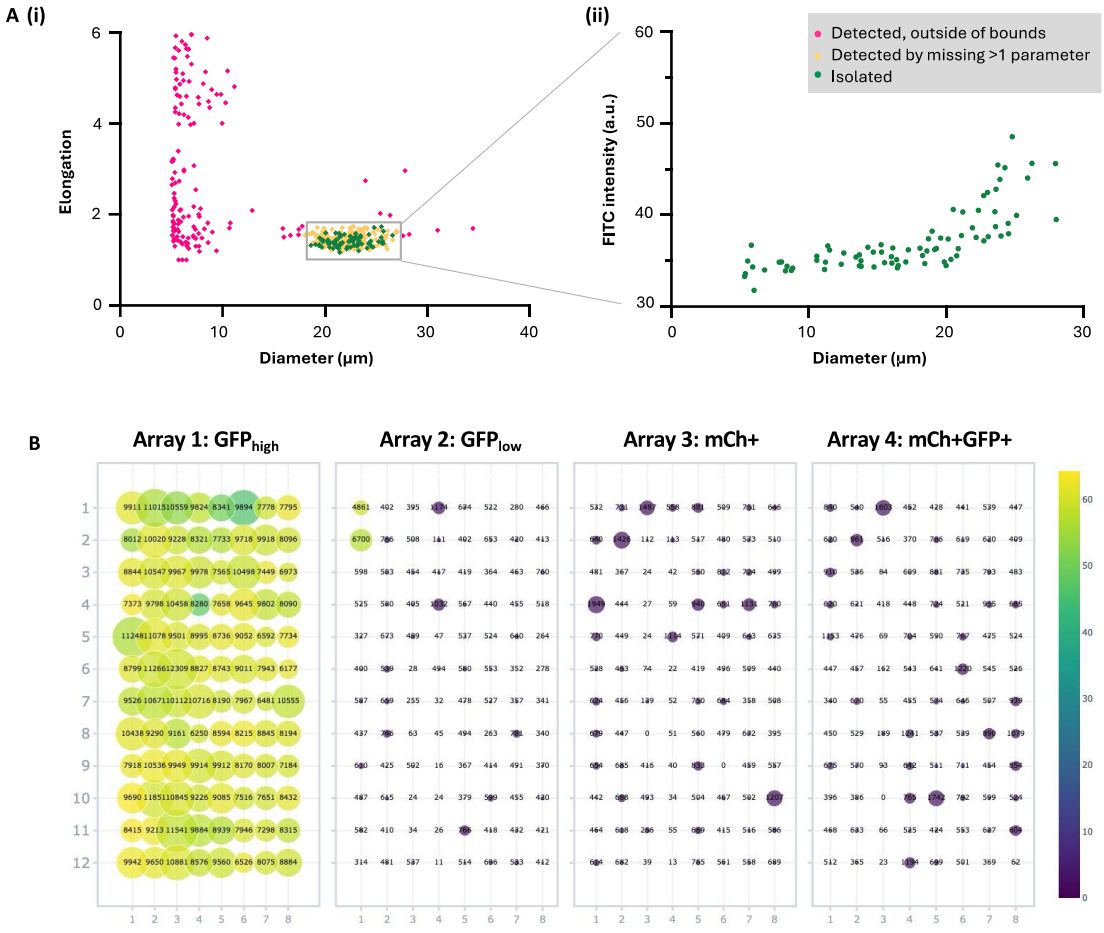


Figure 5.3 ***Overview of Cellenone dispensing test and cellenCHIP experiment. A (i)*** *FACS-sorted GFP+ cells were dispensed into a 96-well plate using the cellenone FITC detection channel. Despite having a population comprising solely of GFP+ cells, only a small population of cells were detected within the FITC thresholds, highlighting difficulty in detection and differentiation between cells of different intensities.* ***(ii)*** *The lowest FITC intensity was detected at 30 a.u. B Mapping rate and reads for sequenced day 77 cellenCHIP. As seen from the mapping rate (colour of each well), and the number of genes detected across each well (number), no cells were dispensed into all but two wells of array 2 and in arrays 3 and 4. As cell barcoding occurs post-dispensing into the cellenCHIP, there was no way to differentiate between the different cell fractions. The FACS-sorted population that each array was to correspond to is highlighted above the cellenCHIP.*

during dispensing, notably with cellenCHIP detection prior to dispensing cells (for day 44) and with dispensing the different fractions into separate arrays (as seen for day 77; figure 5.3 B). Issues with cellenCHIP detection were especially problematic, as this delayed the experiment by approximately one hour (half of the recommended dispensing time), which, despite ensuring that the cellenCHIP was kept cool, resulted in significant reagent evaporation – identified only upon pooling samples post-cDNA synthesis from a small total sample volume. Furthermore, no cDNA was detected during quantification prior to library preparation, likely due to inefficient cell lysis and reverse transcription as a result of reagent evaporation, and as such the sample was not processed further. Software issues experienced during the day 77 collection meant that while each cell fraction was set to be dispensed into separate arrays, they were all dispensed into the same wells, as seen in figure 5.4 B, wherein the number on each well corresponds to the number of genes detected, and the colour of the well corresponding to the mapping rate. Given that cell barcoding was to occur only after dispensing, it was not possible to distinguish between the cell cycle states and unique transcriptomes with multiple cells per well. The final awakening timepoint (at approximately day 104) was not collected due to these identified issues with the reagent kit and software. Theoretically, if this experiment were successful, given the unique barcodes in each well of the cellenCHIP, cells'

transcriptomes at different stages of the cell cycle, pre-determined by reporter expression, could be deconvoluted, widening our understanding of sporadic cycling at a single-cell level. This experimental protocol was not attempted again as the reagent kit was discontinued.

5.3.3 Developing a multi-omic single-cell analytical pipeline

To optimise gene detection from single cells, an alternative protocol, which couples full-length RNA-sequencing and ATAC-sequencing, from the same cell, was attempted. This was done in collaboration with Max Mossner, from the Genomics and Evolutionary Dynamics Group, Institute of Cancer Research, Sutton, who has significant expertise with working with and developing miniaturised protocols with the Cellenone X1 platform. Given the novelty of this method, experiments focused on optimising the different sequencing techniques and did not involve imaging (section **5.3.1**).

5.3.3.1 Initial processing: tagmentation, cellenone dispensing and G&T-seq

As in section **5.3.2.2**, AI-treated gem21-MCF-7 cells, at different treatment timepoints were detached from culture flasks, FACS-sorted into the four distinct populations and transported to the Institute of Cancer Research, Sutton for all downstream processing. Nuclei were extracted according to the nuclei extraction protocol established by EpiCypher® for their CUT&RUN protocol, previously validated within the lab (data not included), which reproducibly yielded between 0-5% cell viability.[174] As nuclei extraction involves rupture of the cell membrane, a viability of 0 %, wherein no cell membranes can be detected, is desired. In cases where higher viability was obtained, indicative of poor nuclei extraction, significant cell clumping was observed, suggestive of overlysis. Nuclei were initially tagmented *in situ*, in preparation for S3-ATAC-seq processing (section **5.3.3.2**), wherein any unfragmented DNA was fragmented by a Tn5 enzyme and subsequently tagged with universal adapter sequences, before dispensing 96 nuclei from each fraction into a 384-well plate using the Cellenone.[181] Cells were then processed *via* G&T-seq, enabling separation of genomic DNA (gDNA) for mapping chromatin accessibility using S3-ATAC-seq, and mRNA for transcriptomic analysis using FLASH-seq, from the same cell, using streptavidin-conjugated beads. By ensuring that material from the same cell is processed *via* two different sequencing techniques, a more comprehensive, representative view of mechanisms of gene regulation, effects of chromatin architecture and differentially expressed genes can be obtained. This is especially valuable when attempting to decipher the heterogeneity across dormancy and awakening.

5.3.3.2 FLASH-Seq

Once RNA had been separated from gDNA using G&T-seq, the RNA was processed using the FLASH-seq UMI protocol, to generate a scRNA-seq library.[173] Using the switching mechanism at the 5' end of the RNA template (SMART) technique, wherein reverse-complements of template are generated by alternating replication along both the complementary and nascent DNA strand, FLASH-seq yields full-length transcriptomic information with unparalleled sensitivity.[169,182] This, in and of itself, is a significant advantage over the 3'-RNA-seq technique that the cellenCHIPs adopt, which preferentially amplifies the 3' terminus of mRNA, limiting analysis only to distal splicing events.[183] Overall, SMART-seq techniques provide a whole breadth of additional information, including larger mRNA coverage and enhanced identification of alternative transcript isoforms and single-nucleotide polymorphisms, which, given the numerous unknowns and the need to characterise the various cell cycle states that dormant cells transition through to develop resistance, is extremely beneficial.[72,169,184] Most notably, FLASH-seq

has been shown to yield improved gene detection with a reduced number of PCR cycles, without modifying sequencing depth or compromising data coverage and quality.

Unfortunately, initial attempts at generating cDNA from the mRNA captured using streptavidin-beads from treatment-naïve gem21-MCF-7 cells were unsuccessful, exhibited by a lack of cDNA yield post-reverse transcription. While it could be hypothesised that mRNA, and subsequent cDNA amounts were too low to fall within the limit of detection of the tapestation module, it must be noted that an average signal from all 384 cells seeded into the plate was measured. Further steps, including tagmentation and enrichment PCR, library cleanup and quantification were therefore not carried out. Validation and optimisation experiments, including **(i)** increasing the number of PCR amplification cycles, **(ii)** increasing Dynabead incubation and bead release times, and **(iii)** increasing primer concentration and RT mixes, are currently underway, to determine optimal bead capture and reagent combinations for successful mRNA extraction and library preparation.

5.3.3.3 S3-ATAC-seq

Once gDNA had been separated from RNA using G&T-seq, samples were processed using S3-ATAC sequencing. In general, sci assays entail binding of two unique adapters (as forward and reverse primers) at the ends of the target molecule. However, given that successful binding of both primers onto opposing strands occurs only 50 % of the time, traditional sci assays are not always able to generate high quality libraries, with sparse overall read coverage (figure 5.4 A (i)).[171] This yield limitation was overcome by the introduction of single-adapter transposition, in S3-ATAC-seq, wherein only a forward primer sequence, notably containing a terminal uracil residue on the Tn5 mosaic, binds to the gDNA target (figure 5.4 A (ii)). By gap-filling using an uracil-intolerant polymerase, extension can be limited, ending at this terminal uracil residue. A second primer, containing a 3'-blocked locked nucleic acid (LNA) mosaic sequence, complementary to the copied Tn5 mosaic template sequence, and a 5' terminal overhang is then introduced and binds to the copied sequences. Not only does this provide a template for gDNA extension for library generation, but also ensures that all library fragments contain both forward and reverse adapter sequences. Efficiency is further improved over traditional sci-assays, by increasing the number of LNA primer extension rounds prior to indexing PCR. As such, final usable yields can be improved by up to two orders of magnitude per cell. The final tapestation trace of S3-ATAC-seq library reveals the presence of the desired nucleosome laddering bands, indicative of successful tagmentation and library preparation (figure 5.4 B). Four main peaks can be identified between 250 – 1000 bp, corresponding to nucleosome-free, mononucleosome, dinucleosome and multinucleated chromatin fragments (from smaller to larger bp fragments). Furthermore, all peaks have frequency ranging between 150-180 bp, the approximate length of DNA wrapped around each nucleosome.[185] However, no further data has been generated as samples are yet to be sequenced.

5.4 Conclusions

In an attempt to further probe the transcriptional dynamics of previously identified FA and sporadic cycling events and decipher how these states differ from *bona fide* awakening as a result of

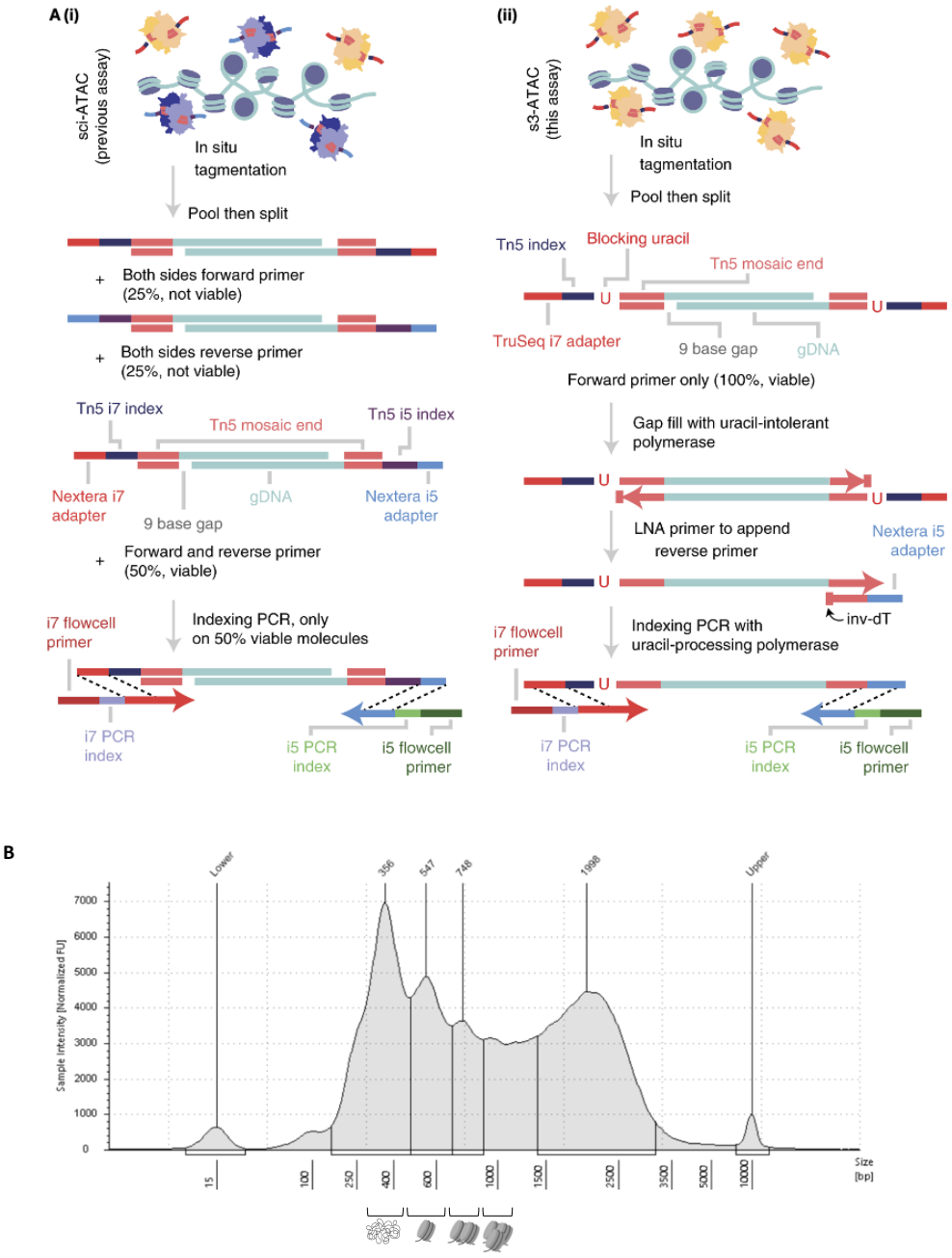


Figure 5.4 ***Working principles of S3-ATAC sequencing protocol and sample nucleosome laddering tapestation profile. A*** *Overview of S3-ATAC seq, highlighting key differences between precursor single-cell combinatorial indexing-ATAC-seq platforms. In both cases, Tn5 binds to the the chromatin, simultaneously fragmenting it and tagging it with unique adapters in the form of forward and reverse primers.* ***(i)*** *However, as both forward and reverse sequences must bind go the target fragment, which only occurs in 50 % of cases, even if indexing PCR and library generation were to occur with 100 % efficiency, only 50 % of starting chromatin will contribute to the overall library yield.* ***(ii)*** *By introducing a single-adapter transposition step, wherein only a forward primer sequence, with a Tn5 mosaic end sequence with a terminal uracil residue and a reaction-specific barcode, binds to the target. Inclusion of this uracil, and gap-filling using a uracil-intolerant polymerase, limits extension. A second primer, containing a 3' blocked locked nucleic acid mosaic then binds to the newly copied target sequence, and acts as a template for further extension for library generation. Using this technique, 100 % forward and reverse primer binding can be obtained. Figure adapted with permission from R. Mulqueen et al., 2021.*[171] *Copyright (2021) Spinger Nature.* ***B*** *Final tapestation profile of untreated gem21-MCF-7 cells (day 0), showing successful tagmentation, as exhibited by nucleosome laddering. All peaks span approximately 150-180 bp, consistent with length of DNA wrapped around the core histone octamer.*[203] *From left to right, the main peaks correspond to nucleosome-free, mononucleosome, dinucleosome and multinucleated chromatin fragments.*[186]

resistance to therapy, a long-term experiment, using a reporter cell line, developed by Hannah Dewhurst (Magnani Lab), was established.[2] Coupled with live imaging for the duration of the experiment, key sporadic cycling events could be identified, as an increase in mCh (and geminin) expression during dormancy, distinct from the dormant

population of GFP+ cells. Transcriptomic analysis was initially to be conducted fully with the Cellenone X1 system, using their 3'-RNA-seq compatible, pre-loaded cellenCHIPs at key timepoints during adaptation. However, due to various technical issues with the software and issues with cDNA synthesis, despite successful collection of cells at these timepoints, it was not possible to isolate individual populations. Due to the kit being discontinued, experiments were not re-attempted in this manner. As such, a multi-omic pipeline was developed in collaboration with Dr Max Mossner. This incorporated use of the Cellenone X1 for cell dispensing, to facilitate the extraction of information regarding both transcriptomic states of cells and chromatin architecture by combining G&T-seq, FLASH-seq and S3-ATAC-seq. Attempts at generating a FLASH-seq library were unsuccessful as exhibited by a lack of cDNA yield. Currently, work is underway to optimise the protocol, to ensure efficient cDNA synthesis and subsequent library preparation using untreated MCF-7 cells. Addition of S3-ATAC-seq into the analytical pipeline further expands the breadth of knowledge that can be extracted, enabling more detailed characterisation of various dynamic cell states. The unique gap-filling mechanism of S3-ATAC-seq, ensures 100 % primer binding efficiency, especially advantageous when working at a single cell scale. While initial library preparation was successful, the sample must first be sequenced and analysed to determine whether this is a reliable technique to implement across key experimental timepoints. Despite not being covered in the scope of this thesis, the phenomenon of FA warrants further work to validate cell size and changes in cell volume as cells progress through the cell cycle, to confirm whether there is in fact chromosomal duplication and thus determine whether the cells have indeed stalled in G2, or whether other complex mechanisms are responsible for this.

Ultimately, being able to understand the transcriptomics and epigenetic mechanisms that drive dormant cancer cells to awaken and evade therapy will hopefully uncover a vast array of targets whose activity could be modulated, perhaps even using combinations of existing small molecules, to prevent relapse. Understanding similarities and differences across single cells within the same cell states (as identified by GFP or mCh expression), as well as across different cell states (GFP+ *vs* mCh+ *vs* GFP+mCh+) would also facilitate the identification of transitions that could drive progression through the cell cycle and awakening. Although further optimisation is still required to increase the sensitivity and reliability of this technique, the preliminary work conducted as part of this thesis presents a promising foundation for dissecting the dynamics of dormancy at single-cell resolution. Integrating FLASH-seq and S3-ATAC-seq into a unified multi-omic pipeline greatly expands the potential applications of this platform, enabling the simultaneous characterisation of transcriptional programs and chromatin accessibility associated with sporadic cycling dynamics and *bona fide* awakening. Moving forward, longitudinal single-cell profiling during dormancy exit, coupled with functional perturbations using small-molecule inhibitors or CRISPR-based epigenome editing, will be critical to validate candidate regulators implicated in awakening and relapse. Ultimately, this work could provide the basis for therapies designed not only to target proliferating cancer cells but also to eliminate or permanently silence the dynamic dormant reservoirs that drive disease recurrence.

CHAPTER 6
EXPLORING THE INTERSECTION BETWEEN AGEING, OESTROGEN EXPRESSION AND CANCER

6.1 Introduction

6.1.1 Age and ERα expression

Despite being the most common subtype of breast cancer, ER+ breast cancers remain understudied. The complexity of this disease model can be underscored by considering the role of ERs in normal breast development, effect of age and menopausal status on incidence and disease outcome, and hormonal fluctuations with the menstrual cycle and menopause, in women. While substantial efforts have been directed toward understanding the molecular mechanisms underlying ER+ breast cancer progression and resistance to ET, considerably less attention has been dedicated to understanding the variations in ER expression in healthy breast, with age, highlighting the need for further exploration. Preliminary findings from spatial transcriptomics studies conducted alongside collaborators revealed that ER+ cells constitute a significant portion of breast tissue in healthy post-menopausal women (data not published). Interestingly, these were found to possess a transcriptional profile that partially overlaps with that of dormant ER+ breast cancer cells. However, rather than being overtly proliferative or pre-malignant, their profile adopts more of a 'retirement-like' phenotype, suggesting that adaptation to age adopts a similar trajectory to how ER+ cancer cells adapt to ETs.

While ideally a dynamic approach would have been adopted to explore variations in ER+ cell states between pre- and post-menopausal individuals, maintaining ER+ expression *in vitro* has proven extremely difficult, with only few studies highlighting success in doing so in patient-derived xenograft organoids thus far.[187] Until recently, most *in vivo* models to study the molecular mechanisms driving the acquisition of resistance to ETs in ER+ BC have also proven inadequate, as, despite the global similarities in hormone regulation between humans and mice, most carcinomas are unable to recapitulate this ER+ status, even with supplemental E2.[188,189] This has since been addressed by the development of a mouse intraductal (MIND) model, which presents a xenografting technique, wherein hormone-receptor positive breast cancer cells injected into the milk ducts of mice, are able to retain their ER status without hormone supplementation.[205] Notably E2 levels in the host mice are comparable to that seen in post-menopausal women, allowing this model to mimic *in vivo* clinical progression, with the ability to develop dormant micrometastatic lesions. However, study of effects in conditions where E2 concentrations are comparable to pre-menopausal women cannot be conducted using this model, without supplementing with additional E2. As such, a more static approach was adopted, wherein formalin-fixed paraffin-embedded (FFPE) samples obtained from a small cohort of pre- and post-menopausal individuals, containing both tumour and adjacent healthy tissue were processed using a small-scale spatial transcriptomic approach to obtain a snapshot overview of the cell states without the need to mimic native conditions. This approach enabled a high-resolution snapshot of ER+ cell states within their native tissue context, without the need to replicate complex hormonal dynamics in vitro or in vivo.

