IMPERIAL

Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution

A thesis submitted by:

Elizabeth Flittner

For the degree of Doctor of Philosophy (PhD)

February 2026

Imperial College London

Department of Metabolism, Digestion and Reproduction

Abstract

Cancer progression is driven by dynamic clonal evolution and interactions across molecular scales, yet experimental dissection of these processes is often limited to a single modality. Comprehensive investigation of tumour evolution requires lineage-tracing systems that are scalable, modality-flexible, and compatible with phenotypic and spatial analysis. This thesis describes the development and validation of X-CODE, a modular RNA barcoding framework that anchors lineage with defined, probe-detectable sequences, enabling integration across sequencing-based, cytometric, and spatial imaging platforms.

X-CODE combines an RNA barcode, detectable by probe hybridisation, with a matching short sequencing barcode to support clonal tracking within a unified experimental architecture without constraining downstream analytical modalities. This thesis shows combinatorial generation of high-diversity X-CODE libraries, enabling tracking of large populations. Isotope-based detection of barcodes via in-suspension time-of-flight mass cytometry is demonstrated across library scales, preserving compatibility with high-dimensional antibody phenotyping.

To extend lineage resolution into intact tissues, imaging-based X-CODE detection was implemented using a repurposed Akoya PhenoCycler-Fusion workflow, enabling spatially resolved clonal mapping. Detection of X-CODE by imaging mass cytometry was also established, broadening compatibility across imaging platforms. Together, this establishes a sequencing-independent route for clonal analysis within cytometry and imaging workflows. Furthermore, MALDI mass spectrometry imaging with sequential, same-section X-CODE detection was demonstrated to support integrative, unbiased molecular phenotyping with linked clonal identity.

Applied to an *in vivo* model of androgen deprivation in prostate cancer, X-CODE resolves clonal architecture and selection dynamics underpinning castration resistance with spatial resolution, revealing clone-specific phenotypic and metabolic plasticity within structured tumour ecosystems. Collectively, this work establishes X-CODE as a scalable, experimentally accessible lineage-tracing platform that integrates cytometric, spatial, metabolic, and sequencing readouts, enabling unified analysis of clonal dynamics across molecular and spatial dimensions.

Statement of Originality

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) at Imperial College London. The work presented in this thesis is my own and has not been submitted in whole or in part for any other degree or qualification at this or any other institution.

All research described herein was conducted by me, except where otherwise stated. Contributions of collaborators, supervisors, or others are clearly acknowledged. All sources of information, data, and previously published material have been appropriately cited and referenced.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

I would first like to express my sincere gratitude to my cross-institution supervisory team, Professor Zoltan Takats and Dr Marco Bezzi, for their continued guidance and support throughout this project. Marco's patience and commitment to mentorship were invaluable to my professional and personal development throughout this PhD.

I am deeply grateful to Roberto Rota for his dedication to the X-CODE project. This work would not have been the same without his insight, commitment, and characteristic optimism. I would like to wholeheartedly thank the Tumour Functional Heterogeneity team for their advice, expertise, and willingness to help at every stage of this research.

I am endlessly thankful to my family, friends, and partner for their constant encouragement and kindness, and for being my greatest cheerleaders throughout this journey. A very special thank you to my mother, whose unwavering belief has been the foundation of everything I have achieved.

Finally, I gratefully acknowledge the Convergence Science Centre and Cancer Research UK for their support in enabling this research.

Table of Contents

Chapter 1 Introduction 14

1.1 Cancer as an evolutionary and ecological disease 14

1.1.1 Cellular relationships and biological organisation 14

1.1.2 The evolutionary dynamics of cancer 14

1.1.3 Tumour heterogeneity: genetic and non-genetic diversity 17

1.1.4 Functional cellular states and plasticity in cancer 19

1.1.5 The tumour microenvironment as a selective landscape 20

1.1.6 Studying cancer evolution: resolving clonal dynamics 23

1.2 Lineage tracing and cellular barcoding 24

1.2.1 Lineage tracing in biology: retrospective and prospective approaches 24

1.2.2 Principles of cellular barcoding 27

1.2.3 Static barcodes 31

1.2.4 Evolvable barcodes 38

1.2.5 Molecular Recorders 42

1.2.6 Cellular barcoding: practical considerations and readout constraints 43

1.3 Toward multimodal clonal tracing 46

1.3.1 Beyond single-modality cellular profiling 46

1.3.2 Spatial RNA and protein detection 47

1.3.3 Spatial metabolic and molecular imaging 51

1.3.4 Suspension-based cytometry profiling 52

1.3.5 Integrated multimodal workflows and platforms 54

1.3.6 Persistent constraints in multimodal clonal tracing 59

1.4 Rationale and scope of this thesis 60

1.4.1 Perspectives and outlook 60

1.4.2 X-CODE as a flexible, scalable barcoding framework 61

Chapter 2 Aim and objectives 63

2.1 Objectives 64

Chapter 3 Materials and methods 65

3.1 Core reagents 65

3.2 Oligonucleotides 65

3.2.1 X-CODE barcode oligonucleotides 65

3.2.2 Primers for X-CODE cloning 65

3.2.3 Hybridisation probes 66

3.2.4 Fluorophore detection probes 66

3.2.5 Thiol detection probes 66

3.3 Antibodies 67

3.4 Tissue culture 68

3.4.1 Cell lines 68

3.4.2 Cell culture 68

3.5 Molecular and genetic methods 69

3.5.1 Plasmid transformation and amplification 69

3.5.2 Lentiviral production and transduction 70

3.5.3 Next-generation sequencing 70

3.6 Generation of simple A, B, and C barcodes 70

3.7 Design of the X-CODE construct 71

3.7.1 Design of X-CODE barcode units 71

3.7.2 Design of barcode unit sets 1–3 72

3.8 Generation of the 15K and 2M X-CODE libraries 73

3.8.1 Assembly of the X-CODE library backbone (CROPseq-derived) 73

3.8.2 Assembly of X-CODE libraries 73

3.9 X-CODE clonal retrieval 75

3.10 Animal work 76

3.10.1 TRAMP-C1 and RM-1 *in vivo* experiments 76

3.10.2 MyC-CaP *in vivo* experiments 77

3.10.3 *In vivo* sample collection and processing 77

3.11 Tissue block generation 78

3.11.1 OCT embedding 78

3.11.2 HPMC/PVP embedding 78

3.11.3 Formalin-fixed paraffin embedding 79

3.12 Tissue block sectioning 79

3.12.1 OCT and HPMC/PVP block sectioning 79

3.12.2 FFPE block sectioning 80

3.13 Haematoxylin and eosin staining 80

3.14 Core X-CODE assay buffers 81

3.15 X-CODE detection in suspension 82

3.15.1 Isotope-oligonucleotide conjugation 82

3.15.2 X-CODE CyTOF assay 82

3.15.3 CyTOF sample acquisition 84

3.16 Spatial X-CODE detection 84

3.16.1 Cell or tissue sample preparations 85

3.16.2 X-CODE spatial assay 86
3.16.3 X-CODE spatial detection 87
3.17 MALDI-mass spectrometry imaging 89
3.17.1 DHB matrix application 90
3.17.2 DHAP matrix application 90
3.17.3 Matrix-assisted laser desorption/ionisation 90
3.17.4 Post-MALDI sample preparation for X-CODE spatial assay 91
3.18 Data analysis and visualisation 91
3.18.1 Next-generation sequencing analysis 91
3.18.2 MALDI-MSI analysis 92
3.18.3 X-CODE spatial analysis and debarcoding 94
3.18.4 Data visualisation and figure generation 96
Chapter 4 Design and validation of the X-CODE framework for probe-based detection of combinatorial barcodes 97
4.1 Research in context 97
4.1.1 Chapter aims 99
4.2 Probe-based detection of X-CODE barcode units 99
4.2.1 Detection of barcode units by flow cytometry and fluorescence imaging 100
4.3 Construction of high-complexity combinatorial X-CODE libraries 103
4.3.1 A combinational cloning strategy for library diversification 104
4.4 Assessment of X-CODE library diversity and balance 107
4.5 Functionalisation of the X-CODE construct for live clone retrieval 108
4.6 Discussion 111
4.6.1 Limitations 113
4.6.2 Conclusions 114
Chapter 5 High-parameter suspension detection of X-CODE barcodes by mass cytometry 115
5.1 Research in context 115
5.1.1 Chapter aims 117
5.2 Detection of modular X-CODE barcode units by CyTOF 117
5.3 Assessment of X-CODE co-detection compatibility with antibody staining 118
5.4 Validation of high-complexity X-CODE detection by CyTOF 120
5.4.1 Confirmation of isotope conjugation to X-CODE detection probes 120
5.4.2 Detection of the X-CODE 15K library by barcode units 1–18 121

5.4.3 Detection of the X-CODE 2M library lines by barcode units 1–27 124

5.5 Discussion 126

5.5.1 Limitations 128

5.5.2 Conclusions 129

Chapter 6 High-parameter spatial detection of X-CODE barcodes by multiplexed imaging 130

6.1 Research in context 130

6.1.1 Chapter aims 131

6.2 Optimisation of hybridisation probe conditions for spatial X-CODE detection 132

6.3 Expansion of fluorescence X-CODE detection capacity using PhenoCycler 134

6.4 Adaptation of X-CODE detection for imaging mass cytometry 137

6.5 Comparison of X-CODE detection across tissue embedding methods 138

6.6 Spatial detection of X-CODE barcode libraries in fresh-frozen tissues 141

6.6.1 High-parameter fluorescence imaging for X-CODE detection in fresh-frozen tissues 141

6.6.2 Detection of X-CODE barcodes in fresh-frozen tissues by imaging mass cytometry 143

6.7 Antibody co-detection considerations in fresh-frozen tissue 145

6.8 Discussion 147

6.8.1 Limitations 149

6.8.2 Conclusions 150

Chapter 7 Integration of X-CODE spatial workflows with unbiased mass spectrometry imaging 152

7.1 Research in context 152

7.1.1 Chapter aims 154

7.2 Establishing post-MALDI *in situ* X-CODE RNA detection on the same tissue section 154

7.2.1 X-CODE detection following MALDI-MSI using a DHB matrix 154

7.2.2 X-CODE detection following MALDI-MSI using a DHAP matrix 157

7.3 Balancing X-CODE RNA detection with MALDI-MSI acquisition parameters 160

7.4 Discussion 162

7.4.1 Limitations 166

7.4.2 Conclusions 167

Chapter 8 Application of X-CODE to an in vivo model of prostate cancer evolution 169

8.1 Research in context 169
8.1.1 Chapter aims 171
8.2 Assessment of clonal dynamics in a MyC-CaP tumour model during androgen deprivation 171
8.2.1 Spatial mapping of X-CODE barcodes reveals clonal organisation during tumour evolution 173
8.3 Clone-associated phenotypic and metabolic characterisation during tumour evolution under androgen deprivation 175
8.3.1 Immunophenotypic characterisation of X-CODE–defined clones 175
8.3.2 Integration of X-CODE with MALDI–MSI reveals clone-associated metabolic features 178
8.4 Discussion 183
8.4.1 Limitations 185
8.4.2 Conclusions 186
Chapter 9 Thesis discussion, impact, and outlook *188*
9.1 X-CODE as a scalable, probe-based cellular barcoding framework 188
9.2 X-CODE enables cross-platform barcode retention and analytical flexibility 190
9.3 Multimodal integration: linking clonal identity and phenotype 193
9.4 X-CODE reveals spatial clonal dynamics during tumour evolution 196
Chapter 10 Thesis conclusions *199*
Chapter 11 Supplementary information *201*
11.1 Simple barcode cloning and detection probes 201
11.2 Probes for β-actin detection 202
11.3 X-CODE library cloning 202
11.4 Barcode unit sequences 1–27 203
11.5 Padlock probes 1–27 and universal probe 204
11.6 X-CODE isotope detection probes 205
11.7 X-CODE fluorescent detection probes 207
11.8 Akoya PhenoCycler-Fusion cycle information 208
Rights and Permissions *209*
References *218*

List of Figures

Figure 1-1. Models of tumour evolution. 16
Figure 1-2. Sources of intratumour heterogeneity. 18
Figure 1-3. Components of the tumour microenvironment. 21
Figure 1-4. Prospective and retrospective methods for lineage tracing. 25
Figure 1-5. Static and evolvable barcoding strategies. 29
Figure 1-6. Representative workflow using a static or optical barcode library in cancer studies. 32
Figure 1-7. CRISPR-mediated DNA editing. 42
Figure 1-8. Spatial transcriptomics techniques. 48
Figure 1-9. Targeted spatial protein detection. 50
Figure 1-10. Mass spectrometry imaging techniques: DESI, MALDI, and SIMS. 52
Figure 1-11. Integrative spatial multiomic workflows. 55
Figure 1-12. Generating an integrated, multimodal picture of tumour heterogeneity. 61
Figure 2-1. Representative workflow using the X-CODE framework. 63
Figure 3-1. Sequence level design of an X-CODE barcode unit with padlock probes. 72
Figure 4-1. Probe-based strategies, using amplification, for spatial profiling of RNA. 98
Figure 4-2. Design of a simple barcode for probe-based detection using a limited set of sequences. 100
Figure 4-3. Probe-based detection of modular X-CODE barcode units by flow cytometry and fluorescence imaging. 101
Figure 4-4. Construction of 15K and 2M combinatorial X-CODE libraries. 105
Figure 4-5. Combinatorial cloning generates diverse and balanced X-CODE libraries. 107
Figure 4-6. Live-cell isolation of barcode-defined clones using the X-CODE framework. 109
Figure 5-1. Schematic representation of the Proximity Ligation Assay for RNA (PLAYR). 116
Figure 5-2. Detection of X-CODE barcode units by CyTOF using simple-barcoded cells. 118
Figure 5-3. Validation of X-CODE detection compatibility with immune-phenotyping by CyTOF. 119
Figure 5-4. Assessment of isotope-conjugation to X-CODE detection probes. 120
Figure 5-5. Detection of barcode units 1–18 in 15K X-CODE library barcoded cell lines. 122
Figure 5-6. Signal intensities for barcode units 1–18 in a heterogeneous barcoded TRAMP-C1 15K population. 123
Figure 5-7. Barcode unit detection across known barcode populations from the 2M X-CODE library. 124
Figure 5-8. Detection of barcode units 1–27 across X-CODE 2M library barcoded cell lines. 125
Figure 6-1. Optimisation of padlock probe and universal probe concentrations for spatial X-CODE readout. 132
Figure 6-2. Validation of high-plex fluorescence detection of X-CODE barcode units using Akoya PhenoCycler. 135
Figure 6-3. Highly multiplexed fluorescence detection of barcode units 1-27 by Akoya PhenoCycler-Fusion. 136
Figure 6-4. Protocol adaptation for X-CODE detection using high-parameter imaging mass cytometry in cells. 137