6.1.3 TNM Staging

In breast cancer, the most commonly used staging system is the tumour, node, metastasis (TNM) system.[190,191] This system provides a standardised method for categorising the extent and severity of disease, an overview of which is shown in figure 6.1.

- **T (tumour)** refers to the size and extent of the primary tumour. Tumour size is graded numerically, ranging from 0 to 4, with size of primary tumour increasing with magnitude. Each numerical category can be suffixed with either the letters a, b or c, which provide further detail on tumour dimensions.

- **N (node)** denotes the regional lymph node involvement. Nodal status is classified numerically (N0-N3), with each stage accompanied by detailed criteria based on the number, size and location of affected lymph nodes.
- **M (metastasis)** describes the presence or absence of distant metastatic spread. This is reported in a binary fashion as either M0, no metastasis, or M1, indicating the presence of metastatic lesions.

In all cases, where a particular component of the TNM system cannot be accurately assessed, it is denoted as TX, NX, or MX, indicating that the tumour, node, or metastasis status is unmeasurable, Generally, when this is the case, the metric is omitted from the TNM descriptor. Furthermore, all of the metrics can be prefixed with either the letters 'c' or 'p', referring to how the metric was determined: clinically, where the classification is assigned based on what is known prior to surgery, or pathologically, where the classification is assigned upon examining the cells after surgical excision. In the panel available for this study, all samples were classified pathologically.

Cancers can also be classified according to grade. The grade describes cancers in terms of how abnormal the constituting tumour cells look and how aggressively they are likely to behave.[192] Most grading systems use a three-tiered scale:

- **Grade 1 (low grade)**: tumour cells resemble normal breast cells and tend to frow slowly.
- **Grade 2 (intermediate grade)**: Cells start to behave abnormally, starting to grow faster and showing abnormal morphology.
- **Grade 3 (high grade)**: Grade 2 characteristics are further magnified, with cells growing very quickly and cells looking very abnormal.

While tumour grade provides valuable prognostic information, it is not directly equivalent to TNM staging. A high-grade tumour may present at an early TNM stage and vice versa, depending on the tumour's size, lymphatic spread, and metastatic status. Therefore, although tumour grade was considered during sample selection, it was not prioritised as the primary inclusion criterion in this study, in favour of TNM staging, which offers a more comprehensive representation of disease extent and progression.

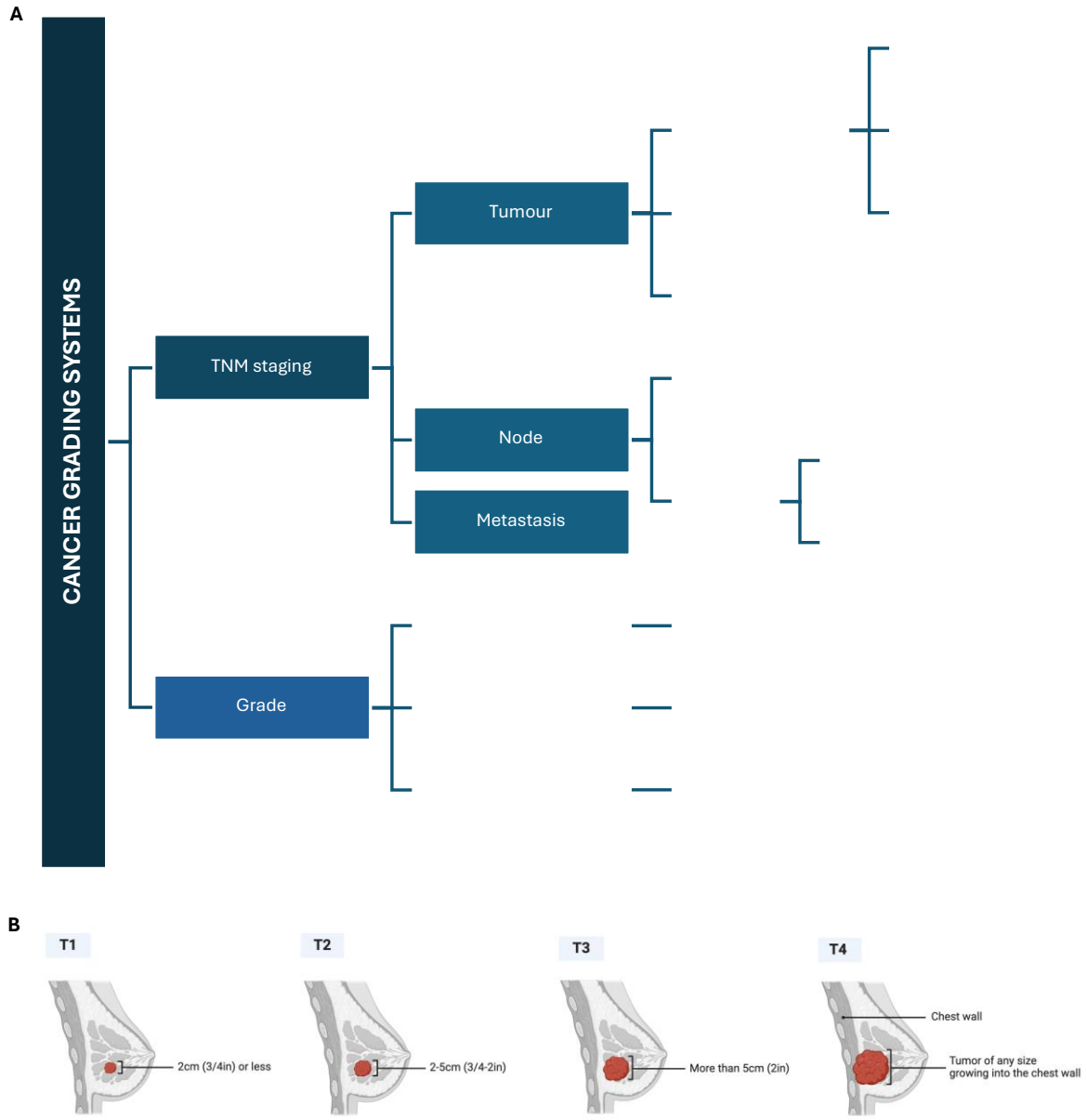


Figure 6.1 ***Overview of the two cancer grading systems: TNM staging and grade.***[190,191] ***A*** *Flowchart outlining characteristics of TNM staging and grade for samples included in the panel. Note that this is not a complete list and is missing all categories and subcategories not relevant to the scope of this study.* ***B*** *Diagram showing relative size of tumour in the breast, at the four TNM stages. T4, not mentioned in the flowchart, involves spread of tumour of any size into the chest wall. Figure adapted from Biorender template.*

The likelihood of developing luminal A breast cancer increases with age, and while menopause has been found to halt this rising risk, it does not reverse it [193]. To further explore why this is the case, and to examine how variations in certain genes may regulate ER expression with age, a HiPlex RNAscope assay was conducted. This assay aimed to predominantly assess differences in *ESR1* expression, alongside select ER-responsive genes (outlined in table 6.1), across tumour and adjacent healthy regions. The primary objective was to determine whether changes in ER expression with age occur in a uniform, tissue-wide manner or reflect inherent cellular heterogeneity, whereby specific subpopulations may retain or lose ER expression variably over time. While cutting-edge spatial transcriptomic platforms such as Nanostring's CosMx are capable of profiling up to 6,000 genes at subcellular resolution, RNAscope HiPlex was selected as the method of choice in this study for both technical and biological reasons. Firstly, although CosMx offers high multiplexing capabilities, it is highly sensitive to RNA integrity and

tissue morphology, requiring freshly cut FFPE slides, ideally processed within 1–2 weeks.[194] Unfortunately, the slides used in this study were prepared months in advance from archival paraffin blocks. Even when stored at 4 °C, tissue sections begin to experience RNA degradation within days to weeks due to exposure to oxidation, humidity, and air, making them unsuitable for CosMx analysis and potentially compromising transcriptomic accuracy. In contrast, RNAScope is significantly more robust to RNA degradation, leveraging short (approximately 20 nt) double Z probes to facilitate the detection oof transcripts, even when RNA is partially degraded, and is therefore better suited for older FFPE samples.[195] Moreover, RNAscope offers exceptional sensitivity and specificity, particularly for low-abundance transcripts that are often missed by broader transcriptomic approaches like CosMx. This becomes particularly important when studying ER-expression in post-menopausal or dormant tissues, where expression of ER-responsive genes may fall below the detection thresholds of existing protein assays, such as immunohistochemistry and Akoya's Phenocycler, primarily due to limited antibody sensitivity.[196] While Akoya's Phenocycler can be considered a suitable alternative to RNAScope due to its multiplexing capabilities, it is strongly limited by its decreased sensitivity and incompatibility with archival slides. Additionally, as RNAScope detects mRNA rather than proteins, it permits the detection of early transcriptional changes before they can be detected at a protein level – a notable advantage when aiming to resolve subtle variations in gene activity which may regulate key stages in tumour dormancy and adaptation in hormone-depleted environments.

6.2 Materials & Methods

Variation in ERα expression in tumour and healthy tissues with age, were probed using RNAScope. ACDBio's proprietary HiPlex RNAScope (v2) assay enables the specific detection of up to 48 different targets, including a pre-defined panel of 12 housekeeping genes.[197] Utilising the same double-Z probe technology as in their Multiplex v2 assay (section **2.3.5**), iterative hybridisation and fluorophore cleavage allows for smaller panels of up to four targets, including but not limited to short RNA targets and miRNAs, to be built up to generate spatially defined libraries in highly heterogenous samples. Images can be superimposed using the RNAscope HiPlex Image Registration Software v2.1 to enable visualisation of all or select genes (as distinct dots), facilitating data interpretation.

Given the novelty of the technique, there is still a limited library of target probes available for this assay which restricted the scope of this study. This was further complicated by the necessity for different targets to populate unique channels to enable this HiPlex detection, which was not always possible. As such, the target library had to be amended to ensure that all the chosen targets populated different channels, and that those chosen were relevant to the study.

6.2.1 Samples

5 µm FFPE slices from 40 archival breast cancer tissue samples, from patients with untreated luminal breast cancer between the ages of 30 and 89, were kindly provided by Dr Andrea Vingiani (Instituto Nazionale de Tumori, Milano). The sample panel was carefully selected to ensure that all excised slices contained a portion of cancerous tissue and healthy tissue at its periphery. These included ten samples per decade between 30-49 years and five samples per decade between 50-89 years of age. All samples were provided with annotations including information on patient age, stage and grade of cancer, menopausal status, number of pregnancies, number of abortions, date of last period (if applicable), duration of breastfeeding (if known) and oral contraceptive use. A

summary of the full cohort is shown in supplementary **6.1**. Corresponding hematoxylin and eosin (H&E)-stained slides were also provided for all FFPE samples to facilitate the differentiation of healthy and cancerous tissue.

6.2.2 Sample Selection

One sample with TNM classification pT1cpN0, i.e. tumour size ≤ 2 cm with no spread to lymph nodes and no metastases, and grade 1, was selected for each age group for analysis. Where this was not possible (age groups: 50-59, 60-60 and 80-89), the next closest grading was chosen within the same TNM classification. A summary of selected samples is shown in figure 6.2, which, from hereon, will be referred to by age of patient.

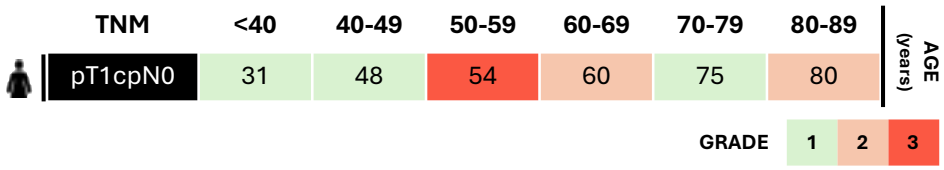


***Figure 6.2* Overview of samples selected for RNAScope analysis of ERα expression.** *Samples were chosen according to their TNM classification – pT1cpN0 – and grading – grade 1. This grade was not available for all age groups, so samples with the most similar classification was chosen. Patient age and tumour grade are shown as numbers and colours, respectively. Grade 1 cancers are shown in green, grade 2 in orange and grade 3 in red.*

6.2.3 Target Selection

Gene targets were selected based on findings from a recent study by Rosano *et al.*, exploring how individual lineages adapt to therapy in oestrogen-receptor breast cancer as well as findings from previous spatial transcriptomics analyses conducted on four women (two pre- and two post-menopausal) using pan-cytokeratin+ markers (data not shown).[2] Given the aim to monitor ER expression with age and cancer, oestrogen responsive genes, and genes exhibiting significant variation during their evolutionary trajectory were selected as part of this panel. Targets of interest therefore included the following: *ESR1, FOXA1, S100A1, MTOR, CDKN1A, CD36, PROM1, MKI67, CD44, ELF5, MYC, CCNE1*.

6.2.4 Hi-plex RNAScope Assay

6.2.4.1 Deparaffinisation

The selected slides were inserted into a Tissue-Tek slide rack and placed inside a dry oven (D06058 200, Memmert) pre-heated to 60 °C for one hour, to melt the paraffin. Once melted, the slide rack was placed into a Tissue-Tek Clearing Agent dish filled with xylene (Sigma-Aldrich) and incubated for five minutes at room temperature, with occasional agitation. The slide rack was transferred to a second xylene-containing dish and incubated for a further five minutes. Slides were then incubated in 100% ethanol for two minutes at room temperature before transfer to another dish containing 100% ethanol for a further two minutes, both with slight agitation. Slides were tapped to remove any residual liquid and returned to the 60 °C dry oven for five minutes to dry completely.

6.2.4.2 Pretreatment

Once slides were dried, RNAScope Hydrogen Peroxide was applied to each slide, making sure to cover the entire section, and incubated at room temperature for ten minutes. Slide were tapped to remove any excess reagent and placed into a Tissue-Tek slide rack submerged in distilled water. Slides were washed by moving the rack up and down approximately five times. This was repeated once more with fresh distilled water.

6.2.4.3 Manual Target Retrieval

1X target retrieval buffer was heated in a microwave (NN-SD27HS, Panasonic) on high power for 5 minutes and then at 30 second intervals, until the temperature reached 98 °C. The buffer was then transferred to a hot plate to maintain the temperature at 98 °C for the duration of target retrieval. The slide rack was placed in the hot target retrieval buffer, covered with foil, and maintained at temperature for 15 minutes. The slide rack was removed, immediately transferred to a beaker containing distilled water and washed by moving the rack up and down. A second wash was conducted, this time in 100% ethanol, allowing them to air dry. A hydrophobic barrier was drawn as close to the boundaries of the tissue sections as possible, using an Immedge hydrophobic barrier pen (H-4000, Vector Labs).

6.2.4.4 Hiplex Assay

Given the imaging capabilities of the microscope used for imaging, slides were processed using the RNAScope™ Hiplex12 (488, 550, 650) Reagent Kit, wherein each different channel is referred to as a tail (T). Unless otherwise specified, all steps outlined below were followed by two wash steps in 1X Wash Buffer (ACDBio) with gentle agitation, and all incubations were conducted in a HybEZ™ II Oven (ACDBio) at 40 °C. When working with a positive control panel, housekeeper genes are supplied pre-mixed (RNAscope HiPlex12 Positive Control Probe – Hs; ACDBio) and can be applied directly to slides. However, as probes are purchased individually at their 50X stock concentrations for each tail for the target panel, probes must first be diluted and combined to form a 1X mix, in RNAscope HiPlex probe diluent, before application to slides. Tails to which probes correspond are summarised in table 6.1.

Initially, a few drops of positive control probe mix were applied to each slide, and incubated for 2 hours. Targets were amplified in three rounds using AMP 1, AMP2 and AMP 3 reagents, respectively. In each case, a few drops of AMP reagent were applied to each slide and incubated for 30 minutes. An initial set of experiments was then conducted using the RNAScope HiPlex FFPE Reagent to determine effects on minimising autofluorescence. When used, a few drops of a 2.5% FFPE reagent dilution (in 4X SSC (Invitrogen)) was applied to slides right before hybridisation of Fluoro reagents. Slides were incubated in the dark for 30 minutes, at room temperature. Following this, enough Fluoro T1-3 reagent was applied to cover each section and incubated for 15 minutes. As this protocol entailed iterative fluorophore stripping and imaging of three probes at a time, only three fluorophores were used. T1, 4, 7 and 10 were labelled with Alexa Fluor 488; T2, 5, 8 and 11 were labelled with Dylight 550 and T3, 6, 9 and 12 were labelled with Dylight 650. Samples were counterstained using a few drops of DAPI and allowed to incubate for 30 seconds at room temperature before carefully flicking off excess liquid. Approximately 30 µL Prolong Gold Antifade Mountant (Thermo Fisher Scientific) was pipetted within the hydrophobic boundaries and a coverslip was placed on top, being careful not to trap any air bubbles. Mountant was left to cure in the dark for a minimum of an hour, before imaging the same day.

Once slides hybridised with T1-3 had been imaged, coverslips were removed by soaking in 4X SSC at room temperature until coverslips either fell off or could be removed easily. 10% cleaving solution (prepared in

4X SSC) was applied to cover each section and allowed to incubate for 15 minutes at room temperature. Slides were washed in PBS-T (0.5% Tween, Sigma-Aldrich), before a second round of cleaving. Optionally, a few drops of 2.5% FFPE reagent was applied to slides before repeating Fluoro reagent hybridisation with Fluoro T4-6 reagent. Slides were counterstained, mounted and imaged on the same day. This cleaving and fluorophore hybridisation was repeated until all tails had been imaged.

If no FFPE reagent was used, slides were prepared for another round of probe hybridisation and imaging. After all tails (T1-12) from the positive control panel had been imaged, coverslips were removed and fluorophores were cleaved by application of 10% cleaving solution and incubating at room temperature for 15 minutes. Slides were washed briefly with PBS-T (0.5% Tween). All remaining probes and amplifiers were stripped from the samples using the RNAscope® HiPlexUp Reagent. Few drops of HiPlexUp reagent were applied to completely cover each section and incubated for 5 minutes at room temperature. This was repeated two more times, before washing slides in 1X wash buffer. The whole HiPlex Assay protocol was then repeated, this time using the target probes.

Table 6.1 Genes assigned to each tail, for both the positive control probe mix and targets of interest. Fluoro reagent used for each set of three genes is specified. No probe was assigned to T6 in the target panel.

	Fluoro Reagent	Gene		
Housekeeping Genes	T1-3	POLR2A	PPIB	UBC
	T4-6	HPRT	ACTB	SDHA
	T7-9	TFRC	LDHA	GAPDH
	T10-12	RPL5	YWHAZ	RPL28
Targets of Interest	T1-3	ESR1	FOXA1	S100A1
	T4-6	MTOR	CDKN1A	-
	T7-9	CD36	PROM1	MKI67
	T10-12	CD44	ELF5	MYC

6.2.4.6 Imaging

All images were acquired using an inverted spinning disk confocal microscope, Zeiss Axio Observer Z1 equipped with a Mariana's CSU-W spinning disk unit (Intelligent Imaging Innovations (3i)), at the Light Microscopy Facility at the Institute of Cancer Research, Chelsea. The system is equipped with six solid state lasers and an Orca C13440 Flash 4.0 Digital Camera (Hamamatsu). With 50 µm pinholes, the CSU-W1 spinning disk unit enables capture of high quality images very quickly (<200 fps). [198] The system is also equipped with an ADEPT Piezo External Input (Applied Scientific Instrumentation) which enables smooth transitions between z-coordinates, especially beneficial when acquiring small z-stacks.

Using the 'vector' version of Slidebook 6 (2023, 3i), cells were initially located through the eyepiece, using the 'Ocular' setting with the DAPI filter at 20x magnification using the default settings of the 'Focus' window. Once focused, the sample was viewed using the FITC and Cy3 filters to confirm that the transcripts were visible and in the focal plane. The viewing modality was then changed to 'ConfocalFLASH4', for confocal visualisation. Again, all channels were viewed individually and focus adjusted as necessary. Once ready to acquire, the capture settings, including exposure time for the selected channels, were modified in the 'Capture' window. For static 2D images, the following filter settings were used: filter set = ConfocalFLASH4.0; filters selected = C405 (DAPI), C488 (Opal520), C561 (Opal570); exposure times = C405 (200 ms), C488 (200 ms), C561 (100 ms) and laser power = 100 %.

For all samples, a 40x image was acquired, as a 3D 1x1 z-stack, a 2x2 z-stack or a 5x5 z-stack, depending on cell dispersion in the chamber slides. Z-stacks for all multiplex images comprised 3 planes, with a 0.6 µm step size, to cover a total depth of 1.2 µm. Z-stacks were set up in the 'Z' tab of the 'Focus' window. The

most focused plane was set as the central reference and upper and lower bounds were set to 2 µm either side of the focal plane. Capture type was set to '3D' in the 'Capture' window prior to acquisition.

6.2.5 Image Processing

6.2.5.1 Image Deconvolution

All images were deconvolved, using the near neighbour deconvolution method, to improve signal resolution using the Slidebook software and saved as .sld files prior to further analysis. Deconvolution parameters used are as follows: µ/pixel = 0.162; numerical aperture = 1.30; working distance = 210; index of refraction = 1.518; spacing (µm) = 0.6; subtraction constant = 0.9 and edge padding = X - 64, Y – 64.