Figure 6-5. Assessment of X-CODE detection across fresh-frozen OCT or HPMC/PVP hydrogel embedded tissue, and FFPE tissue. ---------- 139
Figure 6-6. X-CODE detection of 2M clonal lines and 15K library in OCT and HPMC/PVP embedded tissues by Akoya PhenoCycler. ---------- 142
Figure 6-7. X-CODE 15K library detection in fresh-frozen OCT and HPMC/PVP embedded tissues by imaging mass cytometry. ---------- 144
Figure 6-8. Considerations for co-detection of antibodies with X-CODE staining in fresh-frozen OCT- and HPMC/PVP-embedded tissues. ---------- 146
Figure 7-1. Evaluation of rolling circle-mediated RNA detection of X-CODE barcode units and β-actin following MALDI-MSI using DHB matrix. ---------- 155
Figure 7-2. Assessing detection of X-CODE barcode units and antibody staining following MALDI-MSI using DHAP matrix. ---------- 158
Figure 7-3. Balancing MALDI-MSI data quality with X-CODE RNA detection across laser power settings. ---------- 161
Figure 8-1. Investigating clonal dynamics using X-CODE in a MyC-CaP mouse model of prostate cancer during androgen deprivation. ---------- 172
Figure 8-2. X-CODE resolves clonal dynamics spatially during androgen deprivation in vivo. ---------- 174
Figure 8-3. Immunophenotyping of MyC-CaP tumour states during tumour evolution under androgen deprivation. ---------- 176
Figure 8-4. Clone-associated phenotypic states during tumour evolution under androgen deprivation. - 178
Figure 8-5. X-CODE spot detection and spatial 15K library decoding following MALDI-MSI. ---------- 180
Figure 8-6. Clone-associated metabolic features during tumour evolution under androgen deprivation. 182

List of Tables

Table 1-1. Static and evolving genetic barcoding strategies. ---------- 30
Table 3-1. Antibodies for CyTOF mass cytometry. ---------- 67
Table 3-2. Antibodies for fluorescence microscopy. ---------- 67
Table 3-3. Antibodies for Akoya PhenoCycler-Fusion. ---------- 67
Table 3-4. Core X-CODE assay buffer components. ---------- 81

Supplementary Tables

Supplementary Table 1. Primer sequences for Gibson cloning of barcodes A, B, and C. ----------201
Supplementary Table 2. Barcode A, B, and C padlock and detection probes. ----------201
Supplementary Table 3. Padlock and detection probes for β-actin. ----------202
Supplementary Table 4. Primer sequences for Gibson cloning of X-CODE library backbone. ----------202
Supplementary Table 5. Primer sequences for add-ons of the X-CODE library construct. ----------202
Supplementary Table 6. Barcode sequences 1–27. ----------203
Supplementary Table 7. X-CODE padlock probes 1–27 and the universal probe. ----------204
Supplementary Table 8. Probes 1–18 for isotope detection of the X-CODE 15K library. ----------205
Supplementary Table 9. Probes 1–27 for isotope detection of the X-CODE 2M library. ----------206
Supplementary Table 10. Probes 1–27 for fluorescent detection of the X-CODE libraries. ----------207
Supplementary Table 11. Akoya PhenoCycler-Fusion cycles for X-CODE 15K library readout. ----------208
Supplementary Table 12. Akoya PhenoCycler-Fusion cycles for X-CODE 2M library readout. ----------208

Chapter 1 Introduction

1.1 Cancer as an evolutionary and ecological disease

1.1.1 Cellular relationships and biological organisation

Multicellular organisms are organised systems in which individual cells do not act alone but rather as components of structured, interacting networks. Cellular behaviour is influenced not only by intrinsic genetic programming but also by positional context, lineage history, and interactions with neighbouring cells and the extracellular environment. Tissues maintain architecture and function through coordinated proliferation, differentiation, and communication, forming dynamic yet regulated cellular ecosystems. Therefore, biological organisation emerges from the collective behaviour of cells operating within defined spatial and temporal frameworks [1].

Cancer can be understood as a pathological extension of these same principles. Rather than representing a purely cell-autonomous genetic disorder, tumorigenesis reflects disruption of coordinated cellular systems, in which altered cells proliferate, compete, and adapt within evolving tissue environments [2]. Therefore, interpreting cancer requires consideration of both evolutionary processes operating within cell populations and ecological interactions shaping behaviour within tissues [3]. The following sections examine how these principles underpin contemporary models of tumour evolution and shape our understanding of cancer as a dynamic and adaptive ecosystem.

1.1.2 The evolutionary dynamics of cancer

Cancer progression can be understood through an evolutionary framework in which cells acquire heritable variation, enabling the selection and expansion of cancer cells within the tumour ecosystem [4]. This view parallels Darwin's theory of natural selection: individuals with traits best suited to their environment are more likely to survive and pass on advantageous characteristics to their progeny [5]. Furthermore, tumours face many of the same selection pressures as evolving species; they compete for space and resources, encounter geographic barriers, adapt to changing ecosystems, and evade immune system predation or eradication by therapeutic interventions [3]. Accordingly,

evolutionary principles underpin contemporary interpretations of tumour development, metastatic dissemination, and therapeutic resistance [6].

Peter Nowell formalised the concept of **tumour evolution** in 1976, proposing that most cancers arise from a single neoplastic cell and diversify through sequential selection of somatic mutations [7]. This process drives the expansion of **clones**: descendants that originate from a common ancestor [3]. Early cytogenetic observations revealed substantial genomic diversity within tumours, consistent with ongoing somatic alteration during progression, yet technological limitations constrained detailed molecular characterisation. The advent of next-generation sequencing (NGS) enabled systematic genomic analysis of tumours, revealing that they are not genetically uniform masses but structured populations of related clones. To interpret this clonal complexity, researchers developed evolutionary models describing how mutations accumulate, interact, and shape tumour architecture over time [8].

1.1.2.1 Models of tumour evolution

The simplest model of tumour evolution is **linear**, characterised by sequential accumulation of mutations in which one genotype replaces another within a population [9]. Driver mutations confer a selective advantage, enabling clonal sweeps that outcompete earlier populations (**Figure 1-1**A) [8]. Although linear dynamics may operate in specific contexts, accumulating evidence supports **branched evolutionary trajectories** across many tumour types [5]. In branched models, clones diverge from a common ancestor and evolve in parallel, generating multiple coexisting lineages with variable **fitness**: the ability to survive and proliferate within an ecosystem [8]. These dynamics produce phylogenetic structures composed of distinct subclonal populations detectable at clinical sampling (**Figure 1-1**B).

Early multi-region sequencing studies, including work by *Gerlinger et al.*, demonstrated that 63–96% of somatic mutations were regionally restricted in renal cell carcinoma, providing direct evidence of subclonal diversification [10]. Nonetheless, additional analyses identified patterns consistent with dynamics beyond classical Darwinian principles, such as **neutral evolution** [5]. In this framework, passenger mutations accumulate with limited selective constraint, resulting in extensive genetic drift and comparatively modest differences in clonal fitness [8]. Phylogenetic reconstructions in such tumours often reveal numerous low-frequency lineages (**Figure 1-1**C). Multi-region

sequencing of 300 sites from a hepatocellular carcinoma by *Ling et al.* demonstrated extreme genetic diversity consistent with limited clonal selection [11].

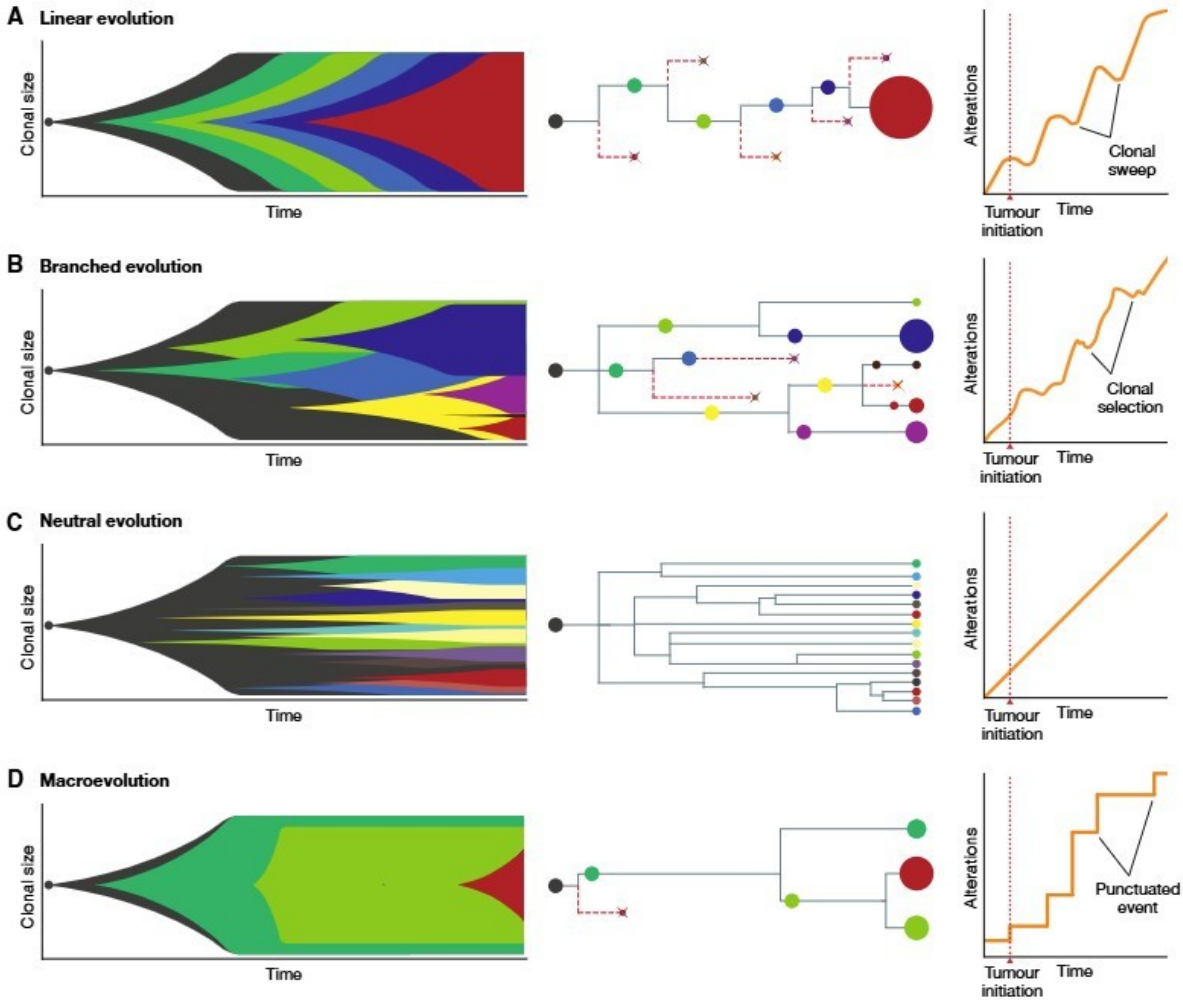


Figure 1-1. Models of tumour evolution.

Models that aim to explain tumour dynamics and cancer progression include A) linear evolution, B) branched evolution, C) macroevolution, and D) neutral evolution. These are depicted in Muller plots (left), phylogenetic trees of clonal lineages (centre), and changes in the number of genomic alterations within the population over time (right). Colours across the plots indicate different clonal lineages. Adapted from Vendramin et al. [5]

Beyond gradual models of mutation accumulation, evidence also supports **punctuated** or **macroevolutionary** dynamics. These involve catastrophic genomic events that rapidly reshape clonal architecture and may introduce multiple driver alterations within a restricted timeframe [8]. The resulting phylogenies lack intermediate populations, with dominant clones expanding after the genomic upheaval (**Figure 1-1**D). Models of sequencing data identifying complex gene rearrangements suggest substantial genomic disruption over a few events, including chromothripsis [12]; chromoplexy in prostate cancers [13]; extensive early aneuploidy in breast cancer [14]; and 'big bang' dynamics in colorectal tumour evolution [15].

Moreover, evidence increasingly supports mixed evolutionary models [8]. *Williams et al.* analysed NGS data across 14 tumour types and estimated that approximately one-third evolve under effectively neutral dynamics following early driver acquisition [16]. This suggests driver mutations are acquired early in tumorigenesis, followed by clonal expansion without further selection. Another study in breast cancer suggests that punctuated evolutionary events occur early and remain stable, whereas point mutations evolve gradually, generating extensive clonal diversity [17]. These contrasting patterns align with known mutational mechanisms: catastrophic chromosomal events arise rarely from severe regulatory failures, whereas single-nucleotide mutations accrue gradually through routine replication and repair errors [8].

Collectively, these evolutionary models describe how genetic diversity arises and becomes structured within tumours over time. The manifestation of this diversity at the tissue and clinical level is tumour heterogeneity.

1.1.3 Tumour heterogeneity: genetic and non-genetic diversity

While tumour evolution provides the conceptual framework for cancer progression, its driver, observable outcome, and measurable attribute is tumour heterogeneity: the remarkable diversity seen between and within tumours [18]. **Intertumour heterogeneity** describes variation among tumours arising in different patients with the same cancer type. Clinically, this is reflected in molecular subtyping frameworks that stratify patients based on recurrent genomic alterations or expression profiles, which are often associated with prognosis, therapeutic interventions, and treatment responses [19]. Prostate cancer exemplifies this diversity, with primary tumours exhibiting distinct molecular subtypes defined by specific genomic alterations and variable clinical behaviour. For example, tumours driven by androgen-responsive gene fusions such as *TMPRSS2–ERG*, which reprogram transcriptional networks, and tumours characterised by *SPOP* mutations that disrupt ubiquitin-mediated protein turnover, define distinct molecular subclasses [20,21].

Such interpatient variation underpins precision oncology strategies; however, patients classified within the same molecular subgroup may still experience markedly different outcomes [22]. Furthermore, most advanced cancers, including those with robust initial responses, eventually acquire resistance to therapies and relapse. These clinical

inconsistencies and treatment inadequacies are attributed to **intratumour heterogeneity** (ITH), the diversity within a tumour from the same patient. Pre-existing heterogeneity and ongoing diversification during treatment may enable some tumour cells to survive and become therapy-resistant [23]. Elevated ITH is consistently associated with worse clinical outcomes across various cancers, such as lung, gastrointestinal, breast, and prostate tumours [24], and predicts increased risk of recurrence across cancer types [25].

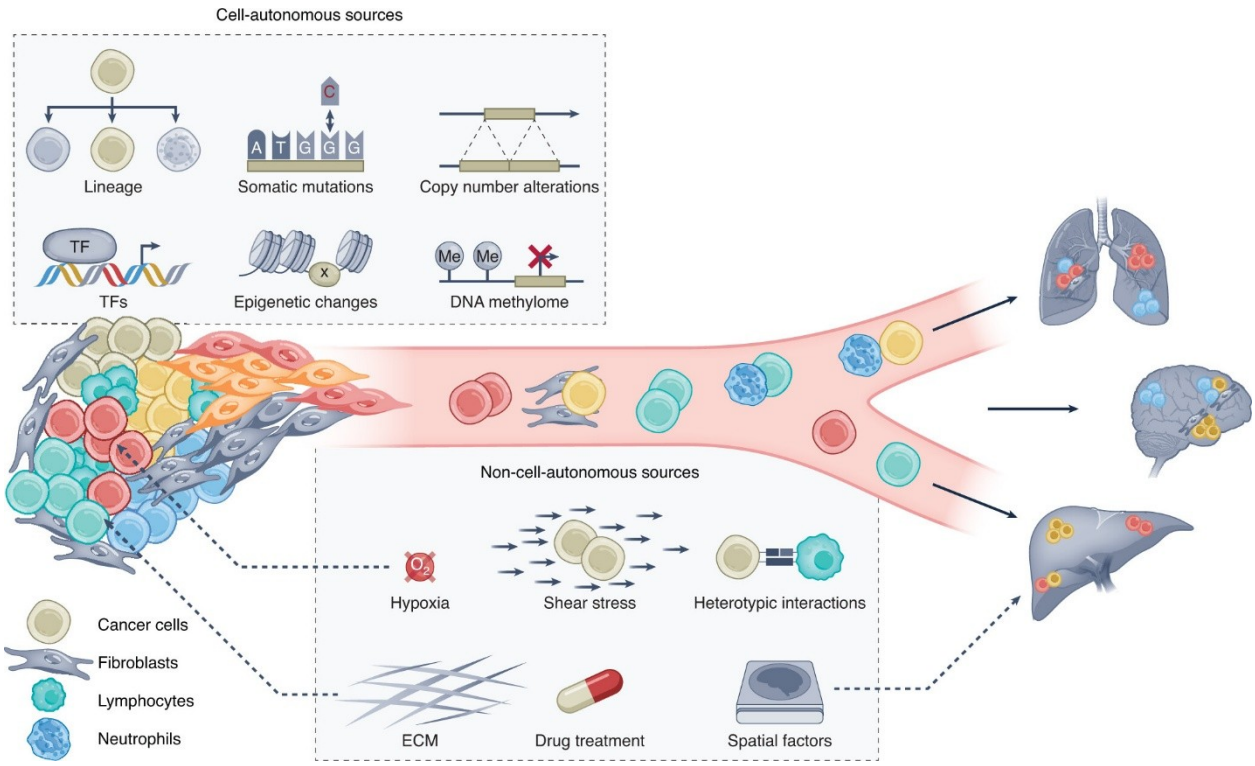


Figure 1-2. Sources of intratumour heterogeneity.

Cell-autonomous and non-cell-autonomous sources of intratumour heterogeneity (ITH) interact to promote tumour initiation and progression, including distant metastases to different organs. Reproduced, with permission, from Li et al. [25]

Extensive multiomic profiling has demonstrated that ITH arises from variation at genetic, epigenetic, and microenvironmental levels. The **phenotype** of an individual cancer cell, encompassing its functional behaviours, reflects the integration of and response to these molecular and environmental inputs [23]. Sources of ITH can be broadly categorised as cell-autonomous or non-cell-autonomous (**Figure 1-2**). Cell-autonomous factors, such as the cancer cell of origin, genetic mutations, epigenetic flexibility, and transcriptional changes, influence phenotypes in mutant cells. Non-cell-autonomous factors, such as environmental stimuli (hypoxia, physical stress, cell interactions, matrix alterations, therapies), can induce altered phenotypes in cells across the tumour

ecosystem by altering cell intrinsic mechanisms. Thus, these sources collaborate and influence each other to promote tumour initiation, progression, and metastasis [25].

At the genetic level, ITH arises from the accumulation of somatic alterations, including point mutations, copy number changes, structural rearrangements, and whole-genome duplications. Although DNA replication and repair are normally highly accurate [26], genomic instability may result from impaired nucleotide or mismatch repair, telomere dysfunction, replication stress, or chromosome segregation errors [19]. In contrast, epigenetic heterogeneity represents an additional layer of diversity that influences gene expression without altering DNA sequence. It may involve promoter hypermethylation, altered enhancer activity, and chromatin remodelling, arising through stochastic variation, genetic disruption of epigenetic regulators, or environmental and metabolic stress [27].

Collectively, genetic, epigenetic, and microenvironmental variation generate molecular diversity within tumours. However, these alterations ultimately exert their phenotypic effects through coordinated transcriptional and signalling programs that define the functional state of a cancer cell.

1.1.4 Functional cellular states and plasticity in cancer

Not all tumour variation is biologically significant; distinguishing alterations that directly influence tumour behaviour, progression, and therapeutic response is essential. Functional heterogeneity reflects the presence of distinct cellular states that shape tumour fitness. A cellular state can be defined as a coordinated configuration of transcriptional networks and signalling modules that confers specific functional properties, such as proliferation, metabolic activity, stress tolerance, or invasive capacity. Importantly, cellular states are not necessarily fixed. Cancer cells may transition between states to produce defined phenotypes in response to intrinsic regulatory changes or external cues, a phenomenon referred to as phenotypic plasticity [28].

Cell-intrinsic plasticity functions through epigenetic modifications to produce alternative lineage commitments or reversible shifts in gene expression, enabling adaptive state transitions under environmental stress or therapeutic pressure [25]. Chromatin

configuration plays a central role in this process: restrictive conditions can block activation of tumour-suppressive programmes, whereas permissive states may facilitate stochastic oncogene activation or non-physiological cell transitions [29]. Furthermore, cell-extrinsic inputs introduce additional diversity through altering transcription factor activity, regulating alternative splicing, differential promoter usage, and RNA editing, thereby expanding functional heterogeneity without novel genetic mutations [27,28].

Prostate cancer exemplifies functional plasticity. Most primary tumours are androgen receptor (AR)-dependent adenocarcinomas; however, under therapeutic pressure, a subset of cells can transition to an AR-independent neuroendocrine phenotype [30]. Treatment-related neuroendocrine prostate cancer (NEPC) is recognised as a mechanism of resistance to AR-targeted therapies and typically emerges in advanced disease. NEPC is associated with poor prognosis, partly due to delayed detection and limited effective treatment options [31]. The initial response to antiandrogen therapy suggests that the majority of tumour cells remain AR-dependent, indicating that only a minority possess the capacity for neuroendocrine transformation [28]. This lineage switch is frequently associated with RB1 and TP53 loss and is supported by epigenetic reprogramming, including alterations in DNA methylation and chromatin state [24].

The existence of plastic cellular states complicates the interpretation of intratumour heterogeneity. Observed phenotypic differences may reflect stable clonal divergence or reversible adaptive transitions. Consequently, tumour evolution cannot be inferred solely from genetic architecture. In this context, the tumour microenvironment plays a critical role in shaping which cellular states are maintained, suppressed, or selectively expanded over time [32].

1.1.5 The tumour microenvironment as a selective landscape

In several tumour types, clonal origins can be traced back decades before clinical diagnosis, yet malignant progression occurs much later. For example, a recent study identified a shared ancestral clone between breast cancer and clonally related benign tissue harbouring the chromosomal translocation der(1;16), likely acquired early in life [33]. The malignant and non-malignant lineages subsequently diverged without consistent acquisition of additional canonical driver mutations. These observations demonstrate that genetically related clones can follow distinct evolutionary trajectories, and that

stable driver alterations alone do not determine malignant fate. Rather, the same genotype may yield divergent outcomes depending on the contextual environment in which it persists. Such that, framing tumours as ecological systems emphasises that cancer cell fitness is shaped by interactions with surrounding cellular and structural components [34].

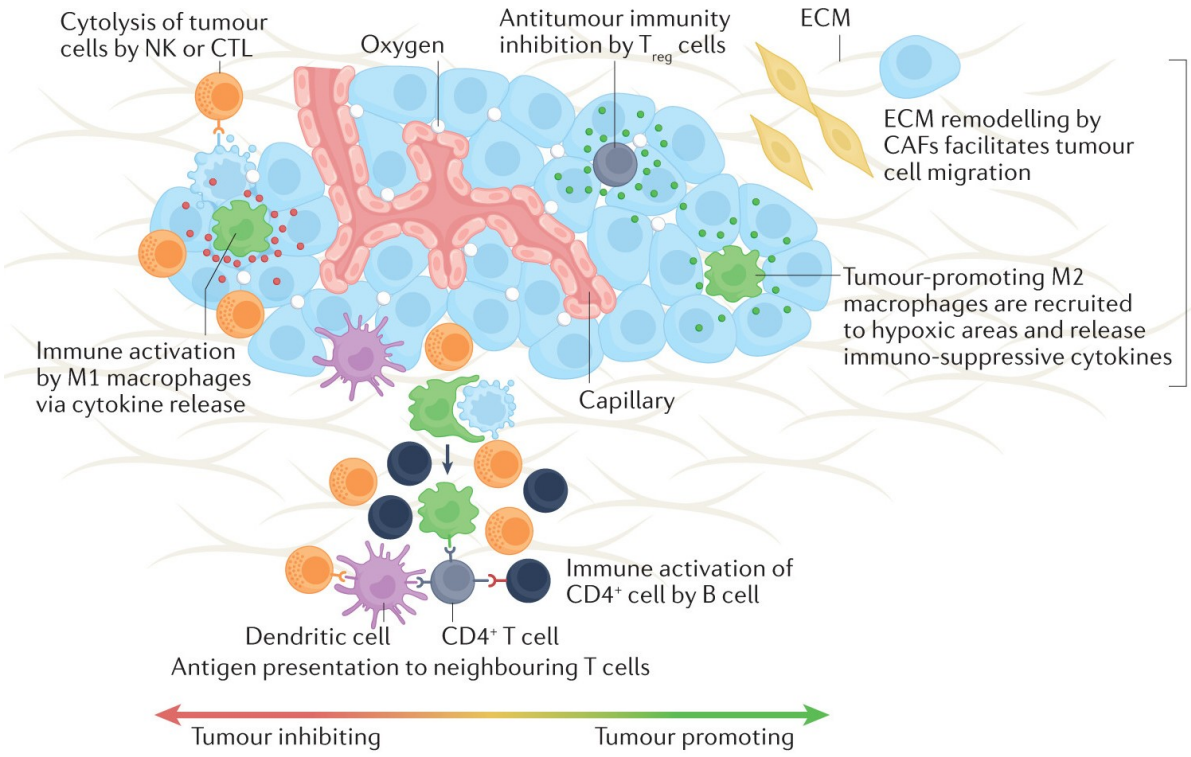


Figure 1-3. Components of the tumour microenvironment.

Schematic of the cellular composition of the tumour microenvironment (TME). The TME comprises various cell types that mediate tumour-inhibiting (left) and tumour-promoting (right) interactions with cancer cells. NK = natural killer cell; CTL = cytotoxic lymphocyte; ECM = extracellular matrix; CAFs = cancer-associated fibroblast. Reproduced, with permission, from Seferbekova et al. [37]

Cancers are complex ecosystems composed of malignant cells and diverse non-malignant populations, embedded within a remodelled extracellular matrix (ECM), collectively termed the **tumour microenvironment** (TME) (**Figure 1-3**). The composition and functional state of the TME vary according to tumour type, anatomical site, developmental stage, intrinsic cancer cell properties, and patient-specific factors [35,36]. The TME comprises immune, stromal, and vascular populations, ECM components, and tissue-resident cell types, such as adipocytes and neurons. These components actively influence tumour initiation, progression, and metastasis, exerting both tumour-suppressive and tumour-promoting functions. For example, antigen presentation by CD4+ T cells and dendritic cells supports CD8+ T cell–mediated

cytolysis of cancer cells, yet these immune populations may acquire dysfunctional or exhausted phenotypes within tumours [36].

1.1.5.1 Microenvironmental modulation by cancer cells

Cancer cells create a supportive TME by recruiting and reprogramming host cells, remodelling vasculature and the ECM. Contact-dependent communication occurs via adhesion molecules and cell-surface receptors, whereas paracrine signals, such as cytokines, chemokines, growth factors, extracellular vesicles, and proteases, facilitate long-range crosstalk. These signals, triggered by intrinsic changes or cellular stress, act through receptor engagement or by reshaping the ECM. For example, cancer-associated fibroblasts are prominent contributors, synthesising matrix components and promoting angiogenesis [36]. Metabolic reprogramming and metabolite communication are also integral to tumorigenesis. Many cancers undergo a shift from oxidative phosphorylation to aerobic glycolysis (the Warburg effect), increasing lactate production and secretion into the TME [38]. Subsequently, *in vivo* lung cancer studies show that lactate, previously assumed to be a waste product, can serve as the preferred tricarboxylic acid cycle carbon source over glucose in other cancer cells [39].