6.2.5.2 Analysis: FIJI & QuPath

Output .sld files were opened in FIJI (Fiji is Just ImageJ) using the Bio-Formats plugin. Filed were viewed as 'Hyperstack' with the 'Default' colour mode and autoscaled. For all images acquired as z-stacks, a maximum intensity projection was generated prior to processing. All images were processed individually, according to the following steps: (i) images were split into the three different channels – DAPI = C1, Opal 520 = C2, Opal 570 = C3; (ii) lookup tables and brightness and contrast were adjusted for each individual channel - C1 = blue, C2 = green, C3 = red – such that distinct nuclei could be observed and transcripts were visible as distinct punctate dots with no background. Images were saves as composite TIFFs, with all channels in one image. These were then opened in QuPath and processed further as outlined in section **2.3.5.8**. Again, no adjustments were made to the pre-defined 'cell detection' thresholds. Specific detection thresholds set for 'subcellular spot detection' are outlined in supplementary **6.3**. All detectable 'dots' were 'split by intensity'. For the purpose of this thesis, transcript counts for only *RPL28*, *ESR1* and *FOXA1* were quantified.

6.3 Results & Discission

6.3.1 Determining optimal assay conditions and imaging parameters

Initially, two samples were processed using just the housekeeping genes and the HiPlex FFPE reagent. ACDBio recommends application of a 2.5 – 5 % HiPlex FFPE solution to increase the signal-to-noise ratio (SNR) when working with samples with high tissue autofluorescence, occurring from the presence of endogenous fluorophore within cells. This is notably problematic when working with single-cell assay, such as RNAScope, where detection involves quantification of punctate dots, whose signal may be obscured by strong background signals. Furthermore, given that breast tissue autofluorescence has a maximum emission wavelength of 526 nm, with large spectral overlap with AF488 (used to visualise tails 1, 4, 7 and 10) which also has a peak emission wavelength of 525 nm, any effort to minimise tissue autofluorescence would be advantageous.[199,200] While use of the FFPE reagent did marginally decrease the autofluorescence visible in the GFP channel (data not shown), it did not seem to yield significant enhancement in signal intensity or reduction in noise for it to be worth continuing. If this were to be used, given its incompatibility with the RNAScope HiPlex reagent, gene expression per cell would be limited to either the 11 housekeeping genes or 12 target probes, limiting the analytical power of this technique. Nonetheless, these samples served as a good test to ensure (i) efficient fluorophore cleavage between imaging rounds and (ii) determined whether the same region could be imaged iteratively, as different targets were visualised, prior to conducting the assay on a larger sample cohort.

Cleavage efficiency was verified by imaging all channels after imaging T10-12 (figure 6.3 A (i)), and doing a final fluorophore strip. As seen in figure 6.3 A (ii), minimal signal was detected in all channels, aside from 405 nm (DAPI), highlighting that there is no signal bleed through and efficient stripping of fluorophores between imaging rounds. The signal that can be observed does not seem to be localized to any cells in particular, and presumably stems from autofluorescence from surrounding fatty tissue. Combined, figure 6.3 A highlights the difficult in imaging the same region iteratively. This is also further identified in figure 6.3 B, wherein, despite attempting to match stage coordinates, only 3 rounds of imaging could be superimposed. As the goal is to ultimately be able to quantify the 24

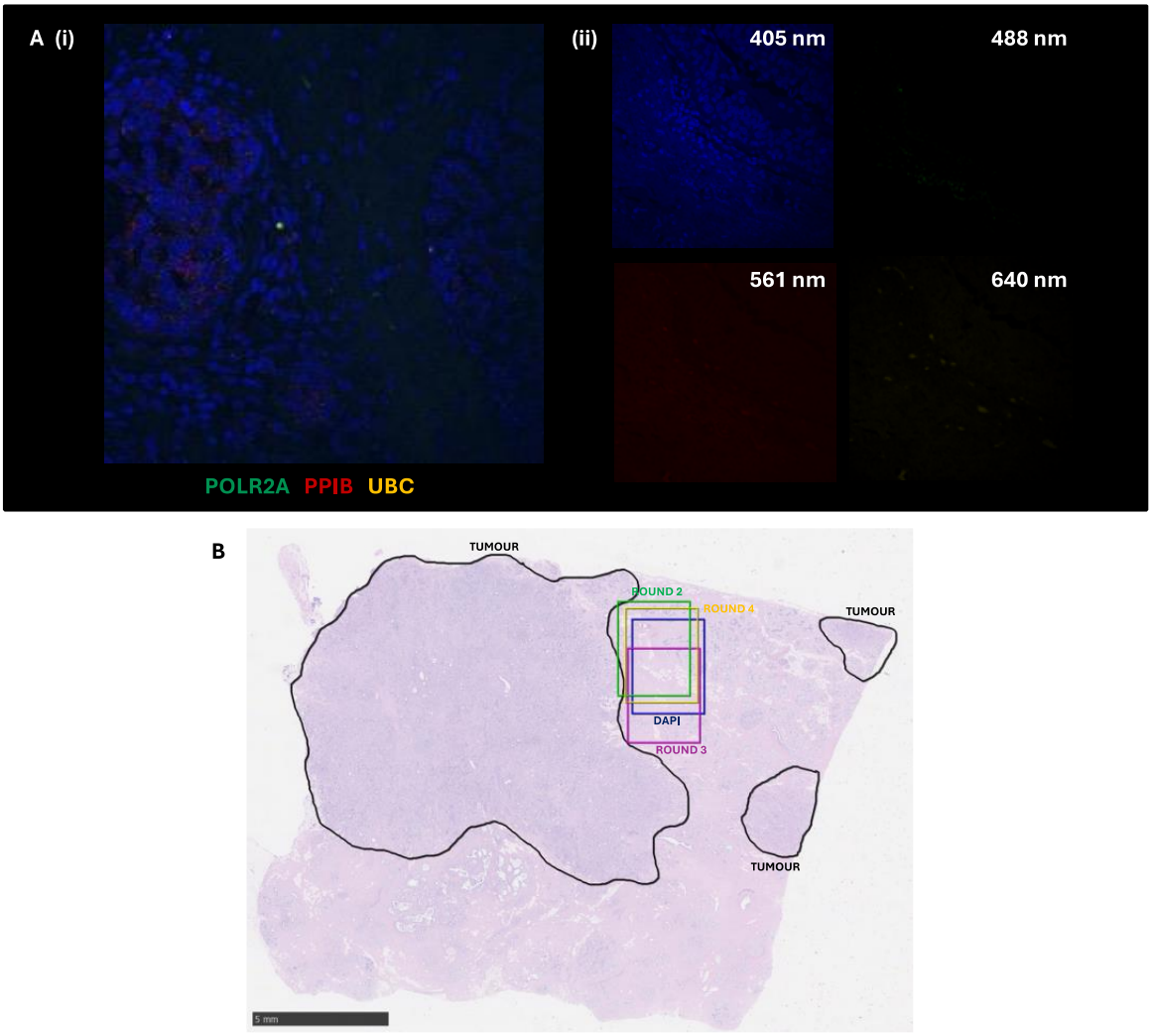


***Figure 6.3** Validating cleavage efficiency between iterative imaging rounds and determining optimal imaging parameters. A Fluorophores were stripped from samples after **(i)** imaging T10-12 (shown on left). **(ii)** While there was initially difficult in iteratively imaging the same region, imaging post-fluorophore stripping shows very little signal in any of the target channels (488 nm, 561 nm and 640 nm). The signals that can be identified are native to the tissue surrounding the cells, highlighting that while the tissue does exhibit autofluorescence despite treatment with the FFPE reagent, fluorophore stripping is efficient. B Initially, attempts were made to image both healthy and tumour region in the capture. However, as seen here, not only was this difficult given the non-uniform tumour boundary, which was quite difficult to identify using solely a DAPI stain at a 40x magnification but overlap between imaging rounds was extremely poor.*

transcripts within the same cells, after multiple fluorophore stripping and rehybridization steps, imaging the same region is extremely important.

6.3.2 Optimised region identification and image processing

To ensure the same region of tissue was imaged during iterative imaging cycles with different probes, the following method was implemented. Initially, from the H&E-stained slides, distinct regions with unique and easily identifiable tissue morphology in the tumour and healthy tissue of the samples were chosen (supplementary **6.1**). These were then mapped onto the DAPI-only images to verify ease of identifiability. Once selected, processed HiPlex slides were positioned onto the stage of the Spinning Disk Confocal microscope to be used for final imaging, and agreed upon regions were identified using the DAPI filter (figure 6.4 A). For ease of identification in subsequent imaging rounds, the location was also marked using a permanent marker on the back of the slide. Images were initially acquired at both a 20x and 40x objective. Initial overall resolution was found to be superior using a 20x objective, however, much more distinct signals were observed at 40x, although some transcripts were out of the plane of focus. Consequently, z-stacks were acquired at 40x. Performing a maximum projection of the z-stacks and comparing them with the images acquired with a 20x objective revealed a significantly greater level of detail (data not shown). Acquiring a stack beyond 1.2 µm provided minimal additional value to the image details. All further imaging iterations were therefore acquired at a 40x objective with a 1.2 µm z-stack.

Prior to quantifying the number of transcripts per cell, images stacks were deconvolved, using the nearest neighbour deconvolution (NND) algorithm (figure 6.4 B). Introduced in 1979 by Castleman, NND simplifies 3D-images into a series of 2D sections, with each focal plane treated as its own independent image.[201,202] This algorithm assumes that all neighbouring image planes are a blurred, out-of-focus transformation of the central plane. Therefore, by subtracting a fraction of this additional light causing this blur and reassigning the light back to the assumed original plane, the image resolution and signal is significantly increased. This method is particularly useful when working with complex tissues where cells of different types or states are located close to each other. By improving the resolution of the data, NND can reveal distinct cellular states and unique changes in gene expression that are fundamental to enhancing our understanding of cellular behaviour and molecular behaviour. Once deconvolved, a single maximum z-projection image was generated for each region and imaging round by collapsing the individual z-slices were collapsed onto the same plane. Image registration was attempted, to superimpose the relevant channels onto the same image and visualise transcript expression from the exact same cell, but due to issues in the registration software, this was not possibly – an example of a successfully registered image is shown in figure 6.4 C. Transcripts were quantified using QuPath's subcellular detection algorithm. However, as the FFPE reagent was not used in the optimised HiPlex assay due to incompatibility when used in conjunction with the HiPlexUp reagent, regions of high autofluorescence were identified. This is especially evident in patient 54's healthy tissue (figure 6.4 A), where clusters of orange signal appear as signels from the green and red channels overlap. Fortunately, QuPath's annotation capabilities allow for exclusion of these high-autofluorescence regions by modifying annotation boundaries, cell and subcellular spot detection can be limited to solely the regions of interest.

With the optimised imaging methodology enabling efficient sequential imaging of the same region, coupled with the ability to strip fluorophores and probes from tissue samples using the HiPlexUp cleaving reagent, all probe imaging can now be conducted on the same region of the same slide. Conducting multiple rounds of imaging and performing different assays on a single sample, rather than on multiple samples in parallel, allows for direct comparison of the same regions by superimposing images and subsequently provides a more comprehensive view of gene and receptor expression.

6.3.3 Quantification of ESR1, FOXA1 and RPL28 expression in luminal breast cancer samples

For the purpose of this thesis, RPL28 was chosen as the housekeeping gene of choice. RNA sequencing data from a recent cancer discovery paper from our group, exploring the adaptation of hormone-dependent breast cancer to endocrine therapies, showed that, of the housekeeping gene panel used for the HiPlex assay, the least variation across replicates and treatment conditions was in *RPL28* (in terms of Z-scores; supplementary **6.4**). As this technique enables absolute quantification of transcripts (as a count of punctate dots), normalisation no housekeeping normalisation was conducted. Due to time constraints, the whole gene panel could not be analysed. Therefore, quantification of *ESR1* and *FOXA1*, whose expression is known to facilitate hormone responsiveness and helps maintain a luminal phenotype in ER+ breast cancer, were prioritised.[203] Number of dots in 150 cells in each sample, from both the tumour and healthy portions, were quantified for the three genes outlined above, using QuPath, using the subcellular spot detection parameters outlined in supplementary **6.3**. It is worth noting that while the number of dots within each cell can be quantified, there is no way to accurately assign the dot area to that of a single transcript and as such, will continue to be referred to as 'dots'. However, as all samples were analysed in the same way, this does not act as a limitation in this regard.

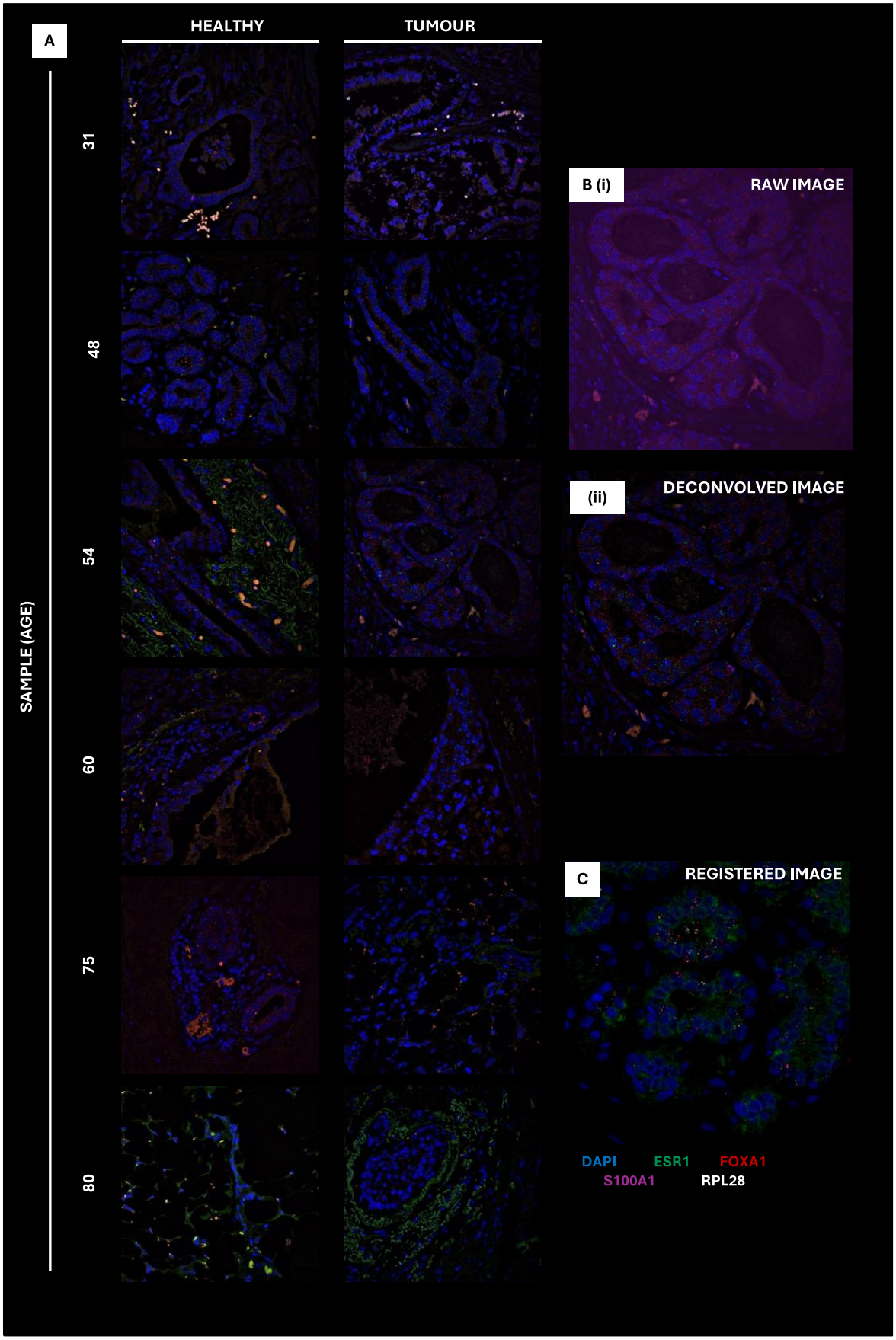

A
HEALTHY
TUMOUR
SAMPLE (AGE)
31
48
54
60
75
80
B (i)
RAW IMAGE
(ii)
DECONVOLVED IMAGE
C
REGISTERED IMAGE
DAPI
ESR1
FOXA1
S100A1
RPL28

Figure 6.4* Optimising the HiPlex RNAScope workflow. A** *Issues in imaging the same region iteratively, with different fluorescent probes, resulted in the generation of a library of easily identificable regions as imaging reference points. Examples of such, all with unique morphology, as shown for both healthy and tumour regions for all samples imaged. Colours of channels refer to the following: DAPI = blue, ESR1 = green, FOXA1 = red, S100A1 = pink.* **B** *Prior to analysis, images were deconvolved to improve the signal-to-noise ratio. An example of* ***(i) *an unaltered and* ***(ii)*** *a deconvolved image are shown. Channels are as in A. C Images can also be registered to overlay specific or all genes of interest onto the same image. Here, DAPI = blue, ESR1 = green, FOXA1 = red and RPL28 = white.*

6.3.3.1 ESR1 expression

Interestingly, of the 150 cells studied for each condition, there is very little variation in the number of nonzero cells for ESR1, both as patient age increases and between the healthy and tumour counterparts of the same patient (figure 6.5). This number ranges mostly between 130 to 145 non-zero cells. However, only 101 and 114 healthy cells were nonzero for patients 31 and 75, respectively, while only 45 % and 60 % of their corresponding tumour cells were nonzero. This large proportion of cells expressing no *ESR1* transcripts could possibly explain the clustering of the number of dots per cell towards the lower end of the spectrum, although in general, those that did express the gene did so in low quantities. Notably, since no ER staining was conducted, cells could not be assigned ER+ or ER- status, and as such, whether the proportion of cells containing no *ESR1* transcripts also contain no ER cannot be determined. Nonetheless, this technique allows binary determination of whether a cell expression *ESR1* or not. While analysis could be conducted upon omission of these non-zero cells, this would result in significant skew due to variations in the number of *ESR1*-containing cells and would further limit the sample pool. A much more significant difference in average *ESR1* expression was observed between the healthy and tumour tissue counterparts for patient 31 compared to all other age groups ($P \leq 0.001$) whilst the greatest variation in the number of transcripts per cell were observed in the tumour counterpart of patients 54 and 60. For patient 60, this heterogeneity in transcript expression across cells is also mimicked in their healthy counterparts, suggesting that *ESR1* expression may also vary significantly even within neighbouring cells with age. However, as these both fall within the post-menopausal age category, wherein there is a marked increase in ER expression, this elevated transcript expression may reflect age-related changes in ER expression and activity.[204] This linear increase in *ESR1* expression with age, can be seen to continue in patient 80, with patient 75 acting as an outlier. The variation in the number of dots per cell for the aforementioned age groups, between healthy and tumour tissue was minimal, with differences being statistically insignificant ($P > 0.05$). Furthermore, all tumour samples exhibited significantly greater *ESR1* expression than patient 31, which expressed the lowest average *ESR1* expression overall ($P \leq 0.0001$). This same trend was somewhat mirrored in the healthy counterparts, with the exception of patient 60 and 75, with patient 60 exhibiting

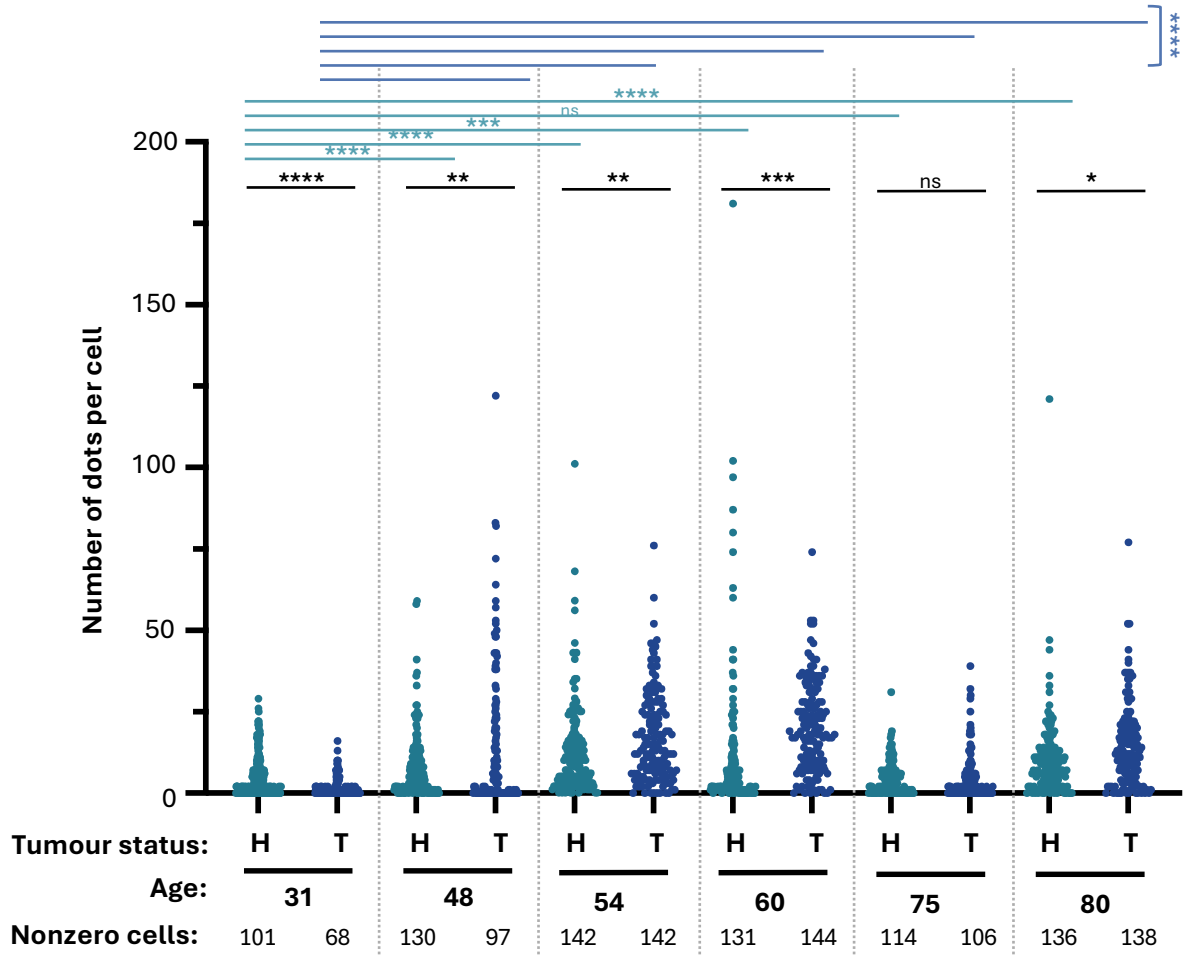


***Figure 6.5** Number of ESR1 dots counted per cell, for 150 cells per condition and age, from HiPlex RNAScope Assay. Box-and-whisker plot of the number of subcellular counts using QuPath. Data is split by age group, with counts from healthy (H; teal) and tumour (T; blue) regions, specified. Number of nonzero counts contributing towards the data plotted is shown under each plot. All results were analysed using a One-way ANOVA test in Excel, with respect to patient 31. Significant differences between data points are shown in black for comparison between the healthy and tumour counterpart of the same patient, in teal for all healthy cell counts and in blue for all tumour cell counts - ns = no significance, $P > 0.05$; * = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$; **** = $P \leq 0.0001$.*

a wide distribution in the number of transcripts per cell and patient 75 displaying a similar overall data shape to patient 31 and no statistically significant difference.