Interactions among cancer cells within the tumour microenvironment are extensive; although many subclonal interactions may be neutral, those that influence phenotype are typically cooperative or competitive. Negative interactions occur in competitions for limited nutrients, space, or oxygen. Clones may secrete factors to suppress or destroy neighbouring cells, eliminating weaker subclones and enabling fitter ones to expand. Positive interactions include commensalism, where one clone supports others by sharing resources, circulating signalling molecules, or diffusing growth factors and cytokines. Synergistic interactions produce characteristics that are not observed in individual populations. Mutualism occurs when clones cooperate to produce factors or gather resources, benefiting all cancer cells and increasing the tumour's overall fitness [2] . *Kornelia Polyak* and colleagues have reported subclonal cooperation in breast cancer *in vivo* models, where subpopulations of cells stimulate overall tumour growth and metastasis by inducing tumour-promoting microenvironmental states [40,41] .

Overall, extensive characterisation of the TME has demonstrated its clinical relevance, influencing clonal fitness, tumour maintenance, and therapeutic response [36]. Cancer cells and tumour-associated myeloid populations frequently overexpress PD-L1,

suppressing immunosurveillance by engaging PD-1 on adaptive immune cells. Immune checkpoint blockade targeting this axis has therefore become an established therapeutic strategy across multiple tumour types [42]. Furthermore, growing evidence indicates that spatial organisation within the TME is prognostically significant. *Schürch et al.* identified distinct immune–tumour architectures in advanced colorectal cancer, demonstrating that high-risk patients exhibited disrupted spatial organisation between tumour and immune cellular neighbourhoods [43]. Critically, this breakdown in local immune–tumour communication was associated with poorer clinical outcomes. These observations support the view that the spatial and ecological structure of the TME is a determinant of clinical behaviour, influencing both disease progression and treatment response.

Therefore, the TME acts not as a passive scaffold but as an active selective force that constrains, promotes, and stabilises specific tumour cell phenotypes. Tumour evolution therefore unfolds within an ecological landscape in which clonal fitness is continuously shaped by spatial organisation, intercellular signalling, cellular interaction, and metabolic exchange.

1.1.6 Studying cancer evolution: resolving clonal dynamics

Cancer progression is a dynamic process shaped by ongoing clonal expansion, selection, phenotypic adaptation, and microenvironmental pressures, rather than a simple linear accumulation of mutations. Genetic diversification, plastic state transitions, and ecological interactions unfold over time, collectively determining tumour behaviour and therapeutic response [23] . Therefore, understanding these processes requires not only characterising tumour composition but also elucidating how cellular populations change and interact across successive disease stages [44] .

However, in clinical settings, ethical and practical constraints limit longitudinal sampling, and most patient-derived analyses rely on single or infrequent biopsies [24] . As a result, tumour diversity is typically interpreted retrospectively from static molecular profiles, treating observed heterogeneity as a historical record of past evolutionary events [24] . While such approaches have provided critical insights into clonal architecture, including the identification of dominant and minor subclones, the functional relevance of ITH remains difficult to discern from these snapshots [25] .

Preclinical model systems partially address these constraints by enabling controlled perturbation, repeated sampling, and experimental manipulation of genetic and environmental variables. Advances in experimental systems that more faithfully reproduce the ITH observed in human tumours, such as genetically engineered mouse models and patient-derived systems, enable the investigation of tumour progression under defined conditions [24,25]. Nevertheless, even in experimental settings, clonal relationships are frequently inferred from endpoint similarity.

Without explicit lineage information, it remains difficult to distinguish clonal expansion from convergent phenotypic states, to identify transient adaptive populations, or to determine whether observed differences reflect stable ancestry or dynamic state transitions driven by plasticity and microenvironmental selection. Comprehensive deconvolution of clonal dynamics and phenotypic adaptation requires approaches that can directly encode and track cellular ancestry, thereby underpinning the use of lineage-resolved strategies in cancer biology.

This thesis focuses on the development and application of tools for direct lineage tracing. The following section briefly outlines the historical foundations and biological relevance of lineage tracing approaches before examining current methodologies and their applications, with particular emphasis on cancer research.

1.2 Lineage tracing and cellular barcoding

1.2.1 Lineage tracing in biology: retrospective and prospective approaches

Understanding how cells relate to one another over time is fundamental to biology. In multicellular organisms, tissues arise through successive rounds of cell division, differentiation, and migration, giving rise to complex architectures and functional specialisation [45]. In development, lineage relationships define how progenitor cells generate diverse cell types with specific fates and functions. Similarly, in adult tissues, lineage dynamics underpin homeostasis, regeneration, and repair, thus enabling efficient biological function [46,47]. In disease, altered lineage trajectories can drive pathological outcomes, such as the degeneration of biological systems or malignant transformation, providing the well-established basis for cancer evolution. Across these contexts, lineage tracing provides a framework for linking cellular ancestry to fate,

function, and behaviour, thereby elucidating the underlying mechanisms that control these processes [48].

Lineage relationships, however, are not directly observable. Cells do not carry explicit records of their ancestry, and lineage must therefore be inferred from measurable features. Different lineage-tracing strategies encode distinct assumptions about how ancestry can be reconstructed and what information can be recovered. Broadly, these strategies fall into two categories: retrospective lineage inference, which reconstructs probable lineage relationships from endpoint measurements, and prospective lineage tracing, which labels cells in advance and follows their descendants over time (**Figure 1-4**).

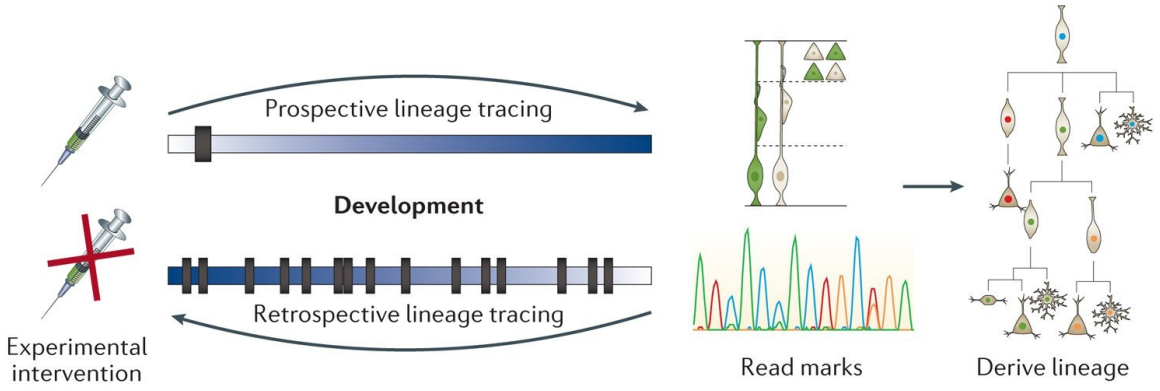


Figure 1-4. Prospective and retrospective methods for lineage tracing.

Prospective lineage tracing involves experimentally applying lineage markers to a cell or population (grey rectangle on the blue timeline), then following cells forward to read their output at a later time. In contrast, retrospective lineage tracing follows cells backwards to read endogenous marks (multiple grey rectangles on the blue timeline) that have accumulated over a specific time period or throughout an organism's life. Adapted, with permission, from Woodworth et al. [49]

Retrospective lineage tracing exploits naturally occurring cellular variation to infer ancestry. Shared features such as somatic mutations, copy-number alterations, mitochondrial variants, or stable epigenetic marks are treated as intrinsic records of lineage history. Advances in bulk and single-cell sequencing, coupled with increasingly sophisticated computational and phylogenetic models, have enabled the reconstruction of lineage relationships across diverse biological systems [48] [50]. Retrospective approaches are powerful for investigating human tissues, where experimental manipulation is not feasible, thereby allowing lineage relationships to be inferred from patient samples. For example, as discussed in Section 1.1, tumour growth is often interpreted as an evolutionary process driven by the accumulation and selection of

somatic mutations, enabling reconstruction of clonal hierarchies and evolutionary trajectories from tumour sequencing data [8].

Despite their impact, retrospective approaches are fundamentally limited by their reliance on inference from a small number of data points or a single endpoint measurement. Lineage relationships are reconstructed computationally from static molecular snapshots, meaning that the order and directionality of branching events must be statistically inferred rather than experimentally observed. Short-lived or low-frequency cellular states may not leave an evident molecular signature or be undetectable in endpoint analyses. Furthermore, because these approaches rely on naturally occurring somatic variation or transcriptional heterogeneity, the amount of recoverable lineage information may be limited and cannot be experimentally controlled. This restricts resolution and complicates functional validation of inferred mechanisms. These limitations are particularly restrictive in systems undergoing active change, such as development, regeneration, or cancer under therapeutic pressure, where plastic state transitions and microenvironmental interactions play critical roles [51].

Prospective lineage tracing addresses key limitations of retrospective inference by introducing heritable labels into cells before or during the biological process of interest. By marking cells in advance and identifying their descendants at later time points, prospective approaches establish lineage relationships directly rather than reconstructing them indirectly from endpoint similarity. This framework enables clonal expansion and lineage structure to be resolved across defined experimental intervals and supports longitudinal analyses in which lineage relationships can be interpreted in relation to perturbations, environmental changes, or developmental context [48]. Consequently, prospective lineage tracing is applied in experimental systems where controlled manipulation and repeated sampling are possible.

Together, retrospective and prospective approaches represent complementary strategies for resolving lineage relationships. Retrospective methods provide broad access to lineage structure in native human tissues, whereas prospective methods enable controlled interrogation of lineage relationships under defined conditions. In the context of cancer, where cellular phenotypes are highly dynamic, experimentally defined lineage markers provide a stable reference against which plastic cellular states can be interpreted. Anchoring ancestry in this way enables direct measurement of clonal

expansion and selection over time, creating the controlled systems necessary to interrogate the mechanistic basis of tumour plasticity and evolutionary dynamics.

This thesis focuses on prospective lineage tracing strategies, collectively referred to as **cellular barcoding**. The following sections review the conceptual principles, technological implementations, and biological applications of these tools.

1.2.2 Principles of cellular barcoding

1.2.2.1 Traditional prospective lineage tracing

Prospective lineage tracing predates modern sequencing technologies that enable the retrospective analysis of samples. Lineage-tracing methods are rooted in developmental biology research dating back to the early 1900s. Investigations sought to understand how individual cells contribute to tissue formation and patterning, initially relying on direct observation of embryonic cell divisions in simple or transparent organisms [45]. To examine increasingly complex biological systems, researchers developed strategies to heritably mark individual cells so that their descendants could be subsequently identified [49].

Historically, prospective lineage tracing has relied on physical, chemical, or genetic markers inherited by a progeny cell from its parent cell. These approaches enabled investigation of lineage relationships, cellular movements, and differentiation trajectories, particularly when combined with experimental perturbations that altered developmental or environmental conditions [47]. Early methods used dyes and radioactive labels, typically introduced by microinjection or local application, which are passively partitioned during cell division. However, this technique is limited by the loss of the marker due to diffusion or dilution during successive cell divisions, by the number of cells that can be labelled and distinguished, and by its inability to be applied to complex organisms [45,49].

By the late 20th century, advances in recombinant DNA technology had greatly expanded the range of prospective lineage-tracing tools. Genomic integration of genetically encoded reporters, exemplified by green fluorescent protein (GFP), enables heritable labelling of cells, thereby stably propagating lineage information across cell divisions. Such reporters could be introduced via viral transduction or activated *in vivo*

through site-specific recombination strategies. Recombinase-based systems enabled temporally controlled lineage labelling by restricting reporter activation to defined cell types or developmental stages. In mouse models, the Cre–loxP system is the most widely used example. Cre recombinase is expressed under a cell- or tissue-specific promoter and excises a loxP-flanked transcriptional STOP cassette, irreversibly activating a downstream reporter gene such as GFP [46] .

Subsequent refinements introduced multicolour reporter systems, increasing the number of simultaneously distinguishable lineages that could be followed. This is exemplified by the Brainbow mouse, which uses Cre-mediated recombination to drive combinatorial expression of multiple fluorescent proteins in neurons [52]. Similarly, a Confetti reporter mouse was generated and used to investigate intestinal crypt stem cells using different fluorophores [53]. Genetic recombination enables permanent marking of selected cell populations and their descendants, supporting long-term lineage tracing in complex organisms. Furthermore, strategies using multiple fluorophores represent a significant technical advance; however, the number of fluorescent proteins and hues that can be distinguished constrains their complexity [45,49] .

Collectively, traditional prospective lineage-tracing methods are well suited to small populations or to defined developmental contexts. Yet, these methods are limited in their ability to trace lineages across generations and lack the diversity required to resolve lineage relationships in large, heterogeneous systems [46] .

1.2.2.2 Modern cellular barcodes and design principles

The advent of advanced genetic and sequencing technologies has enabled advances in prospective lineage tracing by replacing visual markers with nucleic acid sequences as cellular barcodes. In modern barcoding systems, cells are labelled with heritable DNA or RNA sequences that function as unique identifiers, with cells sharing a barcode inferred to descend from a common ancestor [46]. This strategy overcomes the limitations of diversity in traditional reporters and enables lineage tracing of large, complex cell populations, including those relevant to intricate biological systems such as development, homeostasis, and diseases such as cancer [48] .

The defining advantage of nucleic acid barcodes is their capacity for combinatorial diversity. Short nucleotide sequences can encode vast numbers of unique identifiers,

with the theoretical diversity increasing exponentially with sequence length [54]. Conceptually, the number of possible unique barcodes is 4^N, where N is the sequence length, and each position can encode one of four nucleotides. A random 10-base pair (bp) barcode can assume ~10^6 possibilities, while a 30-bp barcode generates ~10^{18} combinations, far exceeding typical experimental cell numbers. This property enables the unique labelling of many cells within a large population in a single experiment.

Collectively, modern cellular barcoding strategies are generally categorised as either static or evolvable based on whether the barcode identity (sequence) changes over time (**Figure 1-5**). **Static barcoding** assigns a collection (library) of labels that remain unchanged over time, enabling the stable identification of clonal descendants. **Evolvable barcoding** introduces genetic mechanisms that induce changes to barcode sequences over successive cell divisions, thereby enabling interpretation of lineage history. Significantly, this distinction reflects the conceptual information recoverable from the different systems when applied experimentally. Static systems encode stable lineage *identity*, whereas evolutionary systems enable recovery of generational ancestry information [51].

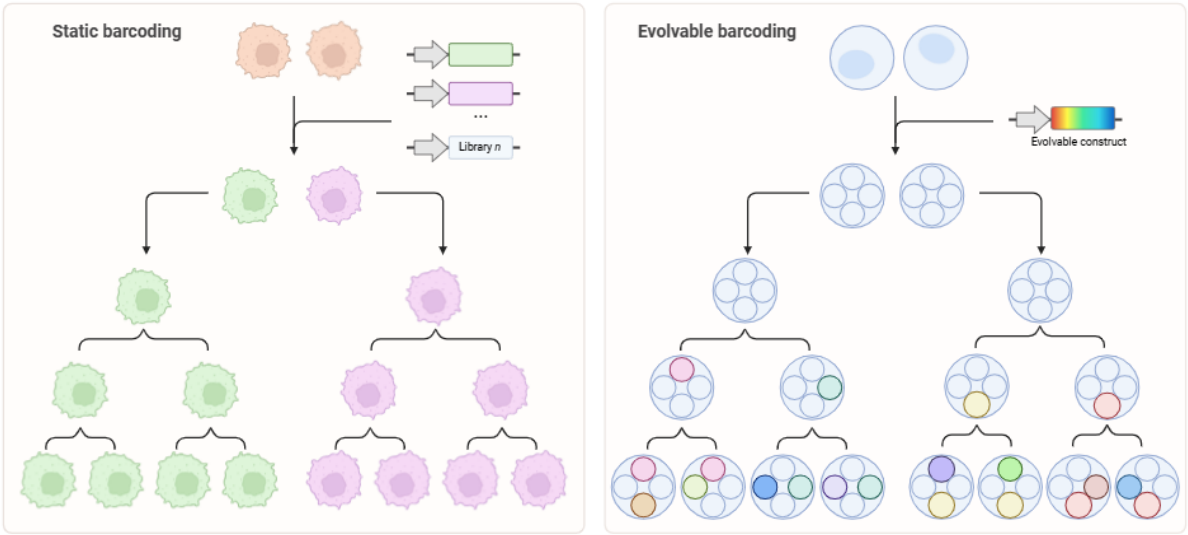


Figure 1-5. Static and evolvable barcoding strategies.

Static barcoding provides stable lineage tracing with fixed identifiers; evolvable barcoding enables interpretation of generational ancestry through identifiers that change during successive cell divisions. Created with BioRender.com.

Across all prospective barcoding strategies, several core functional requirements must be satisfied. Barcodes must be heritable and stably transmitted to all progeny, sufficiently diverse to uniquely label the population of interest, and detectable in downstream assays without perturbing cellular behaviour [54]. **Table 1-1** presents a non-

exhaustive summary of current barcoding systems that employ static libraries and evolving barcode strategies. Overall, these techniques differ in how barcode diversity is generated, how information is encoded over time, and how lineage relationships can be resolved. Use of these tools in biological models requires trade-offs between system stability and flexibility, diversity and interpretability, and informational depth and technical complexity [55] . The following sections examine static barcoding in detail before discussing evolvable systems.

Table 1-1. Static and evolving genetic barcoding strategies.

FISH = fluorescence in situ hybridisation; scRNA-seq = single-cell RNA sequencing; nCas9 = nickase Cas9; BE = base editor; TDT = terminal deoxynucleotidyl transferase; scATAC-seq = single cell assay for transposase-accessible chromatin using sequencing; PE = prime editor; RT = reverse transcriptase; ST = spatial transcriptomics. N/A = non-applicable.

Year	System	Sequence	Instantiation	Functionalisation	REF.
2015	ClonTracer	Static	Lentivirus	N/A	56
2016	GESTALT	Evolving	Cas9 nuclease	N/A	57
2017	MEMOIR	Evolving	Cas9 nuclease	FISH	58
	Polylox	Static (recombination)	Cre-Lox	N/A	59
2018	scGESTALT	Evolving	Cas9 nuclease	scRNA-seq	60
	TracerSeq	Static	Tol2 transposase	scRNA-seq	61
	CellTag	Static	Lentivirus	scRNA-seq	62
	COLBERT	Static, retrieval	Lentivirus	Live cell isolation	63
	LINNAEUS	Evolving	Cas9 nuclease	scRNA-seq	64
	ScarTrace	Evolving	Cas9 nuclease	scRNA-seq	65
2019	Hwang et al.	Evolving	nCas9-BE	N/A	66
2020	CARLIN	Evolving	Cas9 nuclease	scRNA-seq	67
	CaTCH	Static, retrieval	Lentivirus	Live cell isolation	68
	CloneSifter	Static, retrieval	Lentivirus	Live cell isolation	69
	LARRY	Static	Lentivirus	scRNA-seq	70
	PolyloxExpress	Static (recombination)	Cre-Lox	scRNA-seq	71
	Zombie	Evolving	nCas9-BE	FISH	72
2021	IntMEMOIR	Evolving	Integrase	FISH	73
	Rewind	Static	Lentivirus	FISH	74
	CHYRON	Evolving	Cas9-TdT	N/A	75
	ClonMapper	Static, retrieval	Lentivirus	scRNA-seq, live cell isolation	76
	macsGESTALT	Evolving	Cas9 nuclease	scRNA-seq	77
	Watermelon	Static	Lentivirus	scRNA-seq	78
2022	SPLINTR	Static	Lentivirus	scRNA-seq	79
2023	CellTag-multi	Static	Lentivirus	scRNA-seq, scATAC-seq	80
	DARLIN	Evolving	Cas9-TdT	scRNA-seq	81,82
2025	OLIGO-CALL	Static, retrieval	Lentivirus	scRNA-seq, live cell isolation	83
	LoxCode	Static (recombination)	Cre-Lox	scRNA-seq	84
	PEtracer	Evolving	nCas9-PE-RT	scRNA-seq, ST	85

1.2.3 Static barcodes

Static barcode libraries are typically generated *in vitro* by synthesising semi-random, fully random, or defined nucleotide sequences, which are then cloned into viral vectors or plasmids [54]. These synthetic libraries can contain thousands to millions of unique sequences, with library complexity defining the maximum number of uniquely labelable cells. Delivery of static barcodes is typically achieved through lentiviral transduction. In practice, barcoding experiments use a low multiplicity of infection (MOI), typically <0.3, to minimise multiple barcode integrations per cell. Then, constructs include fluorescent proteins or antibiotic resistance markers to enable selection of labelled cells by flow cytometry or drug selection [51]. Therefore, an important experimental consideration is maintaining barcode singularity while ensuring efficient integration.

Static barcode diversity can also be generated *in vivo* through recombinase-based strategies. As discussed previously, Cre–loxP systems enable conditional activation of genetic reporters within defined cell populations [45,52] . Extensions of this logic allow recombinase activity to rearrange predefined DNA segments, generating unique heritable sequence configurations within individual cells [59] . In such systems, barcode diversity arises from controlled recombination events that shuffle or excise predefined DNA elements to produce distinct sequence combinations. Temporal control of recombinase activity restricts this diversification to defined developmental or disease windows, after which the barcodes remain unchanged and support long-term lineage tracing, functioning as a static clonal identifier.

Therefore, *in vivo* recombination-based strategies enable labelling of endogenous cells within intact systems, preserving native microenvironmental interactions. For the successful implementation of these strategies, pre-engineered recombination target arrays must be designed and integrated into a system with controlled recombinase expression. Thus, achievable barcode diversity is inherently limited by the number of recombination sites encoded within the system [54] . Alternatively, *in vitro* barcode library generation offers high diversity, compact sequence design, and compatibility with short-read sequencing platforms. Delivery is generally efficient and scalable, making these systems well-suited for clonal tracking in cultured cells or for transplantation into *in vivo* models [86] .

The stages in performing a typical static barcode lineage tracing experiment involve generating or acquiring a barcode library, introducing barcodes into a target cell population at low MOI, and expanding the labelled population. Barcoded cells can then be used for *in vitro* assays or implanted into model organisms for *in vivo* investigations [62,70,76,87]. Following experimental perturbation, barcode sequences are recovered, and quantification of barcode abundance is performed according to the barcode type [86]. Although static barcodes are characterised by their stability, they are typically classified according to the biological level at which they are identified. **DNA barcodes** are obtained from genomic DNA; **RNA barcodes** are transcribed and detected at the transcript level; **protein barcodes** are represented as distinguishable epitopes or reporter combinations.

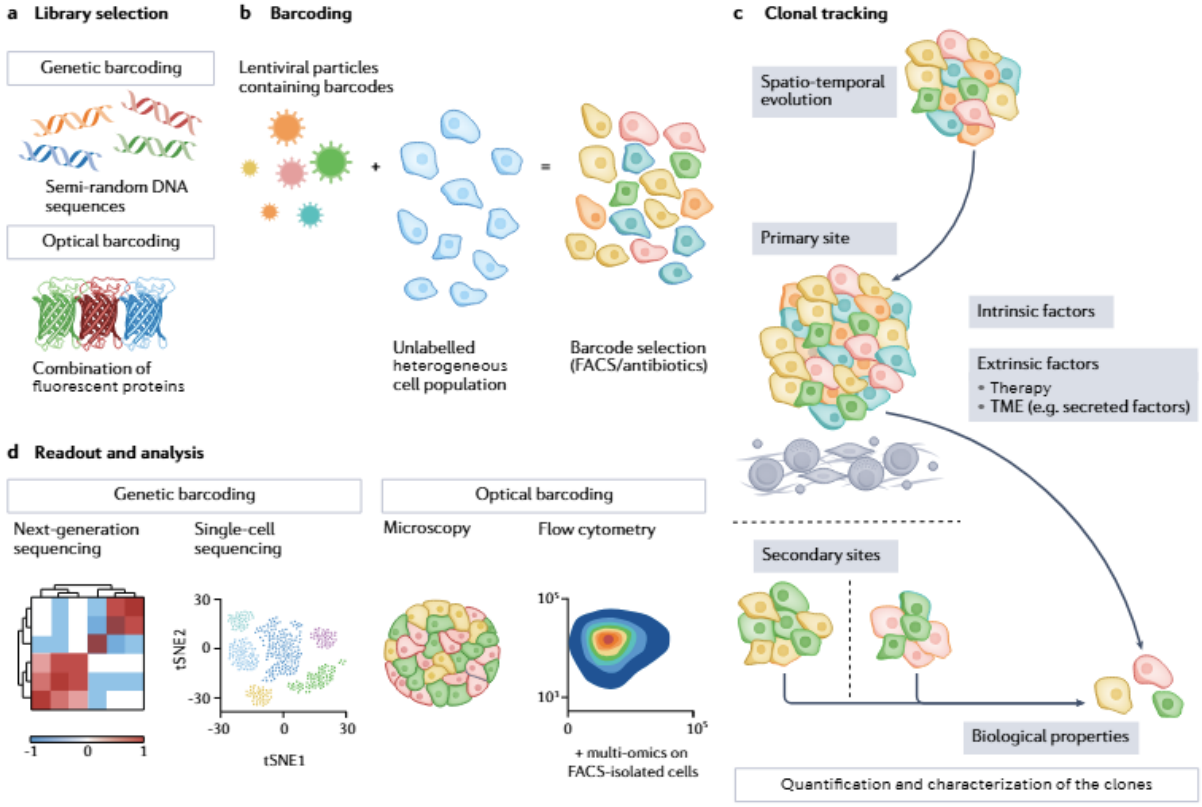


Figure 1-6. Representative workflow using a static or optical barcode library in cancer studies.

A barcode library is selected (a) for labelling of cancer cells via lentiviral transduction. Subsequent selection of barcoded cells is performed by fluorescence-activated cell sorting (FACS) or with antibiotics (b). The fate of infected cells and their progeny can be traced in space and time in clonal-tracking experiments (c). Clonal fate is determined by intrinsic or extrinsic factors, including therapy and secreted factors from the tumour microenvironment (TME). Readout and analytical methods for identifying and quantifying barcodes depend on the cellular barcoding strategy employed (d). Reproduced, with permission, from Serrano et al. [86]

Importantly, this distinction is not merely technical, as the readout level determines which measurement platforms are technically compatible and defines the additional

phenotypic information that can be acquired alongside lineage identity. DNA and RNA barcodes are measured through sequencing technologies. Protein-level barcodes can be detected by suspension cytometry or visualised *in situ* by microscopy and imaging methods, via the inherent fluorescence of reporter proteins or with targeted labelled antibodies. A representative workflow for a static barcoding experiment to investigate cancer clonal dynamics *in vivo* is shown in **Figure 1-6**.

Collectively, static barcodes differ in sequence design and in how and where barcode diversity is generated. These design choices create trade-offs among library complexity, experimental flexibility, and readout compatibility. The following subsections examine exemplary static barcode systems by readout level and their applications in complex biological systems, with particular emphasis on those that have illuminated clonal dynamics in cancer.

1.2.3.1 DNA barcodes

The ClonTracer, reported by *Bhang et al*.[56] , was among the first high-complexity static barcoding systems used in cancer research, employing more than 10^7 unique DNA barcodes. Barcodes were virally integrated and subsequently recovered from genomic DNA for quantitative analysis via sequencing. In epidermal growth factor receptor (EGFR)-mutant non-small-cell lung cancer (NSCLC) models, ClonTracer enabled investigation of clonal resistance to EGFR inhibitors. Researchers sought to determine whether resistance mutations arose during treatment or originated from pre-existing resistant subclones. Experimentally, this involved splitting barcoded cells into treatment replicates, enabling comparison of clonal dynamics across replicates to distinguish reproducible selection from stochastic outgrowth.

Bhang et al. reported that erlotinib treatment consistently drove expansion of an extremely rare pre-existing subpopulation across independent replicates, indicating that resistance was dominated by selection of rare clones already present at baseline [56] . In parallel, *Hata et al.* observed a similar selection of pre-existing clones harbouring defined resistance mechanisms. However, they also described an additional trajectory where drug-tolerant clones lacking a canonical resistance mutation could persist under treatment and subsequently acquire resistance [88] . Together, these studies demonstrated that resistance in EGFR-mutant NSCLC can result from rare pre-existing

clones and evolutionary adaptation of drug-tolerant populations, revealing both deterministic and dynamic resistance processes.