6.3.3.2 FOXA1 expression

Similar trends in transcript expression are observed for *FOXA1* as with *ESR1*, with significantly higher expression levels in tumour samples compared to healthy tissue across all age groups, with the exception of patient 75, wherein the mean *FOXA1* expression appears lower in its tumour counterpart, but with a broader spread of data (figure 6.6). An overall increase in *FOXA1* expression is observed across tumour samples, though sample 75 shows an anomalously low expression that seems to rebound in sample 80. Healthy samples also show a linear increase in *FOXA1* expression, albeit to a lesser extent than their tumour counterparts. However, there is substantial variability in the number of nonzero cells. The number of nonzero healthy cells generally ranges between 102 to 132, although approximately 50 % of the analysed healthy cells for patients 31 and 48 did not express any transcripts. The tumour counterpart also show variation in the number of cells expressing *FOXA1* transcripts, with

counts for patients 31 and 48 shifting significantly from the average range of 132 to 144 *FOXA1*-expressing cells. Interestingly, despite the lack of spread of datapoints for the tumour sample of patient 75 (with only 81 nonzero cells), nearly all of the 150 analysed cells for patients 60, 75 and 80, express *FOXA1* and is quite consistent with the number of cells expressing *ESR1*. Furthermore, the healthy tissue from patient 60 shows a broad spread in *FOXA1* expression, similar to *ESR1*, although levels remain lower overall than in the tumour counterpart. No significant differences in *FOXA1* expression are observed between healthy and tumour tissue in for patient 77, nor between the tumour regions of patients 31 and 70, again consistent with *ESR1* expression patterns. Minimal variation exists between healthy tissue samples of patients 31 and 75, while no significant difference can be identified between the healthy and tumour tissues of patient 31. In all other healthy samples, differences are statistically significant (at least $P \leq 0.001$). Among all tumour tissues, significant differences are observed overall when compared to patient 31, except for patient 75, which shows nearly identical data spread and characteristics.

This identified overlap in features between *FOXA1* and *ESR1* expression can be explained by *FOXA1*'s known role in modulating *ESR1* binding sites in ER+ breast cancer and its key role in mammary gland development, with 50% of *ESR1*-regulated genes requiring prior recruitment of *FOXA*.[203,205,206] Accordingly, similar or proportional expression levels of *FOXA1* and *ESR1* would be expected, and is observed here with ER+ cells displaying approximately half the number of *FOXA1* dots per cell compared to *ESR1*. Overall, the lower *FOXA1* expression in the healthy counterparts of all ages can be attributed to its role in ER transcriptional reprogramming, as high levels have been shown to induce a pro-metastatic secretome and subsequent metastastic progression in ET-resistant breast cancer. [224]

6.3.3.3 RPL28 expression

RPL28 was chosen as the housekeeping gene for analysis in this cohort due to its generally consistent expression across different treatment conditions, as outlined in a recent study conducted within our lab (supplementary **6.4**).[2] While ribosomal proteins have been implicated in the tumour suppressive pathway, by association with the tumour suppressor protein, p53, the role of *RPL28* breast cancer progression is unclear.[207] However, the role that *RPL28* plays in protein synthesis and cell

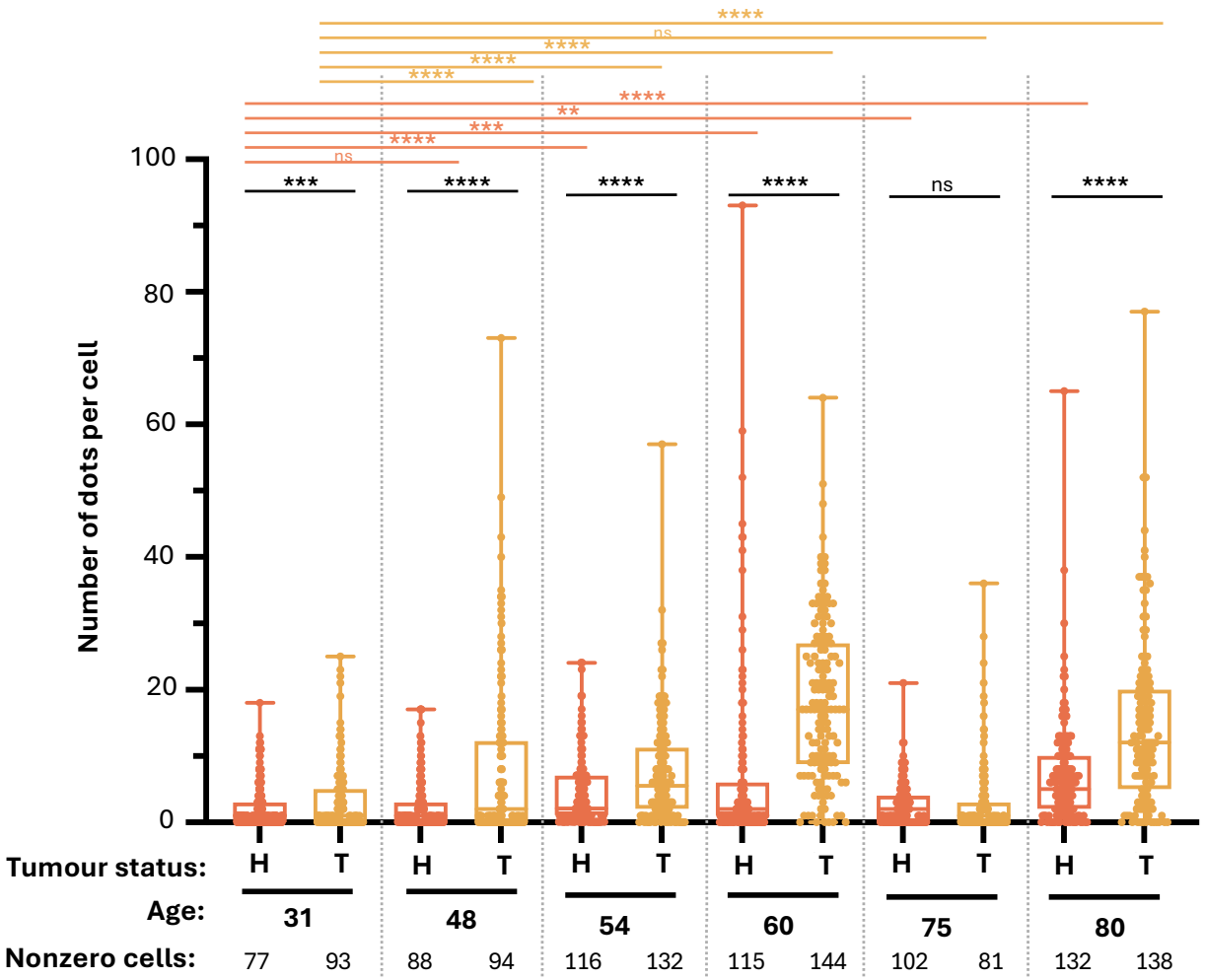


***Figure 6.6** Number of FOXA1 dots counted per cell, for 150 cells per condition and age, from HiPlex RNAScope Assay. Box-and-whisker plot of the number of subcellular counts using QuPath. Data is split by age group, with counts from healthy (H; orange) and tumour (T; yellow) regions, specified. Number of nonzero counts contributing towards the data plotted is shown under each plot. All results were analysed using a One-way ANOVA test in Excel, with respect to patient 31. Significant differences between data points are shown in black for comparison between the healthy and tumour counterpart of the same patient, in orange for all healthy cell counts and in yellow for all tumour cell counts - ns = no significance, $P > 0.05$; * = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$; **** = $P \leq 0.0001$.*

proliferation, and its ubiquitously expression in breast cancer, suggests a possible role in encouraging its proliferative characteristic and function as a proliferative biomarker.[208] Despite relatively stable expression on a bulk, and single-cell level, vast heterogeneity in transcript expression can be identified at a sub-cellular level, both between healthy and tumour tissue and between cells of the same condition (figure 6.7). This notably excludes the healthy counterparts of patients 48 and 80, which have comparable expression to patient 31, which does not align with the patterns observed for *ESR1* and *FOXA1*. Furthermore, almost all analysed cells express RPL28, albeit to varying degrees. Historically, choice of housekeeping genes are validated using techniques that cannot differentiate between expression in single cells, such as RT-qPCR. However, the data presented here, as well as in section **2.3.6.2**, reveals significant fluctuations not only between individual cells but also across different treatment conditions, although most cells ubiquitously express the gene, making them less than ideal for use as housekeeping genes. This highlights the need for further refinement and validation, particularly given the increased use of single-cell and subcellular studies.

6.4 Conclusions

In order to probe how the expression of key ER-responsive genes varies with age, in both healthy and tumour tissues, a HiPlex RNAScope assay was conducted on FFPE samples from luminal breast cancer patients. Including both tumour and healthy regions within each sample allows for a

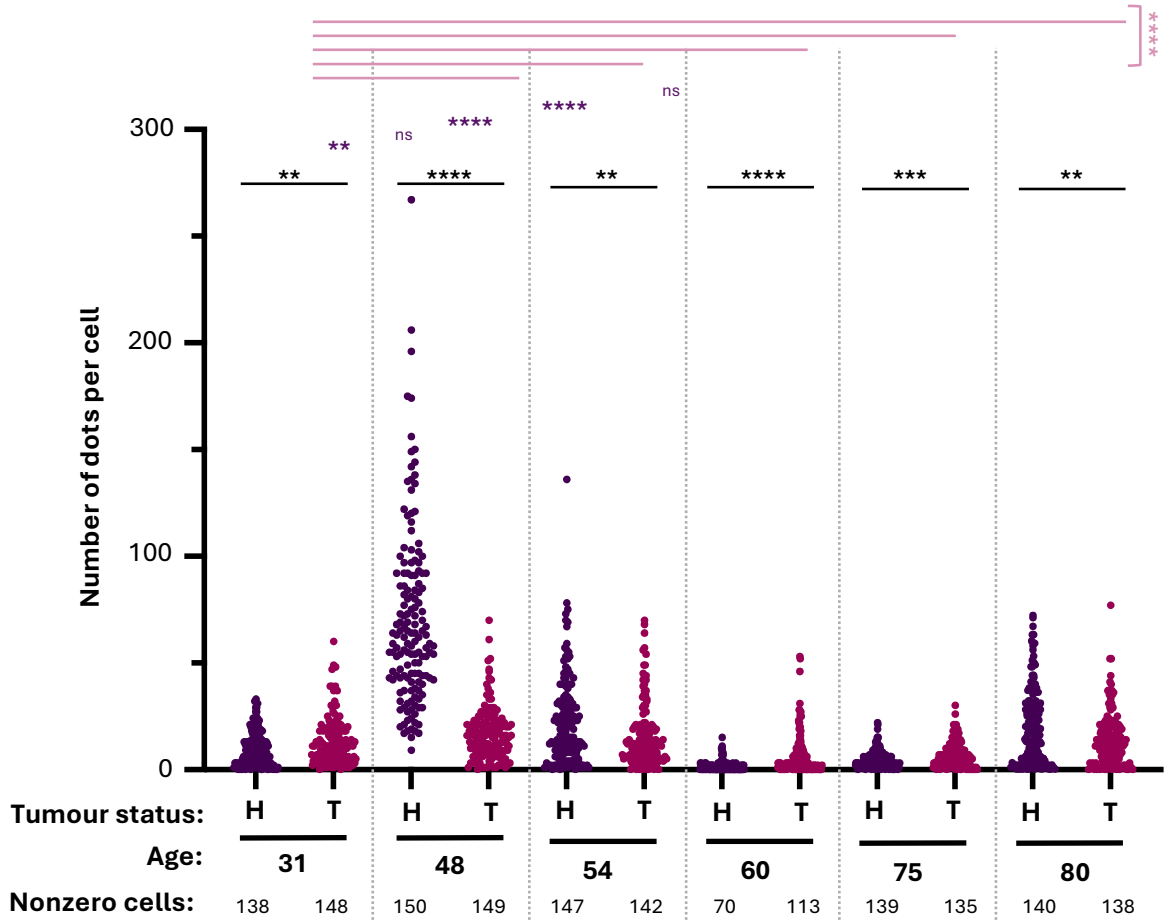


Figure 6.7 ***Number of RPL28 dots counted per cell, for 150 cells per condition and age, from HiPlex RNAScope Assay. Box-and-whisker plot of the number of subcellular counts using QuPath.*** *Data is split by age group, with counts from healthy (H; purple) and tumour (T; pink) regions, specified. Number of nonzero counts contributing towards the data plotted is shown under each plot. All results were analysed using a One-way ANOVA test in Excel, with respect to patient 31. Significant differences between data points are shown in black for comparison between the healthy and tumour counterpart of the same patient, in purple for all healthy cell counts and in pink for all tumour cell counts - ns = no significance, $P > 0.05$; * = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$; **** = $P \leq 0.0001$.*

direct comparison of tumour status within the same individual, and with ageing, between individuals. This involved hybridisation with a panel of 23 probes, including 12 housekeeping genes and 11 target genes, although only three genes were analysed and quantified for the purpose of this chapter. Notably, this analysis was only conducted in a small cohort of cells and only one sample per patient and age group. Furthermore, due to this small cohort size, it has not been possible to unbiasedly compare the effects of tumour grade in this case, and has not been attempted in the context of this preliminary study. The observed heterogeneity in *ESR1* and *FOXA1* expression, particularly the age-associated variability and marked differences between healthy and tumour tissues, suggests biologically relevant trends that warrant further investigation. Despite some patients (e.g., patient 75) deviating from the overall pattern, these findings highlight the critical need to account for both inter- and intra-patient variability when assessing gene expression landscapes in cancer. Moreover, the detected differences in

ESR1 expression and its correlation with patient age could inform hypotheses regarding age-driven changes in ER-signalling and tumour biology, particularly in the post-menopausal setting. Expanding this analysis across a larger patient cohort and stratifying by age, menopausal status, and tumour grade, would enable a deeper exploration of *ESR1* heterogeneity as a biomarker for tumour progression, endocrine resistance, or cellular plasticity in ER+ breast cancer. Additionally, RPL28 expression, while ubiquitous, varied significantly across and between samples. Thus, all housekeeping genes tested within the housekeeping panel will require validation to prior to selection of an optimal gene for use across the remaining samples.

Given more time, analysis would be conducted using a pseudo-multiplexed analytical approach, by overlaying H&E-stained and fluorescent images. This would avoid the estimation of membrane boundaries using QuPath but given that it does so with a 5 µm nuclear expansion (approximately that of an MCF-7 cell), this may not necessarily be a limitation in this case. To gain deeper insights into gene expression differences between healthy and tumour tissues across age groups, the gene and sample panel will eventually be expanded to include all imaged probes and imaging all samples that form part of the larger cohort, respectively, alongside an increased number of analysed cells to provide a more representative overview. This will subsequently form the basis for a larger scale study, involving integration with Nanostring's CosMx, to broaden the pool of analysed targets. Notably, ER expression was not quantified in any of the samples, and as such, subsequent transcript quantification could not be associated with ER expression. Further work in this regard, to differentiate between expression of the key genes analysed in this section, would involve immunofluorescence staining of the samples for the ER, and subsequent analysis of transcripts from solely ER+ cells. Further analysis of MKI67, a known proliferation and cancer marker which has already been incorporated into the HiPlex target panel, would provide insights into cell proliferation dynamics. Additional analysis will explore how expression varies with the morphological characteristics of imaged regions, for example, between ductal and alveolar areas. Overall, this HiPlex RNAScope assay provides a reliable technique for longitudinal studies and allows for spatial mapping of single cells, addressing the challenges in developing dynamic methods.

This initial dataset, while limited in sample size and scope, offers valuable proof-of-concept insights and lays the groundwork for larger-scale investigation, with the promise of yielding novel biological insight into age-related changes in ER+ breast cancer. Ultimately, the data generated as part of this continued project could directly inform downstream mechanistic studies, selection of therapeutic targets and possibly even inform the development of personalised treatment strategies based on the spatial transcriptomic landscape of individual tumours.

CHAPTER 7
CONCLUSIONS

Even after apparently successful treatment, patients suffering from HDBC experience relapse. It is believed that cancer cells that have survived their primary treatment, including micro-disseminated cells and those with a pre-adapted survival fitness, hibernate and enter a dormant state where they are frozen in the cell cycle, before reawakening to yield more aggressive disease. However, how they do this and what causes them to wake up is still somewhat a mystery. Recent advances in single-cell technologies have enabled exploration of the numerous cell states that cancers adopt throughout their evolutionary timeline with unrivalled resolution. However, these notably require that samples to be explored are removed from their native culture conditions and are lysed to access sub-cellular material. Dynamic epigenetic processes, which have recently been attributed to the emergence of resistance to therapy, can therefore not be studied using these techniques, limiting their exploration. Furthermore, given that parallel replicates are found to diverge with time *in vitro*, exploring related events in parallel experiments becomes unreliable. Combined, these severely limit our understanding of how dormant cancer cells adapt and evade therapy. The work presented in this thesis outlines some key techniques and advancements in sub-cellular transcript detection. This was initially attempted using nanotweezers and nanopipettes, both of which are compatible with living cells, before working with more static techniques, that, much like current sequencing techniques, involve sacrificing samples. While a notable limitation, due to the difficulty in transcript detection and the influence of a numerous confounding factors present within complex culture media, the decision was made to halt attempts at optimising these dynamic techniques.

Chapter 2 highlighted that particles can be manipulated, both in solution and in cells, using DEP, regardless of their net overall charge, allowing for pre-concentration from their native environment. This technique, involving NTs, was applied to the extraction of mRNA from MCF-7 cells but faced numerous issues with subsequent transcript detection. As a result, various modifications were attempted, including reducing aperture sizes and inter-electrode distances to increase the DEP force generated without the application of increased voltages, surface modification and pooling multiple biopsies into the same reaction. NT surfaces were modified with poly-T linkers, complementary to the poly-A terminal caps of mRNA. This revealed a substantial increase in capture ability when working with mRNA in solution, but could not be recapitulated in cells. Furthermore, pooling multiple extracts for processing as a single reaction also did not yield the expected decrease in Ct value, hovering around and sometimes even exceeding 35 amplification cycles. Variability in detection efficiency can be attributed to inconsistency in nanotweezer aperture size or carbon deposition, controlling for which is not trivial, despite best attempts to regulate the manual NT fabrication process. Further work, involving the exploration of RNases within culture media, and more accurate quantification of cell penetration and capture efficiency when within the cell are required, especially if this technique is to be used to conduct long-term dynamic studies in AI-treated MCF-7 cells, wherein gene expression of key ER-responsive genes is known to decrease.

An alternative single-cell biopsy technique is presented in **chapter 3**. Using an angular SICM approach to enhance NP manoeuvrability and minimise tip breakage caused by rapid cell approach, this technique provides reliable determination of the nanoprobe location with regard to the cell. While different cell penetration and cytosol extraction techniques were attempted, including application of a driving force post-membrane indentation, application of electroporation pulses upon positioning the NP on the membrane proved most reliable and reproducible. However, as with NTs, challenges with quantifying cytoplasmic extracts persist. Future work in this regard should therefore involve maximising extraction volumes or isolating entire cells for analysis before transitioning to sub-cellular analyses. This should be coupled with recovery and viability assessments to determine how cells withstand stress exerted upon cytoplasmic extraction and its impact on transcriptional profiles. Although initial validation through lambda phage (λ)-DNA sampling has optimised detection, further sensitivity improvements

are needed. Ultimately, this chapter presents an optimised technique to extract cytosolic material from cells that can enable continuous tracking of single-cell transcriptomic changes in response to therapy, crucial for evaluating treatment efficacy and advancing therapeutic developments in cancer research.

Chapter 4 presents efforts to optimise a nucleic acid amplification technique, NASBA. Given the difficulties in transcript detection using the NT platform, this was attempted with the idea that increasing RNA concentrations extracted using nanobiopsies would facilitate subsequent transcript detection using RT-qPCR. Amplification was attempted using two reagent kits, a commercially available kit (AMS-BIO) and an optimised kit with increased sensitivity (INSIGHT). However, despite varying reagent components by switching from one kit to another, varying primer sequencies, increasing enzyme concentrations and isothermal incubation durations, RNA could not be reliably amplified. No further attempts using alternative amplification techniques were attempted due to time constraints. However, the concept of RNA amplification combined with nanobiopsies presents a possible solution to the current limitations in transcript detection sensitivity.