Nevertheless, DNA barcoding alone may not explain why particular clones dominate. It provides robust quantitative lineage structure, but linking clonal expansion to cellular phenotype requires coupling barcode identity to molecular state measurements. This motivation underpins the development of expressed RNA barcodes, which enable the capture of clonal identity alongside transcriptional programmes at single-cell resolution.

1.2.3.2 RNA (expressed) barcodes

Subsequently, methods were developed to express static DNA barcodes as RNA transcripts, enabling their capture alongside transcriptomic profiles via single-cell RNA sequencing (scRNA-seq). This advance enables direct linking of clonal identity to cellular state, aiding in distinguishing whether phenotypic heterogeneity arises from lineage-related differences or from state changes within the same lineage. This is particularly important in cancer, where transcriptional cell states can be plastic and environmentally responsive [89]. Thus, without lineage information, similar transcriptomes can mask distinct ancestries, and, conversely, a single clone can occupy multiple states over time or across conditions.

Watermelon is an expressed static barcoding framework developed to relate clonal identity to drug-induced dynamics in cancer. *Oren et al.* investigated non-genetic mechanisms underlying therapy failure, focusing on the emergence of reversible drug-tolerant persister states [78] . Lineage tracing was achieved by integrating a clone-specific DNA barcode into the 3′ untranslated region (UTR) of an mNeonGreen reporter transcript, enabling barcode capture during scRNA-seq. Additionally, the proliferative history was monitored using a doxycycline-inducible H2B-mCherry construct via the Tet-On system. In this design, doxycycline transiently induces expression of a histone H2B–mCherry fusion protein that incorporates into chromatin. Following removal of doxycycline, mCherry fluorescence is progressively diluted through successive cell divisions, enabling inference of recent proliferation history from fluorescence intensity.

The Watermelon library, comprising over five million unique barcodes, was introduced into clinically relevant *in vitro* and *in vivo* models of lung, breast, and colorectal cancer. This dual system enabled lineage tracing to be interpreted alongside growth dynamics,

allowing assessment of whether drug tolerance is associated with distinct proliferative behaviours. The authors demonstrated that cycling and non-cycling persister cells originate from distinct lineages characterised by divergent transcriptional and metabolic programs. In particular, antioxidant responses and shifts toward fatty acid oxidation were associated with persister proliferation across multiple cancer types. Subsequent targeting of oxidative stress or metabolic reprogramming altered the percentage of cycling persisters, revealing lineage-specific vulnerabilities relevant to minimal residual disease [78] .

Another study used Single-Cell Profiling and Lineage Tracing (SPLINTR) to explore non-genetic factors influencing clonal fitness and dominance in acute myeloid leukaemia (AML) [79] . *Fennell et al.* created diverse libraries to track multiple cell populations and examine how transcriptional programs affect cell fate in AML mouse models. These libraries used distinct barcode structures embedded in the 3′ UTRs of fluorescent reporters, allowing simultaneous lineage tracing and cell profiling via scRNA-seq. The authors demonstrated that malignant clonal dominance is primarily a heritable, cell-intrinsic trait characterised by the repression of antigen-presenting pathways and by high expression of the secretory leukocyte peptidase inhibitor (Slpi) gene. Additionally, increased transcriptional heterogeneity was associated with improved clonal fitness across tissues, immune environments, and competitive contexts involving genetically distinct clones.

RNA barcodes are exceptionally valuable for linking lineage to state, but they do not inherently allow direct experimental testing of particular clones of interest. This limitation motivates the incorporation of clonal retrieval tools, enabling the retrieval of live clones and direct downstream experimental re-interrogation of specific lineages of interest.

1.2.3.3 Functionalisation via clonal retrieval

The ClonMapper system is an exemplary expressed barcoding system in which RNA barcodes can be repurposed for live-cell retrieval, enabling downstream molecular and phenotypic characterisation of interesting clonal lineages [76] . In the context of cancer evolution, this is particularly valuable for downstream functionalisation studies to determine the properties of resistant or metastatic clones and to validate candidate mechanisms identified by transcriptomic profiling.

Guitierrez et al. generated a highly diverse (10^7) barcode library in which barcodes are expressed as guide RNAs (gRNAs) using a modified gRNA expression vector [76]. Consequently, the ClonMapper system can implement the COLBERT retrieval strategy, which isolates clones from a barcoded population using the stably integrated barcode-gRNA [63]. In practice, clonal isolation is achieved by introducing a transcriptional activator variant of Cas9 (dCas9-VPR) and a 'Recall' plasmid containing a matching target barcode sequence positioned upstream of a reporter cassette containing green fluorescent protein (GFP). Therefore, only cells carrying the corresponding barcode-gRNA can recruit dCas9-VPR to the Recall plasmid, activating GFP reporter expression. This enables the isolation of that specific clone by fluorescence-activated cell sorting [76] .

The authors applied ClonMapper to investigate clonal dynamics in response to chemotherapy in an *in vitro* model of chronic lymphocytic leukaemia (CLL). They identified reproducible high- and low-survivorship subpopulations with distinct transcriptional programs and differential sensitivity to chemotherapy within the CLL model. Crucially, clonal retrieval enabled the physical isolation and re-expansion of these individual clones, allowing direct functional validation of the survival differences. This confirmed that differential fitness was a durable, clone-intrinsic property, thereby demonstrating that stable non-genetic states can shape therapeutic outcomes in CLL [76] .

Expressed RNA barcoding has been transformative in linking clonal identity to transcriptional heterogeneity, thereby elucidating the non-genetic factors driving cancer evolution. However, reliance on sequencing-based readouts can limit system accessibility and sample scalability due to costs. Furthermore, destructive sample preparation precludes analysis of clonal dynamics within intact tissues, thereby preventing assessment of direct microenvironmental influences [86] .

1.2.3.4 Protein barcodes

Protein-level barcoding provides an alternative to nucleic acid-based systems, offering distinct advantages: barcode detection can be achieved without sequencing via antibody- or imaging-based measurements. Therefore, barcode readout can be integrated directly into high-dimensional protein phenotyping workflows. Additionally, barcode identity can be visualised *in situ*, preserving spatial context and allowing mapping of clonal identity within the native tumour architecture. However, protein

barcodes are limited in their complexity, offering significantly lower diversity than nucleic-acid-based barcode libraries, as the number of resolvable epitopes or optical channels is finite [86] .

In epitope barcoding, integrated constructs encode short linear epitopes that are translated and displayed on the cell surface. *Brian Brown* and colleagues developed Pro-Codes, synthetic modules encoding triplet combinations of linear epitopes, generating 364 unique protein barcodes [90]. These barcodes can be detected using isotopically tagged antibodies for high-dimensional readout via suspension time-of-flight cytometry (CyTOF) or imaging mass cytometry, enabling parallel barcode detection with targeted antibody phenotyping. While the original Pro-Code studies used the library for CRISPR screenings [90,91], *Garry Nolan* and colleagues adapted the Pro-Codes into EpicTags for *in situ* clonal tracking with spatial phenotyping in cancer models [92]. Protein barcode readout was mediated via isotope-labelled antibodies via multiplexed ion beam imaging (EpicMIBI).

EpicMIBI was applied to investigate clonal architecture and spatial organisation in a human small-cell lung cancer (SCLC) line, known to harbour both neuroendocrine (NE) and non-NE features. Twenty uniquely barcoded human SCLC cell lines were pooled and implanted into mice, enabling simultaneous tracking of genetically distinct clones within the same tumour. Spatially resolved imaging revealed that barcoded clones formed spatially coherent patches corresponding to both NE and non-NE phenotypic states, with non-NE patches displaying preferential local expansion. Furthermore, PTEN deletion altered the spatial expansion of both mutant and non-mutant clones [92] , indicating interclonal interactions that influence tumour architecture and clonal fitness *in situ*.

Finally, recent optical barcode strategies encode clonal identity using combinatorial expression of fluorescent proteins. A notable advancement was the lentiviral gene ontology (LeGO) system, employing modular lentiviral vectors to combine up to six fluorescent proteins and generate 41 distinct hues [93] . *Berthelet et al*. used a LeGO-based optical barcoding approach to map the clonal composition of clinically relevant sites of breast cancer metastasis [94] . The authors showed that metastatic sites differ markedly in clonal composition, with lung lesions frequently comprising highly polyclonal mixtures and liver metastases exhibiting more restricted clonal representation. They

further demonstrated that these site-specific clonal architectures are linked to reversible differences in transcriptomic programs, indicating that tumour cell plasticity and local microenvironmental context jointly shape metastatic clonal dynamics.

Together, these studies demonstrate that protein-level barcoding combined with spatially resolved phenotyping enables direct interrogation of clonal architecture within its native microenvironment, revealing how spatial organisation, phenotypic plasticity, and interclonal interactions shape tumour evolution. More broadly, static barcoding has been instrumental in revealing cellular lineage, clonal dynamics, and fate decisions across development and disease. While static barcodes robustly encode clonal identity, they do not provide information on generational order or ancestral history. This has motivated the parallel development of evolvable barcoding strategies, which will now be examined further, before the experimental considerations and constraints pertaining to both static and evolvable systems are discussed in Section 1.2.6.

1.2.4 Evolvable barcodes

Evolving barcodes enable the reconstruction of cellular ancestry through progressive, heritable changes in the barcode sequence that accumulate over successive generations. This cumulative editing is performed using CRISPR–Cas9–based tools. Typically, Cas9 induces double-strand breaks in genomic DNA, which are repaired by nonhomologous end joining (NHEJ). This process is intrinsically error-prone and frequently introduces short insertions or deletions (indels) at the cut site. In practice, synthetic CRISPR target sequences, with guide RNA (gRNA) recognition motifs, are integrated into the host genome as defined arrays. The binding of Cas9, guided by a gRNA, produces stochastic indels at these sites, and the resulting mutation patterns are inherited by all descendant cells. As edits accumulate over time, shared mutation patterns define common ancestry, enabling reconstruction of lineage trees based on relative edit similarity [54] .

1.2.4.1 Early CRISPR-based lineage tracing in development

The first demonstration of CRISPR–Cas9 evolving barcodes for lineage tracing is credited to the GESTALT system (genome editing of synthetic target arrays for lineage tracing) [57] . Cas9 and gRNAs were injected into zebrafish embryos to induce barcode scarring during early development, allowing for the analysis of how early embryonic

cells contribute to adult zebrafish organs. This study established that cumulative indel patterns in genomic DNA could be used to reconstruct lineage relationships across developmental timelines. Subsequent adaptations coupled evolving barcodes to single-cell RNA sequencing (scRNA-seq), enabling simultaneous recovery of CRISPR scars and transcriptomic profiles to link lineage identity and cellular states. For example, the LINNAEUS system (lineage tracing via nuclease-activated editing of ubiquitous sequences) was applied to zebrafish larvae and adult organs to reconstruct developmental lineage trees [64].

These studies showed that evolving barcodes were effective and functional, and they could be combined with high-resolution phenotyping to offer both ancestry reconstruction and molecular insights. However, target arrays are susceptible to progressive loss of editable sequence space, as cumulative indels disrupt gRNA recognition motifs and can ultimately lead to target-site saturation or collapse [47,54]. Thus, array design directly influences recording depth, in which finite editing capacity constrains the temporal resolution of lineage reconstruction.

1.2.4.2 Engineering strategies to extend barcode diversity

Cas9 editing saturation constrained early CRISPR–Cas9 evolvable lineage-tracing tools, resulting in generally short editing periods and limiting their application to rapidly developing organisms or to short experimental windows. Furthermore, because indel generation by traditional Cas9 is stochastic, there was a risk that identical edits could occur independently in different cells, potentially leading to false-positive lineage relationships [47]. To extend barcode diversification, alternative CRISPR architectures have been developed. Homing guide RNA (hgRNA) systems mutate their own spacer sequence, allowing each edit to generate a new targeting configuration and thereby enabling iterative, self-propagating diversification. This approach delays target exhaustion relative to fixed arrays and increases cumulative diversity over extended lineage depth [54].

In parallel, modifications to CRISPR editing chemistry have been introduced to reduce toxicity and improve edit stability (**Figure 1-7**). Most early lineage tracing systems relied on wild-type Cas9, which generates double-strand breaks that can lead to large deletions or chromosomal rearrangements [95]. To mitigate these effects, base editors (BE) fuse a catalytically impaired Cas9 to a nucleoside deaminase, enabling predictable

single-nucleotide substitutions without a double-strand break. Furthermore, prime editing (PE) systems have been implemented that use a Cas9 nickase fused to a reverse transcriptase to install defined edits guided by a prime editing guide RNA (pegRNA). These alternative designs and editing chemistries provide greater control over mutation patterns and reduce editing-associated toxicity, thereby expanding the feasibility, reliability, and longevity of cumulative lineage tracing across diverse systems [51,95].

1.2.4.3 Application of evolvable barcoding to mammalian systems

Many early demonstrations of evolving barcoding used embryonic injection of Cas9 components, and the systems' recording capacity saturated within a relatively short time [66]. Therefore, application to mammalian homeostasis and disease models, such as cancer, required inducible and genetically encoded editing systems [75]. Inducible Cas9 strategies enabled temporal control of editing, allowing lineage tracing during defined developmental windows or during tumour initiation, progression, or treatment exposure [47].

Several studies applied evolvable lineage tracing to investigate development in mice; however, this required the manipulation of embryos for each experiment [47,96]. Instead, the CARLIN mouse was engineered to incorporate a single, stably integrated mutable locus into the germline, allowing inducible editing without extensive embryonic manipulation [67]. By activating Cas9 in adult tissues, CARLIN enabled temporally controlled lineage tracing across diverse systems. The subsequent DARLIN mouse further refined this approach by employing a Cas9 base editor, reducing double-strand break–associated toxicity and increasing barcode diversity [81]. Together, these systems illustrate how editing architecture and chemistry directly influence lineage depth, diversity, and tissue compatibility.

1.2.4.4 Evolvable lineage tracing in cancer models

The capacity to record lineage branching over time is particularly powerful in cancer, where clonal expansion, metastasis, and microenvironmental adaptation unfold dynamically [28,48] .