Deviating slightly from dynamic measurements, **chapter 5** presents optimisation of a multi-omic pipeline to probe both the genome and transcriptome of the same cell. This involved using a reporter cell line developed by Hannah Dewhurst (Magnani Lab) to probe the transient cycling dynamics of dormancy. Coupled with live imaging experiments, cycling and dormant cells were isolated from a larger population at key timepoints, initially to be processed using the CellenChip 3'-RNA-seq kit. However, due to technical difficulties entailing issues with dispensing into the respective wells and reagent evaporation, this proved unsuccessful. As such, an alternative platform was developed in collaboration Max Mossner (Genomics and Evolutionary Dynamics Group), involving the dispensation of cells using the Cellenone, processing using G&T-seq to separate gDNA and RNA, prior to processing using S3-ATAC-seq and FLASH-seq respectively. While no sequencing of the samples has been conducted, quality control from S3-ATAC-seq processing reveals successful isolation, amplification and library preparation. Generation of a FLASH-seq library was slightly more challenging as exhibited by the lack of cDNA within the pooled RTed sample, prompting optimisation of reagent combinations. This integrated FLASH-seq and S3-ATAC-seq pipeline enhances single-cell analysis and enables detailed comparisons of transcriptional and chromatin differences between sporadic cycling and true awakening, expanding applications in understanding cell state dynamics.

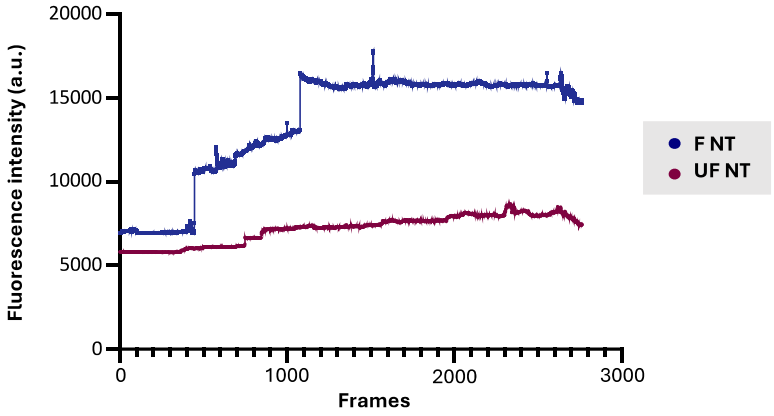
Due to the challenges faced in dynamic single-cell analytical technique, an alternative technique, enabling spatial profiling of RNA transcripts was attempted in **chapter 6** – RNAScope. While HDBC remains understudied, even less attention has been given to the role of ER expression in regulating and protecting against such. Preliminary experiments were therefore conducted using a small cohort of FFPE samples, from patients of age ranging from 31 to 80 years, with all samples containing a region of tumour and healthy tissue. Expression of two oestrogen-responsive genes, *ESR1* and *FOXA1*, alongside *RPL28* as a housekeeping gene were quantified in 150 cells across both tumour and healthy tissue. This revealed an overall increase in the number of *ESR1* and *FOXA1* transcripts with age and significant variability in *RPL28* expression between samples and ages. Further work, in regard to ER expression and proliferation dynamics is however required, alongside exploring a larger cohort of samples, possibly even of varying cancer grades, to establish a relationship between ER expression and understanding the 'retirement' phenotype that ER+ cells adopt with age.

Overall, this thesis presents various techniques, albeit some still requiring further optimisation to achieve their full potential, that can be implemented to enhance our understanding of transient processes underlying cancer evolution. The two dynamic processes can be considered stepping stones in the development of more enhanced sampling techniques, allowing for more longitudinal studies, in a vast array of heterogenous disease models to be conducted.

SUPPLEMENTARY

Supplementary Chapter 2

Supplementary 2.1 Sample fluorescence trace of functionalised and unfunctionalised nanotweezers in solution containing AF-488 conjugated RNA.



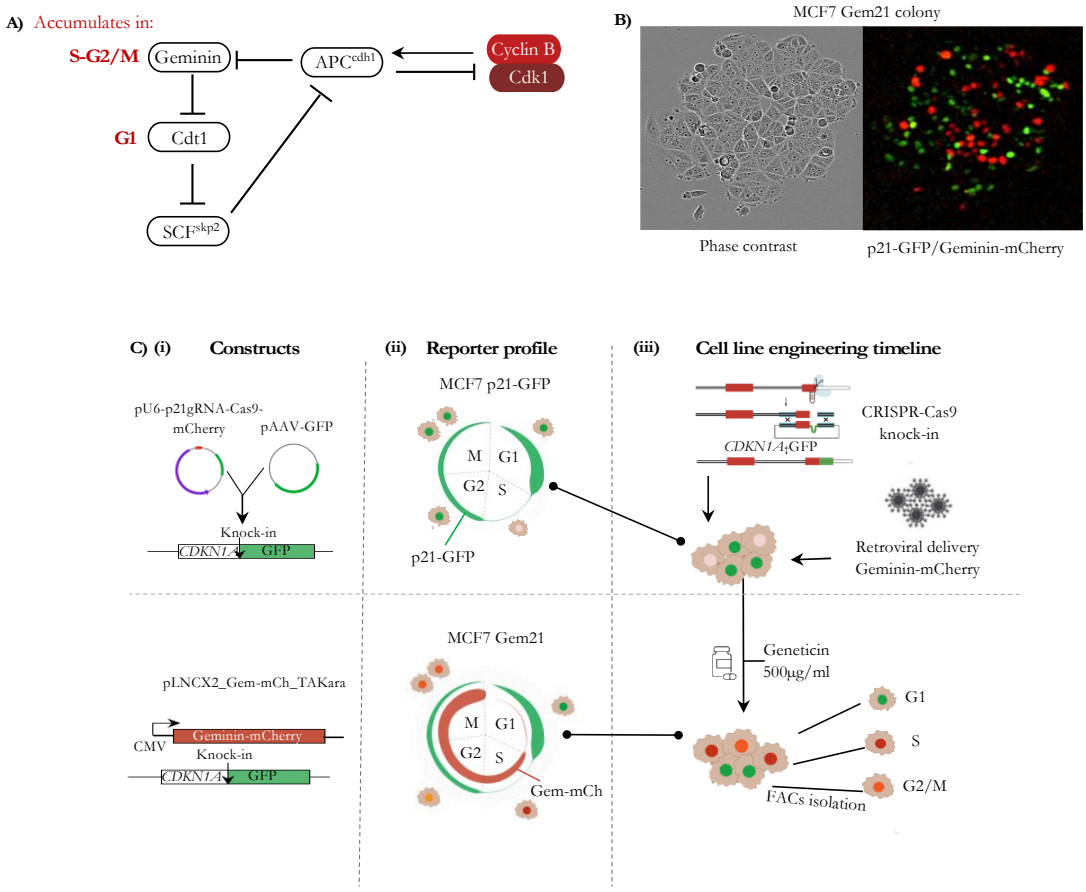
Supplementary Figure 1 ***Fluorescence intensity traces for NT validation.*** *The longer the NT is held in solution, the more fluorescent RNA is trapped at the tip, seen as an increase in fluorescence signal. This fluorescence accumulation was seen to be larger in poly-thymine F NTs (purple), as compared to UF NTs (blue).*

Supplementary 2.2 FIJI Macro for RNAScope preliminary image processing.

```
//Open image manually


//Select location where images/results are to be stored
	output = getDirectory("Output directory for results");


//Store name of current image
	img_name=getTitle();


//Create Maximum Z-projection
	run("Z Project...", "projection=[Max Intensity]");


//Make composite image
	selectImage("MAX_"+img_name);
	run("Make Composite");


//Save image
	out = output + "MAX_"+img_name;
	saveAs("Tiff",out);
	close()
	close(img_name)
```

Supplementary Chapter 5

Supplementary 5.1 Gem21-MCF-7 cell line generation



Supplementary Figure 2 ***Engineering of gem21-MCF-7 cells, by Hannah Dewhurst, Magnani Lab, Imperial College London.*** ***A*** *Geminin and Cdt1 are inversely regulated throughout the cell cycle, controlled by the proteosome regulatory loop.* ***B*** *Phase-contrast and GFP/RFP images of a representative cell colony, captured on the Incucyte Zoom. Background subtraction was performed using FIJI.* ***C*** *Summary of engineering of cell cycle reporter cell line.* ***(i)*** *Constructs used to engineer the cell line, enabling* ***(ii)*** *visualization of a p21-GFP, or combined MCF7 Gem21 fluorescent reporter profile. Band thickness correlates to the level or protein detected at each stage of the cell cycle, with colour indicative of fluorescent reporter – p21 = GFP and geminin = mCherry (mCh).* ***(iii)*** *Chronological order of cell line engineering, resting in generation of gem21-MCF-7 cells, used for transient awakening experiments. Figure and legend adapted from Figure 4.5, Investigating the transcriptional drivers of dormancy in luminal A breast cancer, Hannah Dewhurst (unpublished).*

Supplementary 5.2 FIJI Macro for merging images from different timepoints into a single TIFF.

```
//Select directory of images to open

	input = getDirectory("Input directory where images are stored");


//Select location where images/results are to be stored

	output = getDirectory("Output ditrectory for results");


//Get list of files

	blank=newArray("end");

	list = getFileList(input);

	list=Array.sort(list);

	list=Array.concat(list,blank);


//Create image stack

	ser=substring(list[0],0,list[0].length-22);

	for (image=0;image<list.length-1;image++){

	while (startsWith(list[image],ser)==true){

	full = input +  list[image];

	open(full);

	image=image+1;}

	run("Images to Stack", "use");


//Save image

	out = output + ser + "series";

	saveAs("Tiff",out);

	close();

	if (list[image]!="end"){

	ser=substring(list[image],0,list[image].length-22);

	image=image-1;

	}

	}
```

Supplementary 5.3 FIJI Macro for quantifying number of cells per cell cycle phase.

```
//Analyse full stack in one pass
	run("Set Measurements...", "area mean min stack display redirect=None decimal=3");

//Intensity limits for positive cells
	RFP_limit=5;
	GFP_limit=10;

//Initialise
	fn=getTitle();
	roiManager("reset");
	run("Clear Results");
	run("Select None");
	getDimensions(ImageWidth, ImageHeight, ImageChannels, ImageSlices, ImageFrames);


//Arrays
	Total_cell_number=newArray(ImageSlices);
	Total_RFP_cells=newArray(ImageSlices);
	Total_GFP_cells=newArray(ImageSlices);
	Total_both=newArray(ImageSlices);
	Fr_list=Array.slice(Array.getSequence(ImageSlices+1),1,ImageSlices+1);


//Create mask
	run("Duplicate...", "title=mask duplicate");
	run("Make Composite");
	run("Gaussian Blur...", "sigma=2 stack");
	run("Subtract Background...", "rolling=25 stack");
	run("Duplicate...", "title=overlay duplicate");
	selectWindow("overlay");
	Stack.setChannel(1);
	run("Enhance Contrast", "saturated=0.35");
	Stack.setChannel(2);
```

```
run("Enhance Contrast", "saturated=0.35");
run("RGB Color", "slices");

selectWindow("mask");
run("Split Channels");
close("C3-mask");

imageCalculator("Max create stack", "C1-mask","C2-mask");
selectWindow("Result of C1-mask");
rename("mask")

selectWindow("C1-mask");
rename("RFP");
selectWindow("C2-mask");
rename("GFP");

selectWindow("mask");
setAutoThreshold("Triangle dark");

setOption("BlackBackground", true);
run("Convert to Mask", "method=Triangle background=Dark black");
//run("Watershed", "stack");
run("Adjustable Watershed", "tolerance=0.1 stack");//adjustable watershed value
run("Erode", "stack");

run("Analyze Particles...", "size=15-Infinity add stack");
n=roiManager("count");

//Measure red channel
selectWindow("RFP");
roiManager("Measure");
selectWindow("Results");
RFP=Table.getColumn("Mean");
Fr=Table.getColumn("Slice");
```

```
	run("Clear Results");

//Measure green channel
	selectWindow("GFP");
	roiManager("Measure");
	selectWindow("Results");
	GFP=Table.getColumn("Mean");
	run("Clear Results");



	Roi_list=Array.slice(Array.getSequence(n+1),1,n+1);


	for (roi=0;roi<n;roi++){
		frame=Fr[roi]-1;
		Total_cell_number[frame]=Total_cell_number[frame]+1;
		if (RFP[roi]>RFP_limit){
			Total_RFP_cells[frame]=Total_RFP_cells[frame]+1;
		}
		if (GFP[roi]>GFP_limit){
			Total_GFP_cells[frame]=Total_GFP_cells[frame]+1;
		}
			if (RFP[roi]>RFP_limit && GFP[roi]>GFP_limit){
			Total_both[frame]=Total_both[frame]+1;
		}

	}


	Table.create(fn+" Raw Data");
	Table.setLocationAndSize(150, 150, 920, 600);
	Table.setColumn("Frame", Fr);
	Table.setColumn("ROI", Roi_list);
	Table.setColumn("RFP mean", RFP);
	Table.setColumn("GFP mean", GFP);
	Table.setColumn("Frame (summary)",Fr_list);
```

```
		Table.setColumn("total cell number",Total_cell_number);
		Table.setColumn("total RFP cells",Total_RFP_cells);
		Table.setColumn("total GFP cells",Total_GFP_cells);
		Table.setColumn("total both cells",Total_both);


//Finish up
close("RFP");
close("GFP");
close("mask");


		Table.create(fn+" Counts Summary");
		Table.setLocationAndSize(200, 200, 620, 500);
		Table.setColumn("Frame (summary)",Fr_list);
		Table.setColumn("total cell number",Total_cell_number);
		Table.setColumn("total RFP cells",Total_RFP_cells);
		Table.setColumn("total GFP cells",Total_GFP_cells);
		Table.setColumn("total both cells",Total_both);

		selectWindow("overlay");
		roiManager("Show All with labels");
```

Supplementary Chapter 6

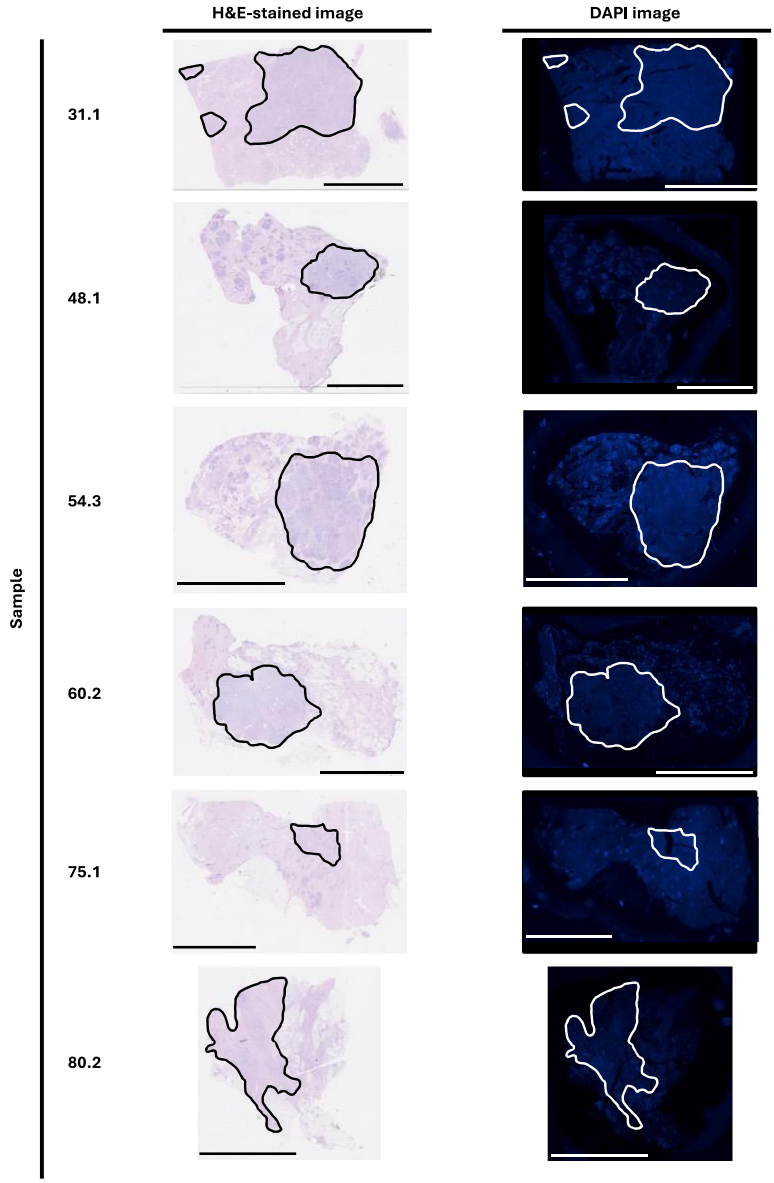
Supplementary 6.1 Full cohort of FFPE breast samples.

Supplementary Table 1 Full cohort of FFPE breast tissue samples provided by Dr Andrea Vingiani (Instituto Nazionale de Tumori Milano) from which a select panel was chosen for processing.

Slide	Decade	Age	Stage	Grade	Menopause	Last period	Pregnancy	Breastfeeding	Note
S19-6408 D1	< 40	31	pT1cpN0	G1	N		1		
S21-3901 B1	< 40	33	pT1cpN0	G1	N		0	-	
S19-8061 C1	< 40	35	pT1cpN0	G2	N	26/11/2019 (HISTOLOGY: 14/11/19)	3		1 abortion + 2 childbirths
S18-2160 3-1	< 40	36	pT2pN0	G3	N		0	-	
S17-6953 4-4	< 40	37		G2	N	28/10/2017 (HISTOLOGY: 27/9/17)	2		
S20-8266 D1	< 40	37	pT1cpN1a	G2	N	19/10/2020 (HISTOLOGY: 11/11/20)	2	Y	2 C-sections; oral contraceptives intake
S17-3267 2-2	< 40	37	pT1bpN0	G2	N		0		
S20-5871 B3	< 40	37	pT1cpN0	G2	N	01/08/2020 (HISTOLOGY: 28/08/20)	2	N	oral contraceptives intake (for 13 years)
S17-659 3-4	< 40	38	pT1cpN0	G2	N	01/2017 (HISTOLOGY: 27/01/17)	1		
S20-3553 B3	< 40	39	pT2pN1a	G2	N		0	-	oral contraceptives intake (for 10 years)
S23-11381	40-49	48	pT1cpN0	G1	N	21/10/2023 (HISTOLOGY: 21/12/23)	0	-	
S23-9367	40-49	48	pT1bpN0	G1	N	19/09/2023 (HISTOLOGY: 23/10/23)	3	Y (1,5 y/child)	2 childbirths + 1 C-section; oral contraceptives intake (for 20 years)
S23-10039	40-49	48	pT1cpN0	G3	N	28/08/2023 (HISTOLOGY: 11/11/23)	2	Y (4 m /child)	oral contraceptives intake (for 20 years)
S24-1517	40-49	46	pT2pN1a	G2	N	23/12/2023 (HISTOLOGY: 15/02/24)	N.A.	N.A.	
S22-742	40-49	47	pT1cpN0	G3	N	08/12/2021 (HISTOLOGY: 02/02/22)	0	-	oral contraceptives intake (for 1 year)
S24-1063	40-49	48	pT1bpN0	G1	N (pre)	22/11/2023 (HISTOLOGY: 02/02/24)	1	Y (1y)	1 C-section
S23-10043	40-49	49	pT1cpN0	G1	N	20/09/2023 (HISTOLOGY: 11/11/23)	2	Y	oral contraceptives intake (for 1 year)
S23-9841	40-49	48	pT1bpN0	G1	N	18/09/2023 (HISTOLOGY: 07/11/23)	0	-	
S23-10348	40-49	45	pT2pN1a	G2	N	10/10/2023 (HISTOLOGY: 20/11/23)	2	Y (only 2 m)	oral contraceptives intake (for 10 years)
S23-5430	40-49	46	pT2pN0	G3	N	11/07/2023 (HISTOLOGY: 08/06/23)	1	N.A.	1 twin birth
S23-10027	50-59	51	pT1bpN0	G1	PERI	4 months ago	2	Y (8 m)	
S23-11074	50-59	54	pT1cpN0	G3	PERI	N.A.	N.A.	N.A.	
S23-10172	50-59	52	pT1bpN0	G2	Y	8 years ago	2	Y (20 m)	2 C-sections; oral contraceptives intake
S23-8199	50-59	51	pT1apN0	G1	Y	1 year ago	1	Y	oral contraceptives intake (for 2 years)
S23-8238	50-59	54	pT3pN1a	G3	Y	N.A.	0	-	
S21-3908	60-69	60	pT1cpN0	G2	Y	7 years ago	1		1 C-section
S21-3741	60-69	61	pT1cpN1a	G2	Y	10 years ago	2	Y (only 1 m)	1 abortion + 1 childbirths; oral contraceptives intake (for 3 years)
S21-4824	60-69	62	pT2pN0	G2	Y	14 years ago	0	-	oral contraceptives intake (for 4 years)
S21-3740	60-69	64	pT2pN0	G2	Y	12 years ago	3	Y (9 m)	oral contraceptives intake (for 1 year)
S22-2264	60-69	65	pT2pN1a	G3	Y	11 years ago	1	Y (9 m)	oral contraceptives intake (for 5 years)
S21-4599	70-79	75	pT2pN1mi	G3	Y	26 years ago	4		
S21-4601	70-79	75	pT1cpN0	G1	Y	N.A.	N.A.	N.A.	
S21-4186	70-79	77	pT2pN0	G2	Y	N.A.	N.A.	N.A.	
S21-3498	70-79	77	pT1cpN0	G3	Y	34 years ago	2		
S21-3755	70-79	77	pT1cpN0	G2	Y	N.A.	N.A.	N.A.	
S21-4563	80-89	80	pT1cpN0	G2	Y	28 years ago	1	-	1 abortion
S21-2751	80-89	82	pT1bpN0	G1	Y	29 years ago	3	Y (5 m + 9 m)	2 childbirths + 1 abortion
S22-1704	80-89	83	pT1cpN0	G2	Y	early menopause	N.A.	N.A.	
S21-2851	80-89	83	pT2pN0	G3	Y	45 years ago	4	few weeks	3 childbirths + 1 abortion
S21-3270	80-89	86	pT1apN0	G1	Y	N.A.	N.A.	N.A.	

Supplementary 6.2 Whole Slide Imaging.

Whole slide imaging was conducted using the Nanozoomer XR (Hamamatsu) coupled with an LX2000 Lightningexciter (Model: L11600-05, Hamamatsu). All images were acquired using NDP.Scan v3.4.3. Initially, images of the H&E-stained slides corresponding to the selected samples were captured with a 20x objective in brightfield mode. Similarly, once the RNAScope assay had been conducted on the slides, they were imaged for DAPI in fluorescence mode, with a 20x objective. For both brightfield and fluorescent images, scans were conducted as semi-automatic batches, as single layers with a manual scan area and 9pt autofocus. The laser intensity, exposure time and gain for fluorescence images were set to 100%, 13.5 ms and 5x, respectively. All images were visualised using the NDP2.view software. Tumour and healthy regions of tissues were identified and annotated with the assistance of Dr Ioannis Roxanis (Molecular Pathology, ICR).



Supplementary Figure 3 ***H&E-stained slides and DAPI-stained nuclei of imaged samples.*** *All slides are annotated, with tumour regions encircled in black for the H&E images, and in white for DAPI images. Scale bars represent 10 mm.*

Supplementary 6.3 QuPath transcript detection parameters: HiPlex RNAScope.

Supplementary Table 2 ***Subcellular transcript detection parameters for RPL28, ESR1 and FOXA1 using QuPath.***

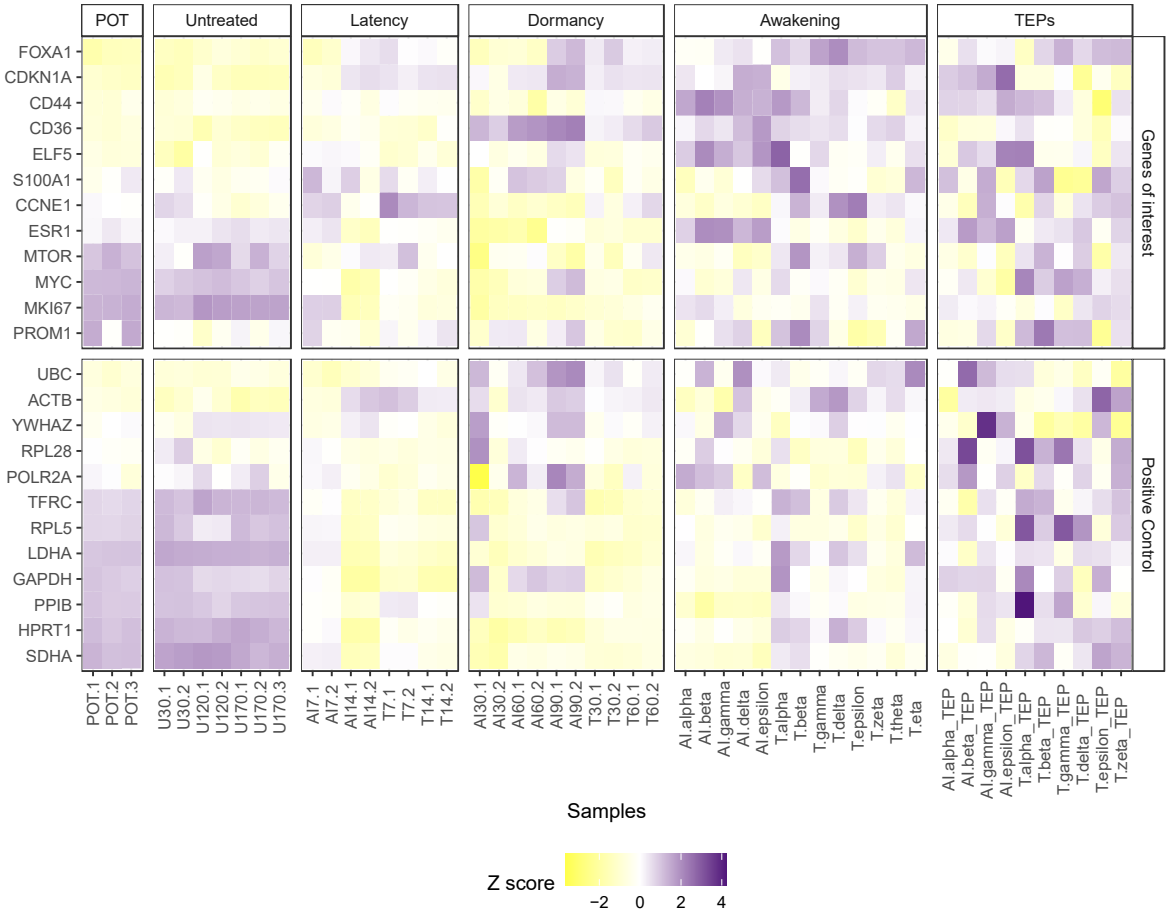
RPL28

Sample		Threshold	Expected spot size	Minimum spot size	Maximum spot size
31.1	**Tumour**	200	0.3	0.2	1
	Healthy	150	0.3	0.2	1.5
48.1	**Tumour**	500	0.3	0.1	1.5
	Healthy	50	0.3	0.1	1.5
54.3	**Tumour**	50	0.3	0.1	1.5
	Healthy	50	0.3	0.1	1.5
60.2	**Tumour**	50	0.3	0.3	1.5
	Healthy	50	0.3	0.3	1.5
75.1	**Tumour**	20	0.3	0.3	1.5
	Healthy	20	0.3	0.3	1.5
80.2	**Tumour**	50	0.3	0.3	1.5
	Healthy	50	0.3	0.3	1.5

ESR1					
Sample		**Threshold**	**Expected spot size**	**Minimum spot size**	**Maximum spot size**
31.1	**Tumour**	700	0.3	0.3	1
	Healthy	250	0.3	0.3	1.5
48.1	**Tumour**	125	0.3	0.1	1.5
	Healthy	175	0.3	0.1	1.5
54.3	**Tumour**	75	0.3	0.25	1.5
	Healthy	75	0.3	0.25	1.5
60.2	**Tumour**	150	0.3	0.1	1.5
	Healthy	50	0.1	0.3	0.6
75.1	**Tumour**	250	0.3	0.1	1.5
	Healthy	120	0.3	0.1	1.5
80.2	**Tumour**	75	0.3	0.1	1.5
	Healthy	75	0.3	0.25	1.5

FOXA1					
Sample		**Threshold**	**Expected spot size**	**Minimum spot size**	**Maximum spot size**
31.1	**Tumour**	400	0.3	0.1	1
	Healthy	150	0.3	0.3	1.5
48.1	**Tumour**	125	0.3	0.1	1.5
	Healthy	100	0.3	0.1	1.5
54.3	**Tumour**	50	0.3	0.25	1.5
	Healthy	50	0.3	0.25	1.5
60.2	**Tumour**	100	0.3	0.1	1.5
	Healthy	75	0.1	0.3	0.6
75.1	**Tumour**	100	0.3	0.1	1.5
	Healthy	85	0.3	0.1	1.5
80.2	**Tumour**	50	0.3	0.1	1.5
	Healthy	50	0.3	0.25	1.5

Supplementary 6.4 Heatmap of gene expression of all housekeeping and target genes used in the HiPlex RNAScope Assay.



Supplementary Figure 4 ***Heatmap of all genes of interest and housekeeping genes used in the HiPlex RNAScope Assay.*** *Gene expression from scRNA-sequencing experiments conducted by Rosano et al, in all treatment replicates are shown.[2] POT = founder population; U = untreated population' AI = aromatase inhibition; TEP = terminal endpoint. RPL28 shows least variation across all treatment conditions and analytical timepoints, and as such, was the chosen housekeeping gene to quantify alongside ESR1 and FOXA1 expression in FFPE samples.*

Copyright Permission

Supplementary CP.1 Copyright permission table.

Supplementary Table 3 Copyright permission table

Figure Number	Source	Copyright holder	Permission	Licence Number	Section Reference
1.1	Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 12 (2022). https://doi.org/ 10.1158/2159-8290.CD-21-1059	Cancer Discovery	Granted	1541305-1	CP.2
1.3	Harbeck, N., Penault-Llorca, F., Cortes, J. et al. Breast cancer. Nat Rev Dis Primers 5, 66 (2019). https://doi.org/10.1038/s41572-019-0111-2	Springer Nature	Granted	5895460086104	CP.3
1.4	Smittenaar, C., Petersen, K., Stewart, K. et al. Cancer incidence and mortality projections in the UK until 2035. Br J Cancer 115 (2016). https://doi.org/10.1038/bjc.2016.304	Springer Nature	Open Access (Creative Commons 4.0)	n/a	n/a
1.6 and 5.1	Rosano, D., Sofyali, E., Dhiman, H. et al. Long-term Multimodal Recording Reveals Epigenetic Adaptation Routes in Dormant Breast Cancer Cells. Cancer Discov 14 (2024). https://doi.org/10.1158/2159-8290.CD-23-1161	Cancer Discovery	Open Access (Creative Commons 4.0)	n/a	n/a
1.7	Xiong, X., Wang, X., Liu, CC. et al. Deciphering breast cancer dynamics: insights from single-cell and spatial profiling in the multi-omics era. Biomark Res 12 (2024). https://doi.org/10.1186/s40364-024-00654-1	Biomarker Research	Open Access (Creative Commons 4.0)	n/a	n/a
1.8	Actis, P., Maalouf, M. M., Kim, H. J., Lohith, A. et al. Compartmental Genomics in Living Cells Revealed by Single-Cell Nanobiopsy. ACS Nano 8 (2014) https://doi.org/10.1021/nn405097u	American Cancer Society	Granted	n/a	CP.4
1.8	Zhang, Y., Takahashi, Y., Hong, S.P. et al. High-resolution label-free 3D mapping of extracellular pH of single living cells. Nat Commun 10, 5610 (2019). https://doi.org/10.1038/s41467-019-13535-1	Springer Nature	Granted	1532963-1	CP.5

1.9	Dufrêne, Y., Ando, T., Garcia, R. et al. Imaging modes of atomic force microscopy for application in molecular and cell biology. Nature Nanotech 12 (2017). https://doi.org/10.1038/nnano.2017.45	Springer Nature	Granted	5898291368219	CP.6
1.9	Meister, A., Gabi, M., Behr, Pascal. et al. FluidFM: Combining Atomic Force Microscopy and Nanofluidics in a Universal Liquid Delivery System for Single Cell Applications and Beyond. Nano Letters 9 (2009) https://doi.org/10.1021/nl901384x	American Chemical Society	Granted	n/a	CP.7
1.9	Guillaume-Gentil, O., Grindberg, R. V., Kooger, R. et al. Tunable Single-Cell Extraction for Molecular Analyses. Cell 166 (2016). https://doi.org/10.1016/j.cell.2016.06.025	Elsevier	Granted	5898301131773	CP.8
1.9	Chen, W., Guillaume-Gentil, O., Rainer, P.Y. et al. Live-seq enables temporal transcriptomic recording of single cells. Nature 608 (2022). https://doi.org/10.1038/s41586-022-05046-9	Springer Nature	Open Access (Creative Commons 4.0)	n/a	n/a
1.10 and 2.6	Nadappuram, B.P., Cadinu, P., Barik, A. et al. Nanoscale tweezers for single-cell biopsies. Nature Nanotech 14 (2019). https://doi.org/10.1038/s41565-018-0315-8	Springer Nature	Granted	1530834-1	CP.9
2.1	Higgins, S. G., Stevens, M. M. Extracting the contents of living cells. Science 356 (2017). https://doi.org/10.1126/science.aan0228	AAAS	Granted	6081450934146	CP.10
2.1	Sahota, A., Nadappuram, B. P., Kwan, Z. et al. Spatial and Temporal Single-Cell Profiling of RNA Compartmentalization in Neurons with Nanotweezers. ACS Nano 19 (2025). https://doi.org/10.1021/acsnano.5c02056	ACS Nano	Open Access (Creative Commons 4.0)	n/a	n/a
2.2	Voldman, J. Electrical forces for microscale cell manipulation. Annu. Rev. Biomed. Eng. 8 (2006). https://doi.org/10.1146/annurev.bioeng.8.061505.095739	Annual Reviews, Inc.	Granted	1541321-1	CP.11
3.1	Hamill, O.P., Marty, A., Neher, E. et al. Improved patch-clamp techniques for high-resolution current recording from cells	Springer Nature	Granted	5882540737709	CP.12

	and cell-free membrane patches. Pflugers Arch. 391 (1981). https://doi.org/10.1007/BF00656997				
3.2	Novak, P., Li, C., Shevchuk, A. et al. Nanoscale live-cell imaging using hopping probe ion conductance microscopy. Nat Methods 6 (2009). https://doi.org/10.1038/nmeth.1306	Springer Nature	Granted	5882540995029	CP.13
5.4	Mulqueen, R. et al. High-content single-cell combinatorial indexing. Nat Biotech 39 (2021). https://doi.org/10.1038/s41587-021-00962-z	Springer Nature	Granted	5891951463709	CP.14

Supplementary CP.2

Copyright permission for "Hallmarks of Cancer: New Dimensions".[10]

Marketplace

Order Date	30-Oct-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1541305-1		
ISSN	2159-8274	Publisher	American Association for Cancer Research
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Cancer discovery	Rightsholder	American Association for Cancer Research
Article Title	Hallmarks of Cancer: New Dimensions	Publication Type	e-Journal
Author / Editor	American Association for Cancer Research.	Start Page	31
		End Page	46
Date	01/01/2011	Issue	1
Language	English	Volume	12
Country	United States of America		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of Images / Photos / Illustrations	1	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the Disabled?	No
		Minor Editing Privileges?	No
Who Will Republish the Content?	Academic institution	Incidental Promotional Use?	No
Duration of Use	Life of current edition	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Professor	Institution Name	Imperial College London
Instructor Name	Luca Magnani	Expected Presentation Date	2024-11-02

ADDITIONAL DETAILS

Order Reference Number	N/A	The Requesting Person / Organization to Appear on the License	Debjani Saha

REQUESTED CONTENT DETAILS

Supplementary CP.3

Copyright permission for “Breast cancer”.[18]

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Oct 24, 2024

This Agreement between Debjani Saha ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5895460086104
License date	Oct 24, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Disease Primers
Licensed Content Title	Breast cancer
Licensed Content Author	Nadia Harbeck et al
Licensed Content Date	Sep 23, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Quantifying Breast Cancer Adaptation: Platforms for Studying Transcriptional Changes in Single Cells
Institution name	Imperial College London
Expected presentation date	Nov 2024
Portions	Fig. 1: Breast cancer.
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, other W12 0NN United Kingdom Attn: Miss Debjani Saha
Billing Type	Invoice
Billing Address	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, United Kingdom W12 0NN Attn: Debjani Saha
Total	**0.00 GBP**
Terms and Conditions	

Springer Nature Customer Service Centre GmbH Terms and Conditions

Supplementary CP.4

Copyright permission for “Compartmental Genomics in Living Cells Revealed by Single-Cell Nanobiopsy”.[89]

Supplementary CP.5

Copyright permission for “High-resolution label-free 3D mapping of extracellular pH of single living cells”.[82]

Marketplace

This is a License Agreement between Debjani Saha ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	05-Oct-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1532963-1		
ISSN	2041-1723	Publisher	Nature Publishing Group
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Nature Communications	Publication Type	e-Journal
Article Title	High-resolution label-free 3D mapping of extracellular pH of single living cells.	Start Page	5610
		Issue	1
		Volume	10
Date	01/01/2010	URL	http://www.nature.com/ncomms/index.html
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		
Rightsholder	Springer Nature BV		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of Images / Photos / Illustrations	4	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the Disabled?	No
		Minor Editing Privileges?	Yes
Who Will Republish the Content?	Academic institution	Incidental Promotional Use?	No
Duration of Use	Life of current and all future editions	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Professor	Institution Name	Imperial College London
Instructor Name	Luca Magnani	Expected Presentation Date	2024-11-01

ADDITIONAL DETAILS

The Requesting Person / Organization to Appear on the License	Debjani Saha

REQUESTED CONTENT DETAILS

Supplementary CP.6

Copyright permission for "Imaging modes of atomic force microscopy for application in molecular and cell biology".[83]

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 29, 2024

This Agreement between Debjani Saha ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5898291368219
License date	Oct 29, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Nanotechnology
Licensed Content Title	Imaging modes of atomic force microscopy for application in molecular and cell biology
Licensed Content Author	Yves F. Dufrêne et al
Licensed Content Date	Apr 6, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no

Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Quantifying Breast Cancer Adaptation: Platforms for Studying Transcriptional Changes in Single Cells
Institution name	Imperial College London
Expected presentation date	Nov 2024
Portions	Figure 2: AFM-based imaging of native biological systems to molecular resolution.
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, other W12 0NN United Kingdom Attn: Miss Debjani Saha
Billing Type	Invoice
Billing Address	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, United Kingdom W12 0NN Attn: Debjani Saha
Total	0.00 GBP
Terms and Conditions	

Springer Nature Customer Service Centre GmbH Terms and Conditions

Supplementary CP.7

Copyright permission for "FluidFM: Combining Atomic Force Microscopy and Nanofluidics in a Universal Liquid Delivery System for Single Cell Applications and Beyond".[86]

29/10/2024, 16:42 Rightslink® by Copyright Clearance Center

CCC RightsLink

DS

ACS Publications
Most Trusted. Most Cited. Most Read.

FluidFM: Combining Atomic Force Microscopy and Nanofluidics in a Universal Liquid Delivery System for Single Cell Applications and Beyond

Author: André Meister, Michael Gabi, Pascal Behr, et al
Publication: Nano Letters
Publisher: American Chemical Society
Date: Jun 1, 2009

Copyright © 2009, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and ConditionsComments? We would like to hear from you. E-mail us at customercare@copyright.com

https://s100.copyright.com/AppDispatchServlet 1/1

Supplementary CP.8

Copyright permission for “Tunable Single-Cell Extraction for Molecular Analyses”.[87]

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 29, 2024

This Agreement between Debjani Saha ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5898301131773
License date	Oct 29, 2024
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	Tunable Single-Cell Extraction for Molecular Analyses
Licensed Content Author	Orane Guillaume-Gentil,Rashel V. Grindberg,Romain Kooger,Livie Dorwling-Carter,Vincent Martinez,Dario Ossola,Martin Pilhofer,Tomaso Zambelli,Julia A. Vorholt
Licensed Content Date	Jul 14, 2016
Licensed Content Volume	166
Licensed Content Issue	2
Licensed Content Pages	11
Start Page	506
End Page	516
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of new work	Quantifying Breast Cancer Adaptation: Platforms for Studying Transcriptional Changes in Single Cells
Institution name	Imperial College London
Expected presentation date	Nov 2024
Portions	Figure 1. FluidFM-Based Extraction Approach. (B) Scanning electron micrographs of a FluidFM probe
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, other W12 0NN United Kingdom Attn: Miss Debjani Saha
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP
Terms and Conditions	

INTRODUCTION

Supplementary CP.9

Copyright permission for “Nanoscale tweezers for single-cell biopsies”.[1]

Marketplace

This is a License Agreement between Debjani Saha (“User”) and Copyright Clearance Center, Inc. (“CCC”) on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	28-Sep-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1530834-1		
ISSN	1748-3395	Publisher	Nature Research
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Nature nanotechnology	Publication Type	e-Journal
Article Title	Nanoscale tweezers for single-cell biopsies.	Start Page	80
		End Page	88
Date	01/01/2006	Issue	1
Language	English	Volume	14
Country	United States of America	URL	http://www.nature.com/nnano
Rightsholder	Springer Nature BV		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of Images / Photos / Illustrations	4	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the Disabled?	No
		Minor Editing Privileges?	Yes
Who Will Republish the Content?	Academic institution	Incidental Promotional Use?	No
Duration of Use	Life of current edition	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Professor	Institution Name	Imperial College London
Instructor Name	Luca Magnani	Expected Presentation Date	2024-11-01

ADDITIONAL DETAILS

The Requesting Person / Organization to Appear on the License	Debjani Saha

REQUESTED CONTENT DETAILS

Supplementary CP.10

Copyright permission for “Extracting the contents of living cells”.[93]

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE
TERMS AND CONDITIONS

Aug 10, 2025

This Agreement between Debjani Saha ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	6001450804146
License date	Aug 01, 2025
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Extracting the contents of living cells
Licensed Content Author	Stuart G. Higgins, Molly M. Stevens
Licensed Content Date	Apr 28, 2017
Licensed Content Volume	356
Licensed Content Issue	6336
Volume number	356
Issue number	6336
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Electronic
Portion	Text Excerpt
Number of pages requested	1
Title of new work	QUANTIFYING BREAST CANCER ADAPTATION: PLATFORMS FOR STUDYING TRANSCRIPTIONAL CHANGES IN SINGLE CELLS
Institution name	Imperial College London
Expected presentation date	Aug 2025
Portions	Figure: Four ways for removing cell contents
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Miss. Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, W12 0NN United Kingdom
Billing Type	Invoice
Billing Address	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, United Kingdom W12 0NN
Total	0.00 USD

Terms and Conditions

American Association for the Advancement of Science ("AAAS") TERMS AND CONDITIONS

Regarding your request, we are pleased to grant you non-exclusive, non-transferable permission, to republish the AAAS material subject to the terms and conditions herein. We must be contacted for permission for any uses other than those specifically identified in your request.

Credit/Copyright notice must be provided according to current industry standards and should include "Reproduced/modified from [Lead Author] et al., Science Journal Title, DOI number (Year), AAAS."

All required credit lines and notices must be visible any time a user accesses any part of the AAAS material and must appear on any printed copies or authorized user might make.