Simeonov et al. adapted the GESTALT system to generate macsGESTALT (multiplexed, activatable, clonal and subclonal GESTALT), an integrated, inducible lineage tracer for

use in mammalian systems compatible with single-cell measurement of phylogenetic and transcriptional information [77]. The authors applied macsGESTALT to a pancreatic cancer metastasis mouse model, enabling reconstruction of the dissemination of ~28,000 single cells over several weeks across multiple metastatic sites. This approach demonstrated that metastatic capacity arises from rare transitional cellular states. Furthermore, the gene signatures of these high-fitness states are associated with poor clinical outcomes in patients with pancreatic and lung cancer, highlighting their clinical relevance.

More recently, *Koblan et al.* developed PEtracer, an inducible, prime-editing–based lineage tracer that is compatible with both single-cell sequencing and high-resolution spatial transcriptomics [85]. The authors applied the PEtracer to a syngeneic mouse model of breast cancer–lung metastasis to link phylogeny with cellular state and microenvironmental context. They found that metastatic expansion was sustained by heritable, high-fitness transcriptional programs that were selectively reinforced within defined spatial niches. Spatially resolved lineage reconstruction further showed that increased proliferation at the tumour periphery did not necessarily correspond to long-term clonal dominance, distinguishing transient, microenvironment-driven growth from intrinsically encoded fitness advantages. Thereby, demonstrating how environmental cues and intrinsic transcriptional programs interact over time to shape metastatic growth.

Collectively, evolvable barcoding has demonstrated the power of cumulative editing to resolve temporal ancestry and lineage branching with high resolution. Following initial application in developmental biology systems, subsequent adaptation to mammalian models, particularly in cancer, has extended this framework to dissect clonal dynamics, fate diversification, and metastatic progression *in vivo*, thereby adding a temporal dimension to the study of cancer evolution. However, evolvable barcode implementation and interpretability are shaped by editing architecture and downstream readout constraints, considerations that apply across barcoding strategies and are discussed further in Section 1.2.6.

1.2.5 Molecular Recorders

A rapidly advancing technology related to CRISPR-based lineage tracing is **molecular recording**, in which defined biological events are encoded in synthetic DNA loci over time. In these systems, modification of a barcode element is coupled to specific cellular states, signalling pathways, or environmental stimuli, extending beyond ancestry reconstruction to document cellular experience. Such histories may reveal how distinct exposures shape phenotypic divergence among cells that share a common lineage, an important consideration in development, immunology, and cancer [97].

Mechanistically, molecular recorders typically place Cas9 expression under stimulus-responsive promoters, such that editing occurs only when a defined pathway is active. For example, mSCRIBE (mammalian synthetic cellular recorder integrating biological events) uses an inflammation-responsive promoter to drive a self-targeting Cas9, with barcode divergence reflecting the magnitude or duration of inflammatory signalling [98]. As with evolvable lineage tracers, alternative editing chemistries, such as homing guides [75], and base or prime editors [99–101], can be employed to enable more predictable editing and reduced toxicity over extended recording windows (**Figure 1-7**) [95].

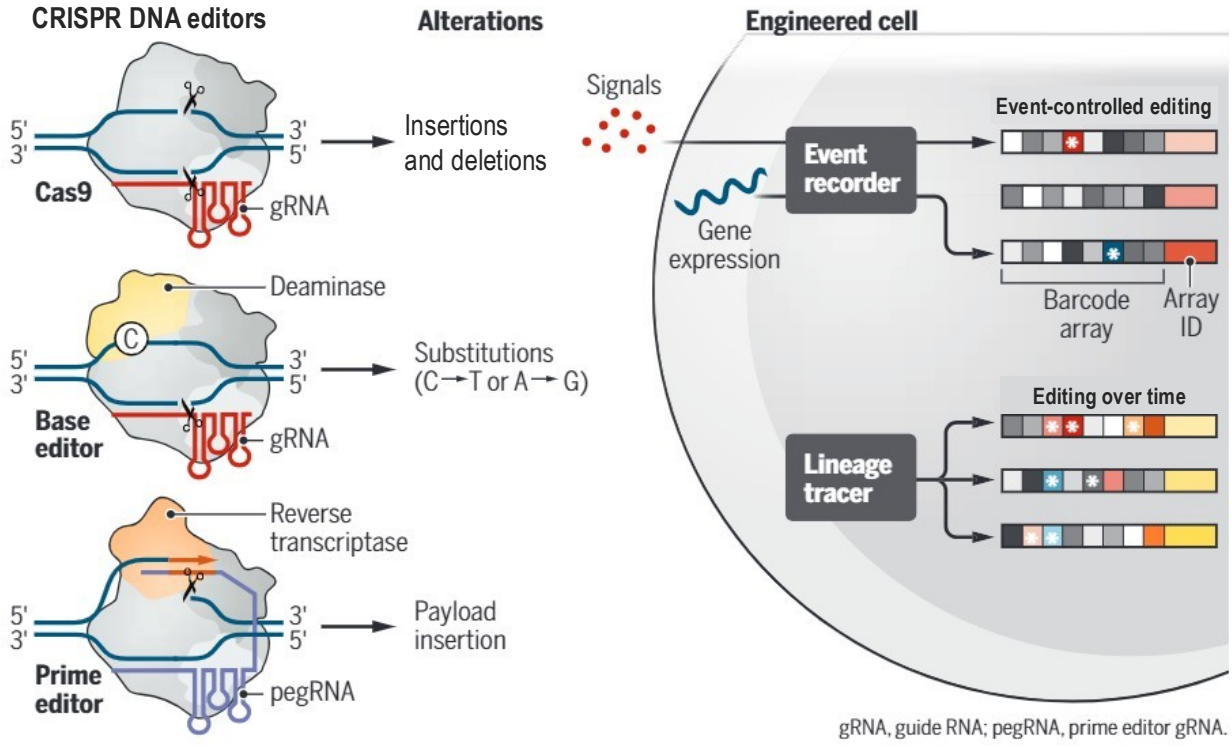


Figure 1-7. CRISPR-mediated DNA editing.

CRISPR-based DNA editors introduce distinct classes of genomic modifications, including insertions and deletions (indels) via Cas9-mediated double-strand breaks, base substitutions via deaminase-coupled base editors, and defined sequence insertions using prime editors. These editing chemistries can be configured either as event-controlled recorders, in which stimulus-responsive editing encodes specific biological signals into synthetic barcode arrays, or as lineage tracers that accumulate heritable edits over time to reconstruct cellular ancestry. Adapted, with permission, from Masuyama et al. [95]

Despite their conceptual appeal, molecular recording systems remain constrained by the field's early stage. DNA-based recorders have a finite information capacity, thereby limiting the complexity and duration of the events they can encode. Moreover, the repertoire of available molecular sensors remains narrow, and stimulus-responsive promoters may exhibit background activity or incomplete specificity, complicating interpretation. In addition, current readout technologies do not provide comprehensive single-cell coverage across large tissues, and computational frameworks for reconstructing high-content biological histories remain underdeveloped [95,102]. Nonetheless, substantial efforts are currently underway to advance these systems, with significant developments anticipated in the near future.

While molecular recording does not directly establish lineage relationships, integrating these rapidly advancing tools with lineage-tracing frameworks presents a promising path for more comprehensive biological analysis. Therefore, allowing recorded events to be mapped onto clonal identity or reconstructed ancestry, linking cellular experience to evolutionary trajectories [103]. Such integration offers a more comprehensive view of how intrinsic programs and microenvironmental influences shape development and tumour evolution. Nevertheless, combining lineage tracing with event recording substantially increases experimental and analytical complexity, requiring careful optimisation of editing architecture, information capacity, and readout compatibility [95,97]. As molecular recording and lineage-tracing technologies continue to advance, they are expected to make an increasingly significant contribution to biological research in the near future.

1.2.6 Cellular barcoding: practical considerations and readout constraints

1.2.6.1 Practical implementation across static and evolvable systems

While static and evolvable barcoding strategies differ in how lineage information is encoded, both are constrained by practical implementation and downstream analytical methods. Library complexity, delivery efficiency, editing kinetics, and reconstruction assumptions impose technical boundaries that determine experimental feasibility and

interpretability [51,54,86]. These considerations determine not only whether a system functions in principle but also whether it produces lineage information that is biologically meaningful.

In static barcoding systems, experimental design is largely governed by library size and delivery parameters. Barcode diversity should exceed the labelling population size to maintain clonal resolution. Infection or transduction efficiency directly influences barcode representation, with high multiplicity compromising clonal uniqueness and low efficiency limiting representation of the starting population [86,89]. Selection steps may introduce bottlenecks that distort barcode representation [51]; therefore, quantifying clonal composition after labelling is essential to establish a baseline for subsequent comparisons. Crucially, usable library size is often constrained by the sensitivity and dimensional capacity of the intended readout platform. Despite these considerations, static systems are experimentally tractable and scalable, enabling stable lineage marking in complex mammalian systems when appropriately implemented [86].

Evolvable barcoding systems introduce additional layers of complexity. Editing efficiency and temporal control determine the depth and resolution of lineage information, while excessive editing can lead to saturation and collapse of phylogenetic signal [54]. In addition, off-target activity and recurrent editing events in CRISPR systems can generate overlapping mutation patterns, complicating the accurate computational reconstruction of lineage relationships. Furthermore, integration with high-dimensional single-cell datasets increases computational burden [51]. While these systems enable inference of lineage branching over time, their implementation requires careful calibration of biological, technical, and analytical parameters to ensure that reconstructed phylogenies remain interpretable.

Together, these considerations emphasise that both static and evolvable systems are constrained by practical implementation rather than barcode architecture alone. The choice of strategy must align with the biological question, experimental context, and required resolution.

1.2.6.2 Readout modality as the dominant constraint on lineage information

Across barcoding systems, a key determinant of biological insight is the modality used to detect barcodes. Readout technique dictates whether lineage can be linked to cellular

phenotypes, such as transcriptional states or protein profiles, or to its native spatial context.

Sequencing-based readouts prioritise molecular depth and barcode diversity. Bulk sequencing enables sensitive quantification of clonal frequency across large populations, facilitating analysis of clonal expansion dynamics [56,88]. However, such approaches lack single-cell resolution and do not directly couple lineage identity to cellular phenotypes. Single-cell sequencing overcomes this limitation by linking barcodes to transcriptomic profiles, thereby enabling analysis of lineage–state relationships at high molecular throughput [76,78,79]. Nevertheless, spatial context is lost during tissue dissociation, protein-level phenotypes are inferred only indirectly, and cost scales with sequencing depth and cell number.

Cytometry-based and spatial readouts permit phenotypic and contextual information. Suspension cytometry platforms can link clonal identity to protein-level phenotypes and functional markers at a large scale [90]. However, barcode dimensionality is often constrained by the number of available detection channels, and spatial dimension is lost. Spatial imaging approaches retain tissue architecture and enable mapping of clonal distribution within microenvironmental niches [85,92,94]. These methods allow interrogation of lineage–ecosystem relationships but often impose constraints on barcode complexity and accompanying phenotyping depth.

Consequently, each readout technique captures a distinct dimension of clonal behaviour. Across modalities, trade-offs exist among barcode complexity, detection sensitivity, spatial preservation, and simultaneous cellular-state information. Thus, a universal constraint encountered by lineage tracing systems is the fragmentation of measurement techniques, in which clonal identity, phenotypic state, and ecological position are typically interrogated through separate experimental frameworks.

Therefore, a more comprehensive characterisation of cellular behaviour within complex systems requires integration of multiple layers of molecular information, linking transcriptional profiles, protein activity, metabolic features, and spatial organisation within a unified analytical framework. Independent of lineage tracing, a growing class of multimodal technologies has emerged to address this broader need for integrative

cellular profiling. The following section reviews these approaches and subsequently examines efforts to incorporate lineage information within such multimodal frameworks.

1.3 Toward multimodal clonal tracing

1.3.1 Beyond single-modality cellular profiling

Cellular identity and behaviour cannot be fully described by a single molecular layer. In complex biological systems, and particularly in cancer, phenotypic state, clonal history, and microenvironmental context are interdependent yet only partially observable through any individual modality [104] . Transcriptomic profiling has transformed the ability to classify cell states at scale, but gene expression alone does not capture protein activity, metabolic function, or tissue-level spatial organisation. As a result, reliance on sequencing-derived measurements alone can obscure functional differences between cells that exhibit similar transcriptional profiles, particularly in contexts where protein activity, cell–cell interactions, and local microenvironmental factors strongly influence behaviour [105] .

This limitation is particularly evident in cancer, where intratumoural heterogeneity arises not only from genetic and clonal diversity but also from spatially structured ecological interactions. The position of a cell within a tumour, its proximity to vasculature, immune infiltrates, or stromal niches, and its local metabolic environment can profoundly influence its phenotype and evolutionary trajectory [106]. Capturing these dimensions requires approaches that extend beyond dissociated sequencing workflows, incorporating spatially resolved measurements and orthogonal molecular readouts [107]. At the same time, high-throughput, suspension-based technologies remain essential for scalable phenotyping, enabling deep protein and RNA profiling across large cell populations and experimental conditions.

Consequently, a growing class of technologies has emerged that interrogates cellular systems across multiple molecular layers and experimental formats, including spatial RNA and protein detection, label-free molecular imaging, and high-parameter cytometric profiling. These approaches provide complementary views of cellular state, function, and context that cannot be accessed through sequencing alone. The following sections review these modalities in turn, before examining efforts to integrate them into unified

multimodal platforms and lineage-tracing frameworks that resolve clonal dynamics alongside cellular phenotype and spatial organisation.

1.3.2 Spatial RNA and protein detection

Spatial context is a critical axis of cellular identity that is lost in dissociative workflows, motivating the development of techniques that preserve tissue architecture while enabling *in situ* molecular profiling. Methods for detecting RNA in tissues have steadily improved in spatial resolution and overall throughput over recent decades, with particularly rapid progress in the last five years driven by advances in probe chemistry, imaging, and sequencing technologies. Early approaches relied on fluorescence *in situ* hybridisation (FISH) to visualise RNA molecules within tissues. As next-generation sequencing technologies advanced, these methods were adapted to retain spatial information while increasing molecular coverage [108].