This permission does not apply to articles/text/figures/photos/artwork or any other content or materials included in your work that are credited to non-AAAS sources. You are responsible for ensuring that the material you wish to use is not controlled by another entity or third-party. If the requested material is sourced to or references non-AAAS sources, you are responsible for obtaining authorization from that other entity or third-party source before using that material. You agree to hold harmless and indemnify AAAS against any claims arising from your use of any content in your work that credits non-AAAS sources.

If the AAAS material covered by this permission was published in Science during the years 1974 - 1994, you must also obtain permission from the author, who may grant or withhold permission, and who may or may not charge a fee if permission is granted. See original article for author's address. This condition does not apply to news articles.

Modifications and adaptations to the AAAS material require permission from the AAAS journal article author. Author permission for any such changes must be secured prior to your use. Minor edits to charts, graphs, figures or tables that do not alter the meaning, purpose, or context of the AAAS material and that are needs, for example to fit to house style, do not require permission from the AAAS journal article author.

Wherever possible, we ask that electronic uses of the AAAS material permitted herein include a hyperlink to the original work on AAAS's website (hyperlink may be embedded in the reference citation).

AAAS material reproduced in your work identified herein must not account for more than 30% of the total contents of that work.

Your use of AAAS material must not imply any endorsement by the American Association for the Advancement of Science.

This permission is not valid for the use of AAAS and/or AAAS journal logos.

AAAS makes no representations or warranties as to the accuracy of any information contained in the AAAS material covered by this permission, including any warranties of merchantability or fitness for a particular purpose.

If permission fees for this use are waived, please note that AAAS reserves the right to charge for reproduction of this material in the future.

Permission is not valid unless payment is received within sixty (60) days of the issuance of this permission. If payment is not received within this time period then all rights granted herein shall be revoked and this permission will be considered null and void.

In the event of breach of any of the terms and conditions herein or any of CCC's Billing and Payment terms and conditions, all rights granted herein shall be revoked and this permission will be considered null and void.

AAAS reserves the right to terminate this permission and all rights granted herein at its discretion, for any purpose, at any time. In the event that AAAS elects to terminate this permission, you will have no further right to publish, publicly perform, publicly display, distribute or otherwise use any matter in which the AAAS material had been included, and all fees paid hereunder shall be fully refunded to you. Notification of termination will be sent to the contact information as supplied by you during the request process and termination shall be immediate upon sending the notice. Neither AAAS nor CCC shall be liable for any costs, expenses, or damages you may incur as a result of the termination of this permission, beyond the refund noted above.

This Permission may not be amended except by written document signed by both parties.

The terms above are applicable to all permissions granted for the use of AAAS material. Below you will find additional conditions that apply to your particular type of use.

FOR A THESIS OR DISSERTATION

If you are using figure(s)/table(s), permission is granted for use in print and electronic versions of your dissertation or thesis. A full text article may be used in print versions only of a dissertation or thesis.

Permission covers the distribution of your dissertation or thesis on demand by ProQuest provided the AAAS material covered by this permission remains in situ.

If you are an original author on the AAAS material being reproduced, please refer to your License to Publish for rules on reproducing your paper in a dissertation or thesis.

FOR JOURNALS/MAGAZINES:

Permission for use of figures/tables in a journal/magazine includes world-wide rights for use in both print and electronic versions of your journal/magazine. The AAAS material may not be used in any manner other than within the context of your article.

FOR BOOKS/TEXTBOOKS:

Permission for use of figures/tables, in a book/textbook includes world-wide rights in all languages for both print and electronic formats (electronic formats are defined below).

Permission for use of a text excerpt or a full text article includes world-wide rights in English only. You have the option of securing either print or electronic rights or both, but electronic rights are not automatically granted and are subject to additional fees. Permission for translations of text excerpts or full text articles into other languages must be obtained separately.

Permission for use of AAAS material in electronic format books/textbooks is granted only in cases where the electronic version is equivalent to or substitutes for the print version of the book/textbook. The AAAS material reproduced as permitted herein must remain in situ and must not be exploited separately (for example, if permission covers the use of a full text article, the article may not be offered for access or for purchase as a stand-alone unit), except in the case of permitted textbook companions as noted below.

You must include the following notice in any electronic versions, either adjacent to the reproduced AAAS material or in the terms and conditions for use of your electronic products: "Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for non-commercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from the publisher."

If your book is an academic textbook, permission covers the following companions to your textbook, provided such companions are distributed only in conjunction with your textbook at no additional cost to the user:

- Password-protected website
- Instructor guides
- Testing materials
- Student resources

All companions must contain instructions to users that the AAAS material may be used for non-commercial, classroom purposes only. Any other uses require the prior written permission from AAAS.

If your license is for the use of AAAS figures/tables, then the electronic rights granted herein permit use of the AAAS material in any custom databases that you distribute (the electronic versions of your textbook through, so long as the AAAS material remains within the context of a chapter of the title identified in your request and cannot be downloaded by a user as an independent image file.

Rights also extend to copies/files of your work that you are required to provide for use by the visually and/or print disabled in compliance with state and federal laws.

This permission only covers a single edition of your work as identified in your request.

FOR NEWSLETTERS:

Permission covers print and/or electronic versions of your Newsletter, provided the AAAS material reproduced remains in situ and is not exploited separately from your Newsletter issue (for example, if permission covers the use of a full text article, the article may not be offered for access or for purchase as a stand-alone unit).

FOR ANNUAL REPORTS:

Permission covers print and electronic versions of your Annual Report provided the AAAS material reproduced remains in situ and is not exploited separately from your Annual Report (for example, if permission covers the use of a full text article, the article may not be offered for access or for purchase as a stand-alone unit).

FOR PROMOTIONAL/MARKETING USES:

Permission covers the use of AAAS material in promotional or marketing pieces such as information packets, media kits, product slide kits, brochures, or flyers limited to a single print run. The AAAS material may not be used in any manner which implies endorsement or promotion by the American Association for the Advancement of Science (AAAS) or of any AAAS journal any product or service. AAAS does not permit the reproduction of its name, logo or text on promotional literature.

If permission to use a full text article is permitted, the AAAS journal article covered by this permission must not be altered in any way. No additional printing may be set onto an article copy other than the copyright credit line required above. Any alterations must be approved in advance and in writing by AAAS. This includes, but is not limited to, the placement of sponsorship identifiers, trademarks, logos, rubber stamping or self-adhesive stickers onto article copies.

Additionally, article copies must be a freestanding part of any information package [i.e. needle slit into which they are inserted. They may not be physically attached to anything, such as an advertising insert, or have anything attached to them, such as a sample product. Article copies must be easily removable from any kits or informational packages in which they are used. The only exception is that article copies may be inserted into three-ring binders.

If you are using AAAS material in presentation slides, the required credit line must be visible on the slide where the AAAS material is reproduced.

FOR CORPORATE INTERNAL USE:

The AAAS material covered by this permission may not be altered in any way. No additional printing may be set onto an article copy other than the required credit line. Any alterations must be approved in advance and in writing by AAAS. This includes but is not limited to the placement of sponsorship identifiers, trademarks, logos, rubber stamping or self-adhesive stickers onto article copies.

If you are making article copies, copies are restricted to the number indicated in your request and must be distributed only to internal employees for internal use.

If you are using AAAS material in presentation slides, the required credit line must be visible on the slide where the AAAS material is reproduced.

Supplementary CP.11

Copyright permission for “Electrical forces for microscale cell manipulation”.[95]

Marketplace

This is a License Agreement between Debjani Saha (“User”) and Copyright Clearance Center, Inc. (“CCC”) on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	30-Oct-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1541321-1		
ISSN	1545-4274	Publisher	ANNUAL REVIEW OF BIOMEDICAL ENGINEERING
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Annual review of biomedical engineering	Publication Type	e-Journal
		Start Page	425
Article Title	Electrical forces for microscale cell manipulation.	End Page	454
		Issue	1
Date	01/01/1999	Volume	8
Language	English	URL	http://arjournals.annualreviews.org/loi/bioeng
Country	United States of America		
Rightsholder	Annual Reviews, Inc.		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of Images / Photos / Illustrations	1	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the Disabled?	No
		Minor Editing Privileges?	Yes
Who Will Republish the Content?	Academic institution	Incidental Promotional Use?	No
Duration of Use	Life of current and all future editions	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	QUANTIFYING BREAST CANCER ADAPTATION: PLATFORMS FOR STUDYING TRANSCRIPTIONAL CHANGES IN SINGLE CELLS	Institution Name	Imperial College London
		Expected Presentation Date	2024-11-02
Instructor Name	Luca Magnani		

ADDITIONAL DETAILS

Supplementary CP.12

Copyright permission for “Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches”.[132]

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Oct 05, 2024

This Agreement between Debjani Saha ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5882540737709
License date	Oct 05, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Pflügers Archiv European Journal of Physiology
Licensed Content Title	Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches
Licensed Content Author	O. P. Hamill et al
Licensed Content Date	Jan 1, 1981
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Quantifying Breast Cancer Adaptation: Platforms for Studying Transcriptional Changes in Single Cells
Institution name	Imperial College London
Expected presentation date	Nov 2024
Portions	Fig. 9. Schematic representation of the procedures which lead to recording configurations - Image on page 93
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, other W12 0NN United Kingdom Attn: Miss Debjani Saha
Billing Type	Invoice
Billing Address	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, United Kingdom W12 0NN Attn: Miss Debjani Saha
Total	**0.00 GBP**
Terms and Conditions	

Springer Nature Customer Service Centre GmbH Terms and Conditions

Supplementary CP.13

Copyright permission for “Nanoscale live-cell imaging using hopping probe ion conductance microscopy”.[136]

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Oct 05, 2024

This Agreement between Debjani Saha ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5882540995029
License date	Oct 05, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Methods
Licensed Content Title	Nanoscale live-cell imaging using hopping probe ion conductance microscopy
Licensed Content Author	Pavel Novak et al
Licensed Content Date	Mar 1, 2009
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Quantifying Breast Cancer Adaptation: Platforms for Studying Transcriptional Changes in Single Cells
Institution name	Imperial College London
Expected presentation date	Nov 2024
Portions	Figure 1: Principles of adaptive HPICM. (a), (b), (e).
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, other W12 0NN United Kingdom Attn: Miss Debjani Saha
Billing Type	Invoice
Billing Address	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, United Kingdom W12 0NN Attn: Miss Debjani Saha
Total	**0.00 GBP**
Terms and Conditions	

Springer Nature Customer Service Centre GmbH Terms and Conditions

Supplementary CP.14

Copyright permission for “High-content single-cell combinatorial indexing”.[171]

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Oct 18, 2024

This Agreement between Debjani Saha ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5891951463709
License date	Oct 18, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Biotechnology
Licensed Content Title	High-content single-cell combinatorial indexing
Licensed Content Author	Ryan M. Mulqueen et al
Licensed Content Date	Jul 5, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Quantifying Breast Cancer Adaptation: Platforms for Studying Transcriptional Changes in Single Cells
Institution name	Imperial College London
Expected presentation date	Nov 2024
Portions	Fig. 1: Symmetrical strand single-cell combinatorial indexing ATAC-seq (s3-ATAC) improves molecular capture rate.
The Requesting Person / Organization to Appear on the License	Debjani Saha
Requestor Location	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, other W12 0NN United Kingdom Attn: Miss Debjani Saha
Billing Type	Invoice
Billing Address	Debjani Saha Imperial College Hammersmith Campus 72 Du Cane Road London, United Kingdom W12 0NN Attn: Debjani Saha
Total	**0.00 GBP**
Terms and Conditions	

REFERENCES

1. Nadappuram, B. P. *et al*. Nanoscale tweezers for single-cell biopsies. *Nature Nanotech* **14**, 80–88 (2019).

2. Rosano, D. *et al*. Long-term Multimodal Recording Reveals Epigenetic Adaptation Routes in Dormant Breast Cancer Cells. *Cancer Discovery* **14**, 866–889 (2024).

3. Bray, F., Laversanne, M., Weiderpass, E. & Soerjomataram, I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer* **127**, 3029–3030 (2021).

4. Sung, H. *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209–249 (2021).

5. Global Health Estimates: Life expectancy and leading causes of death and disability. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates (2024).

6. Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* **91**, 1983–1989 (2004).

7. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat Med* **10**, 789–799 (2004).

8. Ruddon, R. W. *Cancer Biology*. (Oxford University Press, USA, 2007).

9. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).

10. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31–46 (2022).

11. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

12. Smittenaar, C. R., Petersen, K. A., Stewart, K. & Moitt, N. Cancer incidence and mortality projections in the UK until 2035. *Br J Cancer* **115**, 1147–1155 (2016).

13. Honrado, E., Benítez, J. & Palacios, J. The Pathology of Hereditary Breast Cancer. *Hered Cancer Clin Pract* **2**, 131 (2004).

14. Mavaddat, N., Antoniou, A. C., Easton, D. F. & Garcia-Closas, M. Genetic susceptibility to breast cancer. *Molecular Oncology* **4**, 174–191 (2010).

15. Kuchenbaecker, K. B. *et al*. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* **317**, 2402–2416 (2017).

16. Nik-Zainal, S. *et al*. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

17. Cancer, C. G. on H. F. in B. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet. Oncology* **13**, 1141 (2012).

18. Harbeck, N. *et al*. Breast cancer. *Nat Rev Dis Primers* **5**, 1–31 (2019).

19. Brisken, C. & Duss, S. Stem cells and the stem cell niche in the breast: an integrated hormonal and developmental perspective. *Stem Cell Rev* **3**, 147–156 (2007).

20. Musey, V. C., Collins, D. C., Musey, P. I., Martino-Saltzman, D. & Preedy, J. R. Long-term effect of a first pregnancy on the secretion of prolactin. *N Engl J Med* **316**, 229–234 (1987).

21. Polyak, K. Heterogeneity in breast cancer. *J Clin Invest* **121**, 3786–3788 (2011).

22. Gaudet, M. M. *et al*. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. *Breast Cancer Res Treat* **130**, 587–597 (2011).

23. Patel, H. K. & Bihani, T. Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacology & Therapeutics* **186**, 1–24 (2018).

24. Vu, T. & Claret, F. X. Trastuzumab: Updated Mechanisms of Action and Resistance in Breast Cancer. *Frontiers in Oncology* **2**, 62 (2012).

25. Tavčar Kunstič, T., Debeljak, N. & Fon Tacer, K. Heterogeneity in hormone-dependent breast cancer and therapy: Steroid hormones, HER2, melanoma antigens, and cannabinoid receptors. *Advances in Cancer Biology - Metastasis* **7**, 100086 (2023).

26. Parikh, U., Chhor, C. M. & Mercado, C. L. Ductal Carcinoma In Situ: The Whole Truth. *American Journal of Roentgenology* **210**, 246–255 (2018).

27. Wilson, N., Ironside, A., Diana, A. & Oikonomidou, O. Lobular Breast Cancer: A Review. *Front. Oncol.* **10**, (2021).

28. Platet, N., Cathiard, A. M., Gleizes, M. & Garcia, M. Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. *Critical Reviews in Oncology/Hematology* **51**, 55–67 (2004).

29. Lewis-Wambi, J. S. & Jordan, V. C. Estrogen regulation of apoptosis: how can one hormone stimulate and inhibit? *Breast Cancer Res* **11**, 206 (2009).

30. Baumgarten, S. C. & Frasor, J. Minireview: Inflammation: An Instigator of More Aggressive Estrogen Receptor (ER) Positive Breast Cancers. *Molecular Endocrinology* **26**, 360–371 (2012).

31. Bocchinfuso, W. P. *et al*. Induction of Mammary Gland Development in Estrogen Receptor-α Knockout Mice. *Endocrinology* **141**, 2982–2994 (2000).

32. Quaynor, S. D. *et al*. Delayed Puberty and Estrogen Resistance in a Woman with Estrogen Receptor α Variant. *New England Journal of Medicine* **369**, 164–171 (2013).

33. Paterni, I., Granchi, C., Katzenellenbogen, J. A. & Minutolo, F. Estrogen Receptors Alpha (ERα) and Beta (ERβ): Subtype-Selective Ligands and Clinical Potential. *Steroids* **0**, 13 (2014).

34. Klinge, C. M. Estrogen receptor interaction with co-activators and co-repressors☆. *Steroids* **65**, 227–251 (2000).

35. Pavithran, H. & Kumavath, R. Emerging role of pioneer transcription factors in targeted ER positive breast cancer. *Explor Target Antitumor Ther.* **2**, 26–35 (2021).

36. Eeckhoute, J., Carroll, J. S., Geistlinger, T. R., Torres-Arzayus, M. I. & Brown, M. A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. *Genes Dev.* **20**, 2513–2526 (2006).

37. Aromatase Inhibitors: Benefits, Side Effects, and More. https://www.breastcancer.org/treatment/hormonal-therapy/aromatase-inhibitors (2024).

38. Gonadotropin Releasing Hormone (GnRH) Analogues. in *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury* (National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda (MD), 2012).

39. Fulvestrant. https://go.drugbank.com/drugs/DB00947.

40. Osborne, C. K., Wakeling, A. & Nicholson, R. I. Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action. *British Journal of Cancer* **90**, S2 (2004).

41. Mechanisms of tamoxifen resistance in: Endocrine-Related Cancer Volume 11 Issue 4 (2004). https://erc.bioscientifica.com/view/journals/erc/11/4/0110643.xml?body=fullhtml-62568.

42. Burstein, H. J. Systemic Therapy for Estrogen Receptor–Positive, HER2-Negative Breast Cancer. *New England Journal of Medicine* **383**, 2557–2570 (2020).

43. Pan, H. *et al.* 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *New England Journal of Medicine* **377**, 1836–1846 (2017).

44. Bekes, I. & Huober, J. Extended Adjuvant Endocrine Therapy in Early Breast Cancer Patients—Review and Perspectives. *Cancers* **15**, 4190 (2023).

45. Slamon, D. *et al.* Ribociclib plus Endocrine Therapy in Early Breast Cancer. *New England Journal of Medicine* **390**, 1080–1091 (2024).

46. Johnston, S. *et al.* MONARCH 3 final PFS: a randomized study of abemaciclib as initial therapy for advanced breast cancer. *npj Breast Cancer* **5**, 1–8 (2019).

47. Acquired CYP19A1 amplification is an early specific mechanism of aromatase inhibitor resistance in ERα metastatic breast cancer - PubMed. https://pubmed.ncbi.nlm.nih.gov/28112739/.

48. First-in-Human Phase I Study of the Tamoxifen Metabolite Z-Endoxifen in Women With Endocrine-Refractory Metastatic Breast Cancer | Journal of Clinical Oncology. https://ascopubs.org/doi/10.1200/JCO.2017.73.3246?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

49. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer | Nature Genetics. https://www.nature.com/articles/ng.2822.

50. Razavi, P. *et al*. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438.e6 (2018).

51. PhD, S. L. Endocrine Therapy in Breast Cancer: Making Sense of the 'Word Salad'. *American Association for Cancer Research (AACR)* https://www.aacr.org/blog/2022/07/18/endocrine-therapy-in-breast-cancer-making-sense-of-the-word-salad/ (2022).

52. Yates, L. R. *et al*. Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32**, 169-184.e7 (2017).

53. Schmidt-Kittler, O. *et al*. From latent disseminated cells to overt metastasis: Genetic analysis of systemic breast cancer progression. *Proceedings of the National Academy of Sciences* **100**, 7737–7742 (2003).

54. Hanrahan, E. O. *et al*. Combined-modality treatment for isolated recurrences of breast carcinoma. *Cancer* **104**, 1158–1171 (2005).

55. Barclay, N. L. *et al*. Trends in incidence, prevalence, and survival of breast cancer in the United Kingdom from 2000 to 2021. *Sci Rep* **14**, 19069 (2024).

56. Friedman, R. Drug resistance in cancer: molecular evolution and compensatory proliferation. *Oncotarget* **7**, 11746–11755 (2016).

57. Schmidt-Kittler, O. *et al*. From latent disseminated cells to overt metastasis: Genetic analysis of systemic breast cancer progression. *Proceedings of the National Academy of Sciences* **100**, 7737–7742 (2003).

58. Magnani, L. *et al*. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proceedings of the National Academy of Sciences* **110**, E1490–E1499 (2013).

59. Nguyen, V. T. M. *et al*. Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. *Nat Commun* **6**, 10044 (2015).

60. Hinohara, K. *et al*. KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell* **34**, 939-953.e9 (2018).

61. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

62. Clyde A Hutchison, I. I. I. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research* **35**, 6227 (2007).

63. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**, 560–564 (1977).

64. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**, 5463–5467 (1977).

65. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).

66. Nyrén, P. Enzymatic method for continuous monitoring of DNA polymerase activity. *Analytical Biochemistry* **167**, 235–238 (1987).

67. Cuber, P. *et al*. Comparing the accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific Biosciences, for DNA barcode sequencing applications. *Ecological Genetics and Genomics* **28**, 100181 (2023).

68. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17**, 175–188 (2016).

69. Holley, R. W. *et al*. Structure of a Ribonucleic Acid. *Science* **147**, 1462–1465 (1965).

70. Emrich, S. J., Barbazuk, W. B., Li, L. & Schnable, P. S. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res.* **17**, 69–73 (2007).

71. Li, X. & Wang, C.-Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* **13**, 1–6 (2021).

72. Hagemann-Jensen, M. *et al*. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol* **38**, 708–714 (2020).

73. Mahat, D. B. *et al*. Single-cell nascent RNA sequencing unveils coordinated global transcription. *Nature* **631**, 216–223 (2024).

74. Eng, C.-H. L. *et al*. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).

75. NanoString: Experience the Power of Spatial Biology. *NanoString* https://nanostring.com/ (2024).

76. Shevchuk, A. I. *et al*. Simultaneous Measurement of Ca2+ and Cellular Dynamics: Combined Scanning Ion Conductance and Optical Microscopy to Study Contracting Cardiac Myocytes. *Biophysical Journal* **81**, 1759–1764 (2001).

77. Laforge, F. O., Carpino, J., Rotenberg, S. A. & Mirkin, M. V. Electrochemical attosyringe. *Proceedings of the National Academy of Sciences* **104**, 11895–11900 (2007).

78. Dale, S. E. C. & Unwin, P. R. Polarised liquid/liquid micro-interfaces move during charge transfer. *Electrochemistry Communications* **10**, 723–726 (2008).

79. Saha-Shah, A. *et al*. Nanopipettes: probes for local sample analysis. *Chem. Sci.* **6**, 3334–3341 (2015).

80. Novak, P. Nanoscale-Targeted Patch-Clamp Recordings of Functional Presynaptic Ion Channels.

81. Jönsson, P. *et al*. Hydrodynamic trapping of molecules in lipid bilayers. *COMPUTATIONAL BIOLOGY*.

82. Zhang, Y. *et al*. High-resolution label-free 3D mapping of extracellular pH of single living cells. *Nat Commun* **10**, 5610 (2019).

83. Dufrêne, Y. F. *et al*. Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nature Nanotech* **12**, 295–307 (2017).

84. Putman, C. A. J., Van der Werf, K. O., De Grooth, B. G., Van Hulst, N. F. & Greve, J. Tapping mode atomic force microscopy in liquid. *Applied Physics Letters* **64**, 2454–2456 (1994).

85. Osada, T., Uehara, H., Kim, H. & Ikai, A. mRNA analysis of single living cells. *J Nanobiotechnol* **1**, 2 (2003).

86. Meister, A. *et al*. FluidFM: Combining Atomic Force Microscopy and Nanofluidics in a Universal Liquid Delivery System for Single Cell Applications and Beyond. *Nano Lett.* **9**, 2501–2507 (2009).

87. Guillaume-Gentil, O. *et al*. Tunable Single-Cell Extraction for Molecular Analyses. *Cell* **166**, 506–516 (2016).

88. Chen, W. *et al*. Live-seq enables temporal transcriptomic recording of single cells. *Nature* **608**, 733–740 (2022).

89. Actis, P. *et al*. Compartmental Genomics in Living Cells Revealed by Single-Cell Nanobiopsy. *ACS Nano* **8**, 546–553 (2014).

90. Guillaume-Gentil, O. *et al*. Tunable Single-Cell Extraction for Molecular Analyses. *Cell* **166**, 506–516 (2016).

91. Singhal, R. *et al*. Multifunctional carbon-nanotube cellular endoscopes. *Nature Nanotech* **6**, 57–64 (2011).

92. Xu, A. M. *et al*. Quantification of nanowire penetration into living cells. *Nat Commun* **5**, 3613 (2014).

93. Higgins, S. G. & Stevens, M. M. Extracting the contents of living cells. *Science* **356**, 379–380 (2017).

94. Sahota, A. *et al*. Spatial and Temporal Single-Cell Profiling of RNA Compartmentalization in Neurons with Nanotweezers. *ACS Nano* **19**, 18522–18533 (2025).

95. Voldman, J. ELECTRICAL FORCES FOR MICROSCALE CELL MANIPULATION. *Annual Review of Biomedical Engineering* **8**, 425–454 (2006).

96. Pohl, H. A. The Motion and Precipitation of Suspensoids in Divergent Electric Fields. *Journal of Applied Physics* **22**, 869–871 (1951).

97. Pethig, R. Dielectrophoresis: Using Inhomogeneous AC Electrical Fields to Separate and Manipulate Cells. *Critical Reviews in Biotechnology* **16**, 331–348 (1996).

98. Bahaj, A. S. & Bailey, A. G. Dielectrophoresis of microscopic particles. *J. Phys. D: Appl. Phys.* **12**, L109 (1979).

99. Dielectrophoresis: Theoretical and Practical Considerations. in *Dielectrophoresis* 245–308 (John Wiley & Sons, Ltd, 2017). doi:10.1002/9781118671443.ch10.

100. Castellanos, A., Ramos, A., González, A., Green, N. G. & Morgan, H. Electrohydrodynamics and dielectrophoresis in microsystems: scaling laws. *J. Phys. D: Appl. Phys.* **36**, 2584–2597 (2003).

101. Pethig, R. Dielectrophoresis: Using Inhomogeneous AC Electrical Fields to Separate and Manipulate Cells. *Critical Reviews in Biotechnology* **16**, 331–348 (1996).

102. Tuukkanen, S. *et al*. Carbon Nanotubes as Electrodes for Dielectrophoresis of DNA. *Nano Lett.* **6**, 1339–1343 (2006).

103. Freedman, K. J. *et al*. Nanopore sensing at ultra-low concentrations using single-molecule dielectrophoretic trapping. *Nat Commun* **7**, 10217 (2016).

104. Wang, Y., Du, F., Baune, M. & Thöming, J. Dielectrophoresis in aqueous suspension: impact of electrode configuration. *Microfluid Nanofluid* **17**, 499–507 (2014).

105. Modarres, P. & Tabrizian, M. Alternating current dielectrophoresis of biomacromolecules: The interplay of electrokinetic effects. *Sensors and Actuators B: Chemical* **252**, 391–408 (2017).

106. Dávila, G. C., Gómez-Quiñones, J. I., Perez-Gonzalez, V. H. & Martinez-Duarte, R. Assessing the Advantages of Using Square Wave Signals for Particle Trapping in Carbon-Electrode Dielectrophoresis. *ECS Trans.* **72**, 105 (2016).

107. NanoDrop-2000-User-Manual-EN.pdf.

108. TG2000 Series DDS Function Generator, Digital Control | Aim-TTi. https://www.aimtti.com/product-category/function-generators/aim-tg2000series.

109. Wang, F. *et al*. RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *The Journal of Molecular Diagnostics : JMD* **14**, 22 (2012).

110. 2019 Feb 26 - Tips and Tricks for Running the RNAscope™ Multiplex Fluorescent Assay | In Situ Hybridization, RNA-ISH | ACDBio. https://acdbio.com/2019-feb-26-tips-and-tricks-running-rnascope%E2%84%A2-multiplex-fluorescent-assay.

111.Sandford, C. *et al*. A synthetic chemist's guide to electroanalytical tools for studying reaction mechanisms. *Chem. Sci.* **10**, 6404–6422 (2019).

112.Banks, C. E., Compton, R. G., Fisher, A. C. & Henley, I. E. The transport limited currents at insonated electrodes. *Phys. Chem. Chem. Phys.* **6**, 3147–3152 (2004).

113.Batey, R. T., Rambo, R. P. & Doudna, J. A. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed.* **38**, 2326–2343 (1999).

114.Tsuchiya, K. *et al*. Disclosure of a pro-apoptotic glyceraldehyde-3-phosphate dehydrogenase promoter: anti-dementia drugs depress its activation in apoptosis. *Life Sciences* **74**, 3245–3258 (2004).

115.Graven, K. K., Yu, Q., Pan, D., Roncarati, J. S. & Farber, H. W. Identification of an oxygen responsive enhancer element in the glyceraldehyde-3-phosphate dehydrogenase gene. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1447**, 208–218 (1999).

116.Sirover, M. A. New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells. *Journal of Cellular Biochemistry* **95**, 45–52 (2005).

117.Altenberg, B. & Greulich, K. O. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* **84**, 1014–1020 (2004).

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

119.Bankhead, P. *et al*. QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).

120.Lamprecht, M. R., Sabatini, D. M. & Carpenter, A. E. CellProfiler™: Free, Versatile Software for Automated Biological Image Analysis. *BioTechniques* **42**, 71–75 (2007).

121.RNAscope-AI-App-Note.pdf.

122.Moon, H.-S. *et al*. Continual collection and re-separation of circulating tumor cells from blood using multi-stage multi-orifice flow fractionation. *Biomicrofluidics* **7**, 014105 (2013).

123.Shah, K. N. & Faridi, J. S. Estrogen, tamoxifen, and Akt modulate expression of putative housekeeping genes in breast cancer cells. *The Journal of Steroid Biochemistry and Molecular Biology* **125**, 219–225 (2011).

124.Seger, R. A. *et al*. Voltage controlled nano-injection system for single-cell surgery. *Nanoscale* **4**, 5843–5846 (2012).

125.Marcuccio, F. *et al*. Single-cell nanobiopsy enables multigenerational longitudinal transcriptomics of cancer cells. *Sci. Adv.* **10**, eadl0515 (2024).

126.Marcuccio, F. *et al*. Delivery of Biomolecules into Individual Cells and Subcellular Compartments by Localized Electroporation via Nanopipette. *ACS Nanosci. Au* (2025) doi:10.1021/acsnanoscienceau.5c00053.

127.Neher, E., Sakmann, B. & Steinbach, J. H. The extracellular patch clamp: A method for resolving currents through individual open channels in biological membranes. *Pflugers Arch.* **375**, 219–228 (1978).

128.Neher, E. & Sakmann, B. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260**, 799–802 (1976).

129.Takahashi, Y. *et al*. Multifunctional nanoprobes for nanoscale chemical imaging and localized chemical delivery at surfaces and interfaces. *Angew Chem Int Ed Engl* **50**, 9638–9642 (2011).

130.Clark, A. P. Leak current, even with gigaohm seals, can cause misinterpretation of stem cell-derived cardiomyocyte action potential recordings.

131.Hill, C. L. & Stephens, G. J. An Introduction to Patch Clamp Recording. in *Patch Clamp Electrophysiology: Methods and Protocols* (eds. Dallas, M. & Bell, D.) 1–19 (Springer US, New York, NY, 2021). doi:10.1007/978-1-0716-0818-0_1.

132.Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflugers Arch.* **391**, 85–100 (1981).

133.Zhu, C., Huang, K., Siepser, N. P. & Baker, L. A. Scanning Ion Conductance Microscopy. *Chem Rev* **121**, 11726–11768 (2021).

134.Hansma, P. K., Drake, B., Marti, O., Gould, S. A. C. & Prater, C. B. The Scanning Ion-Conductance Microscope. *Science* **243**, 641–643 (1989).

135.Li, C. *et al*. High resolution imaging using scanning ion conductance microscopy with improved distance feedback control. *Progress in Natural Science* **18**, 671–677 (2008).

136.Novak, P. *et al*. Nanoscale live-cell imaging using hopping probe ion conductance microscopy. *Nat Methods* **6**, 279–281 (2009).

137.Shevchuk, A. *et al*. Angular Approach Scanning Ion Conductance Microscopy. *Biophysical Journal* **110**, 2252–2265 (2016).

138.Kornreich, B. G. The patch clamp technique: Principles and technical considerations. *Journal of Veterinary Cardiology* **9**, 25–37 (2007).

139.Steinbock, L. J., Otto, O., Chimerel, C., Gornall, J. & Keyser, U. F. Detecting DNA Folding with Nanocapillaries. *Nano Lett.* **10**, 2493–2497 (2010).

140.Qiao, G., Duan, W., Chatwin, C., Sinclair, A. & Wang, W. Electrical properties of breast cancer cells from impedance measurement of cell suspensions. *J. Phys.: Conf. Ser.* **224**, 012081 (2010).

141.Gong, X. *et al*. Label-Free In-Flow Detection of Single DNA Molecules using Glass Nanopipettes. *Anal. Chem.* **86**, 835–841 (2014).

142.Suchyna, T. M., Markin, V. S. & Sachs, F. Biophysics and Structure of the Patch and the Gigaseal. *Biophys J* **97**, 738–747 (2009).

143.Zhang, Y. *et al*. High-resolution label-free 3D mapping of extracellular pH of single living cells. *Nat Commun* **10**, 5610 (2019).

144.Schanne, F. A. X., Kane, A. B., Young, E. E. & Farber, J. L. Calcium Dependence of Toxic Cell Death: A Final Common Pathway. *Science* **206**, 700–702 (1979).

145.Orrenius, S., Zhivotovsky, B. & Nicotera, P. Regulation of cell death: the calcium–apoptosis link. *Nat Rev Mol Cell Biol* **4**, 552–565 (2003).

146.Thompson, R. E., Lindau, M. & Webb, W. W. Robust, High-Resolution, Whole Cell Patch-Clamp Capacitance Measurements Using Square Wave Stimulation. *Biophysical Journal* **81**, 937–948 (2001).

147.Babiychuk, E. B., Monastyrskaya, K., Potez, S. & Draeger, A. Blebbing confers resistance against cell lysis. *Cell Death Differ* **18**, 80–89 (2011).

148.Ying, L. *et al*. Frequency and Voltage Dependence of the Dielectrophoretic Trapping of Short Lengths of DNA and dCTP in a Nanopipette. *Biophysical Journal* **86**, 1018–1027 (2004).

149.Takeichi, M. & Okada, T. S. Roles of magnesium and calcium ions in cell-to-substrate adhesion. *Experimental Cell Research* **74**, 51–60 (1972).

150.Compton, J. Nucleic acid sequence-based amplification. *Nature* **350**, 91–92 (1991).

151.Notomi, T. *et al*. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res* **28**, e63 (2000).

152.Zhao, W., Ali, M. M., Brook, M. A. & Li, Y. Rolling Circle Amplification: Applications in Nanotechnology and Biodetection with Functional Nucleic Acids. *Angewandte Chemie International Edition* **47**, 6330–6337 (2008).

153.Wu, Q. *et al*. INSIGHT: A population-scale COVID-19 testing strategy combining point-of-care diagnosis with centralized high-throughput sequencing. *Science Advances* **7**, eabe5054 (2021).

154.Gupta, T. *et al*. Evaluation of real-time PCR as an alternative for potency testing of Brucella abortus vaccines. *Animal Biotechnology* **32**, 591–601 (2021).

155.Varshney, R. *et al*. Development of novel iron-regulated *Pasteurella multocida* B: 2 bacterin and refinement of vaccine quality in terms of minimum variation in particle size and distribution *vis-a-vis* critical level of iron in media. *Microbial Pathogenesis* **147**, 104375 (2020).

156.Kundu, S., Varshney, R. & Sulabh, S. Exploration of isothermal nucleic acid amplification techniques in the biomedical field. *Gene and Genome Editing* **7**, 100032 (2024).

157.Compton, J. Nucleic acid sequence-based amplification. *Nature* **350**, 91–92 (1991).

158.Abd El Galil, K. H. *et al*. Real-Time Nucleic Acid Sequence-Based Amplification Assay for Detection of Hepatitis A Virus. *Appl Environ Microbiol* **71**, 7113–7116 (2005).

159.Deiman, B., van Aarle, P. & Sillekens, P. Characteristics and applications of nucleic acid sequence-based amplification (NASBA). *Mol Biotechnol* **20**, 163–179 (2002).

160.Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat Biotechnol* **14**, 303–308 (1996).

161.Ye, J. *et al*. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).

162.Liu, Z. *et al*. Transient stem-loop structure of nucleic acid template may interfere with polymerase chain reaction through endonuclease activity of *Taq* DNA polymerase. *Gene* **764**, 145095 (2021).

163.Rutledge, R. G. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Research* **31**, 93e–993 (2003).

164.ROGERS-BROADWAY, K.-R. & KARTERIS, E. Amplification efficiency and thermal stability of qPCR instrumentation: Current landscape and future perspectives. *Exp Ther Med* **10**, 1261–1264 (2015).

165.Kimchi, O., King, E. M. & Brenner, M. P. Uncovering the mechanism for aggregation in repeat expanded RNA reveals a reentrant transition. *Nat Commun* **14**, 332 (2023).

166.Wohlschlegel, J. A., Kutok, J. L., Weng, A. P. & Dutta, A. Expression of Geminin as a Marker of Cell Proliferation in Normal Tissues and Malignancies. *The American Journal of Pathology* **161**, 267 (2002).

167.Santos, A., Wernersson, R. & Jensen, L. J. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Research* **43**, D1140–D1144 (2015).

168.Wohlschlegel, J. A., Kutok, J. L., Weng, A. P. & Dutta, A. Expression of Geminin as a Marker of Cell Proliferation in Normal Tissues and Malignancies. *The American Journal of Pathology* **161**, 267 (2002).

169.Hahaut, V. *et al*. Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. *Nat Biotechnol* **40**, 1447–1451 (2022).

170.Salmen, F. *et al*. High-throughput total RNA sequencing in single cells using VASA-seq. *Nat Biotechnol* **40**, 1780–1793 (2022).

171.Mulqueen, R. M. *et al*. High-content single-cell combinatorial indexing. *Nat Biotechnol* **39**, 1574–1580 (2021).

172.Macaulay, I. C. *et al*. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat Protoc* **11**, 2081–2103 (2016).

173.Picelli, S. & Hahaut, V. FLASH-seq UMI protocol. (2022).

174.cutana-cut&run-protocol_1.7.pdf.

175.Picelli, S. & Hahaut, V. FLASH-seq UMI protocol. (2022).

176.Karimian, A., Ahmadi, Y. & Yousefi, B. Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair* **42**, 63–71 (2016).

177.FluoroBrite™ DMEM. https://www.thermofisher.com/order/catalog/product/A1896702.

178.cellenONE® – The open platform for single cell isolation and nanoliter dispensing. *Cellenion* https://www.cellenion.com/products/cellenone/.

179.Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S. & Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* **27**, 2380–2396 (2013).

180.Ma, F. *et al*. A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics* **20**, 9 (2019).

181.Adey, A. C. Tagmentation-based single-cell genomics. *Genome Research* **31**, 1693 (2021).

182.Mönttinen, H. A. M. & Löytynoja, A. Template switching in DNA replication can create and maintain RNA hairpins. *Proceedings of the National Academy of Sciences* **119**, e2107005119 (2022).

183.Ramsköld, D. *et al*. Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. *Nature biotechnology* **30**, *777* (2012).

184.Picelli, S. *et al*. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171–181 (2014).

185.van Holde, K. E. The Nucleosome. in *Chromatin* (ed. van Holde, K. E.) 219–288 (Springer, New York, NY, 1989). doi:10.1007/978-1-4612-3490-6_6.

186.What are the peaks present in my ATAC library trace? *10X Genomics* https://kb.10xgenomics.com/hc/en-us/articles/360028207751-What-are-the-peaks-present-in-my-ATAC-library-trace.

187.Oliphant, M. U. J., Akshinthala, D. & Muthuswamy, S. K. Establishing conditions for the generation and maintenance of estrogen receptor-positive organoid models of breast cancer. *Breast Cancer Research* **26**, 56 (2024).

188.Özdemir, B. C., Sflomos, G. & Brisken, C. The challenges of modeling hormone receptor-positive breast cancer in mice. (2018) doi:10.1530/ERC-18-0063.

189.Quinn, H. M., Battista, L., Scabia, V. & Brisken, C. Preclinical Mouse Intraductal Model (MIND) to Study Metastatic Dormancy in Estrogen Receptor-Positive Breast Cancer. in *Cancer Cell Dormancy: Methods and Protocols* (ed. Montagner, M.) 101–112 (Springer US, New York, NY, 2024). doi:10.1007/978-1-0716-3882-8_7.

190. Stages of cancer. *Cancer Research UK* https://www.cancerresearchuk.org/about-cancer/what-is-cancer/stages-of-cancer (2014).

191. TNM staging for breast cancer. https://www.cancerresearchuk.org/about-cancer/breast-cancer/stages-grades/tnm-staging.

192. Stages and grades of breast cancer. https://www.macmillan.org.uk/cancer-information-and-support/breast-cancer/staging-and-grading-of-breast-cancer.

193. Surakasula, A., Nagarjunapu, G. C. & Raghavaiah, K. V. A comparative study of pre- and post-menopausal breast cancer: Risk factors, presentation, characteristics and management. *Journal of Research in Pharmacy Practice* **3**, 12 (2014).

194. NanoString. Tips When Generating Single-Cell Spatial Data: Sample Selection. *NanoString* https://nanostring.com/blog/tips-when-generating-single-cell-spatial-data-sample-selection-cosmx/ (2024).

195. 2019 Feb 26 - Tips and Tricks for Running the RNAscope™ Multiplex Fluorescent Assay | In Situ Hybridization, RNA-ISH | ACDBio. https://acdbio.com/2019-feb-26-tips-and-tricks-running-rnascope%E2%84%A2-multiplex-fluorescent-assay (2024).

196. Lim, H. J., Wang, Y., Buzdin, A. & Li, X. A practical guide for choosing an optimal spatial transcriptomics technology from seven major commercially available options. *BMC Genomics* **26**, 47 (2025).

197. RNAscope™ HiPlex v2 Assays | In Situ Hybridization, RNA-ISH | ACDBio. https://acdbio.com/rnascope-hiplex-assays.

198. Marianas. *3i* https://www.intelligent-imaging.com/marianas.

199. Characterization Of Breast Cancer Tissue Using Autofluorescence and Fluorescence Spectroscopy In-Vitro. *J COMM MED PUB HEALTH REP* (2024) doi:10.38207/JCMPHR/2024/MAR05050439.

200. Alexa Fluor 488 dye - UK. https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/fluorophores/alexa-fluor-488.html.

201. Castleman, K. R. *Digital Image Processing*. (Prentice Hall Press, USA, 1996).

202. Shaw, P. Deconvolution in 3-D optical microscopy. *Histochem J* **26**, 687–694 (1994).

203. Chaudhary, S., Krishna, B. M. & Mishra, S. K. A novel FOXA1/ESR1 interacting pathway: A study of Oncomine™ breast cancer microarrays. *Oncology Letters* **14**, 1247 (2017).

204. Gulbahce, H. E., Blair, C. K., Sweeney, C. & Salama, M. E. Quantification of Estrogen Receptor Expression in Normal Breast Tissue in Postmenopausal Women With Breast Cancer and Association With Tumor Subtypes. *Applied Immunohistochemistry & Molecular Morphology* **25**, 548 (2017).

205. Thorat, M. A. *et al*. Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. *Journal of Clinical Pathology* **61**, 327–332 (2008).

206.Fu, X. *et al*. High FOXA1 levels induce ER transcriptional reprogramming, a pro-metastatic secretome, and metastasis in endocrine-resistant breast cancer. *Cell Reports* **42**, 112821 (2023).

207.Khoury, W. E. & Nasr, Z. Deregulation of ribosomal proteins in human cancers. *Bioscience Reports* **41**, BSR20211577 (2021).

208.Expression of RPL28 in cancer - Summary - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000108107-RPL28/cancer.