This progress has given rise to two broad classes of spatial transcriptomic approaches (STs): imaging-based (**Figure 1-8**A and B) and sequencing-based methods (**Figure 1-8**C and D). Imaging-based techniques achieve high spatial resolution by visualising transcripts within intact tissues, while sequencing-based approaches capture RNA on spatially indexed substrates for downstream sequencing. Both strategies enable measurement of RNA expression within tissues, with some platforms supporting detection of thousands of transcripts or near-transcriptome-wide coverage. However, a fundamental trade-off remains between spatial resolution and molecular throughput, as unbiased transcriptome-wide methods typically struggle to achieve true single-cell resolution, whereas targeted approaches that provide subcellular detail often sacrifice scale [109].

Imaging-based ST approaches are further classified into two strategies for RNA measurements, referred to here as cyclic combinatorial decoding (**Figure 1-8**A) and cyclic *in situ* sequencing (**Figure 1-8**B). Both methods use fluorescent probes or labelled nucleotides to decode the positions of tens to several thousand transcripts within cells, often at single-cell or subcellular resolution. Cyclic imaging is used to overcome the spectral limitations of fluorescent tags. Cyclic combinatorial decoding employs DNA scaffold probes complementary to target transcripts, followed by repeated cycles of hybridisation of tagged detection probes, imaging, and probe stripping,

enabling combinatorial decoding [106,109]. Scaffold probe architectures can generate an efficient signal-to-noise ratio by enabling multiple fluorophore-labelled probes to bind to a target simultaneously [110]. Variations of this method are employed by MERFISH [111], seqFISHv1 and v2 [112,113], CosMx [114], and EEL FISH [115].

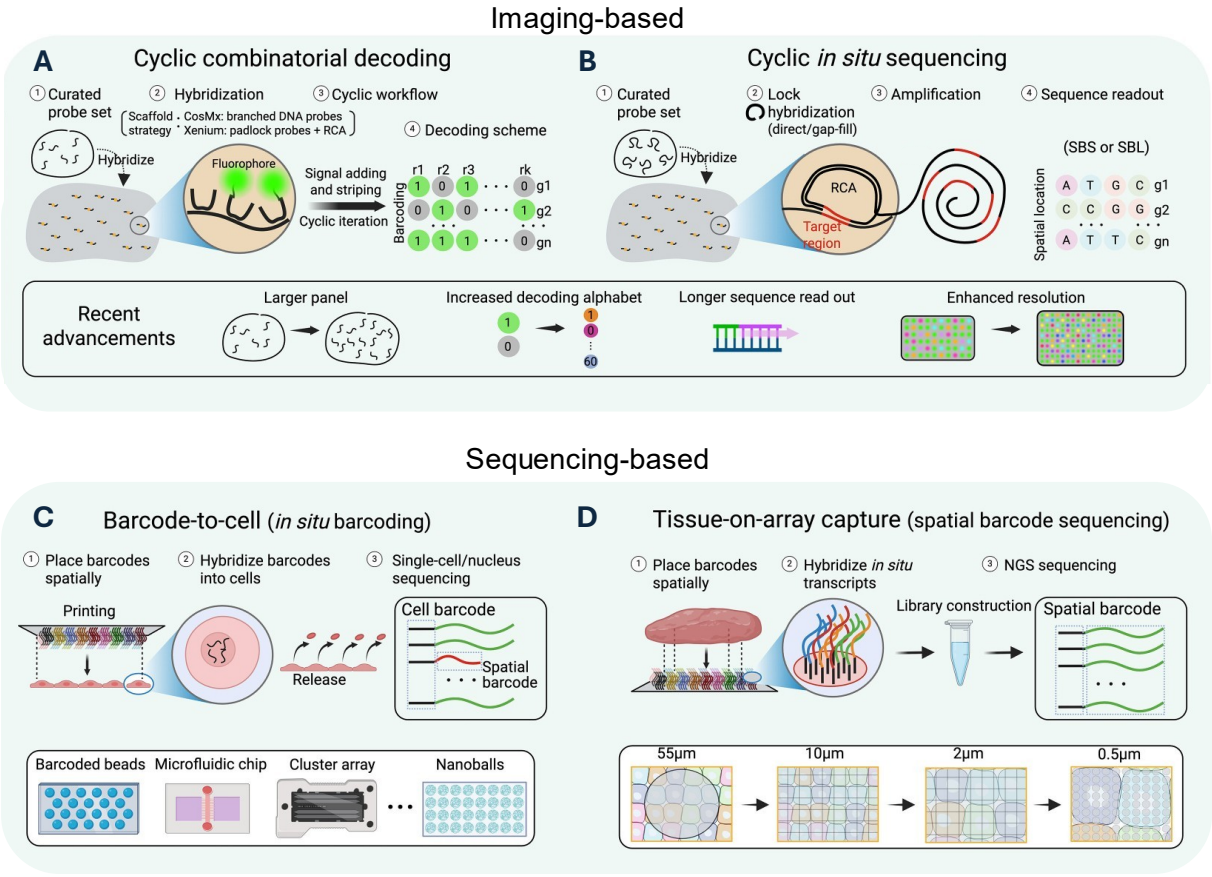


Figure 1-8. Spatial transcriptomics techniques.

In situ (imaging)-based approaches (A) are categorised into cyclic combinatorial decoding and cyclic in situ sequencing (B). Spatial barcoding (sequencing)–based approaches are subdivided into barcode-to-cell capture methods (C, in situ barcoding) and tissue-on-array capture methods (D, spatial barcode sequencing). RCA = rolling circle amplification; SBS = sequencing by synthesis; SBL = sequencing by ligation; NGS = next generation sequencing. *Adapted from Liu et al.* [106]

Alternatively, the 10x Genomics Xenium platform employs rolling-circle amplification (RCA) with padlock probes to target RNA, generating local complementary DNA (cDNA) containing multiple copies of specific sequences used for decoding localised transcript identity [116]. Cyclic *in situ* sequencing approaches also utilise RCA to amplify RNA, thereby increasing signal generation. Then, barcode or transcript amplicons are read optically using sequencing-by-synthesis (SBS) or sequencing-by-ligation (SBL) chemistries, in which fluorophore-labelled nucleotides or oligonucleotides are incorporated or ligated in successive cycles, and the resulting fluorescence is imaged

by fluorescence microscopy. For example, SBS is implemented by FISSEQ [117] and STARmap [118], while BaristaSeq [119] employs SBL.

In contrast, sequencing-based spatial transcriptomic approaches preserve positional information by capturing RNA on spatially indexed surfaces prior to sequencing. These methods can also be broadly divided into two RNA-capture strategies. In barcode-to-cell approaches (**Figure 1-8**C), spatially resolved barcodes are delivered into intact cells or nuclei and subsequently recovered by single-cell or single-nucleus sequencing. For example, specific strategies include spatially barcoded beads (Slide-seq [120], Slide-seqV2 [121]); fluidics-based *in situ* barcoding (DBiTseq [122]); DNA nanoballs (Stereo-seq [123]). In tissue-on-array capture approaches (**Figure 1-8**D), tissue sections are placed directly onto arrays of spatially indexed oligonucleotides that capture transcripts *in situ* prior to library construction and next-generation sequencing. In sequencing-based ST methods, spatial resolution is determined by the size and density of capture features, which may range from multicellular spots to near cellular or subcellular scales [106] .

Sequencing-based STs platforms provide transcriptome-wide coverage without requiring predefined panels, unlike imaging-based approaches, which typically rely on targeted probe sets [109,124]. However, sequencing-based methods can suffer from reduced RNA capture efficiency, particularly in complex or archival tissues [125]. At the same time, the panel capacity of imaging-based spatial transcriptomics platforms continues to expand rapidly. Pre-designed probe sets tailored to specific tissues or microenvironments, together with the option to add custom targets, provide substantial experimental flexibility. For example, the CosMx Whole Transcriptome assay enables subcellular imaging of over 18,000 RNA targets, approaching coverage of the human protein-coding transcriptome [126]. Nevertheless, such large panels require extensive iterative imaging, which can quickly become costly and limit the number of samples that can be analysed.

In parallel with advances in spatial RNA detection, targeted spatial protein profiling technologies have enabled detailed phenotypic characterisation of intact tissues. Microscopy-based approaches visualise proteins using fluorophore-conjugated antibodies (**Figure 1-9**A), allowing spatial localisation at subcellular resolution. Tens of proteins can be imaged by performing iterative cycles of antibody staining, fluorescence imaging, and signal removal. This strategy underpins multiplexed immunohistochemistry

workflows such as CyCIF [127] and IBEX [128], which rely on repeated rounds of fluorophore bleaching or chemical stripping to erase the signal before subsequent staining. While these approaches do not require specialised instrumentation, they are inherently labour-intensive due to repeated manual staining, stripping, and imaging steps.

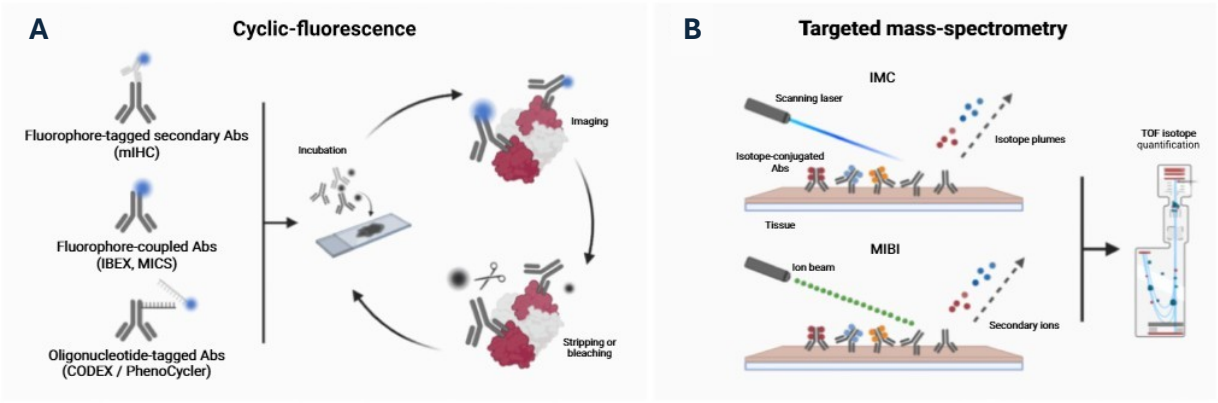


Figure 1-9. Targeted spatial protein detection.

A) *Cyclic immunofluorescence methods for the detection of proteins using antibodies (Abs) labelled with fluorescent tags or short DNA sequences that are recognised by complementary, fluorescently labelled oligonucleotides. Consecutive rounds of incubation, imaging, and stripping allow the detection of a range of epitopes in situ. mIHC = multiplexed immunohistochemistry; IBEX = iterative bleaching extends multiplexity; MICS = MACSima imaging cyclic staining; CODEX = codetection by indexing.*

B) *Targeted mass spectrometry (MS) coupled to time-of-flight (TOF) isotope detection. Isotopes conjugated to Abs are ablated by either a scanning laser (IMC) or an ion beam (MIBI) and registered through TOF-MS. Signals can then be traced back to their point of origin using the scanning apparatus's coordinates. IMC = imaging mass cytometry; MIBI = multiplexed ion beam imaging. Reproduced, with permission, from Toninelli et al.* [109]

To address these limitations, automated platforms for high-parameter protein imaging have been developed that integrate staining, imaging, and signal removal within the instrument. The Lunaphore COMET system automates sequential immunofluorescence by performing cycles of primary antibody binding followed by fluorophore-conjugated secondary antibody detection and on-instrument signal removal [129]. Similarly, co-detection by indexing (CODEX) [130], commercialised as the Akoya PhenoCycler-Fusion, uses antibodies conjugated to unique DNA barcodes. In this approach, all antibodies are applied simultaneously, and fluorescently labelled oligonucleotides complementary to each barcode are iteratively hybridised, imaged, and removed, enabling highly multiplexed protein detection without cumulative antibody stripping.

An orthogonal class of spatial protein-profiling techniques is based on targeted mass spectrometry rather than fluorescence (**Figure 1-9**B). Imaging mass cytometry (IMC) uses antibodies conjugated to metal isotopes, which are detected by time-of-flight mass

spectrometry following laser ablation of the tissue. Each laser pulse vaporises a small region of the sample, and the released isotopes are quantified to generate spatially resolved protein maps [109] . Multiplexed ion beam imaging (MIBI) employs a similar principle but uses a primary ion beam to sputter secondary ions from the tissue surface, which are then analysed by mass spectrometry [131]. By replacing fluorophores with metal tags, IMC and MIBI circumvent spectral limitations entirely and enable simultaneous detection of dozens of proteins within intact tissue sections [109].

Collectively, targeted spatial RNA and protein detection methods provide cellular-resolution measurements and rich phenotypic annotation while preserving tissue architecture. However, these approaches are inherently constrained by predefined marker panels, prior biological assumptions, and practical limits on throughput and scalability. As a result, complementary spatial approaches that provide broader, label-free molecular information across multiple phenotypic layers should be considered.

1.3.3 Spatial metabolic and molecular imaging

In parallel with targeted and sequencing-based spatial profiling approaches, mass spectrometry imaging (MSI) has emerged as a powerful, label-free modality for mapping molecular distributions directly within tissue sections (**Figure 1-10**). MSI enables label-free, spatially resolved detection of molecular species directly from tissue sections, including metabolites, lipids, glycans, and small peptides. In contrast to antibody- or probe-based imaging, MSI does not require prior target selection, providing an unbiased molecular profile of intact tissue [106]. Samples are systematically scanned with light or particle beams to ionise surface molecules, which are subsequently detected by time-of-flight mass spectrometry, allowing spatial mapping of molecular distributions across intact tissues [104].

Among MSI modalities, matrix-assisted laser desorption/ionisation (MALDI) is the most widely used MSI ionisation approach. MALDI-MSI enables detection of a broad range of molecular species, from small metabolites and lipids to peptides and, in some cases, proteins, while preserving their spatial organisation within tissue sections. A chemical matrix is applied to the tissue surface to promote efficient energy absorption and ionisation during laser irradiation, thereby enabling the generation of intact molecular ions across a wide mass range. Spatially resolved mass spectra are acquired by

rastering the laser across the sample, with typical spatial resolutions of ~10–50 µm [133] . Importantly, matrix choice influences ionisation efficiency and molecular coverage, such that different matrices are more suitable for specific classes of analytes [134] .

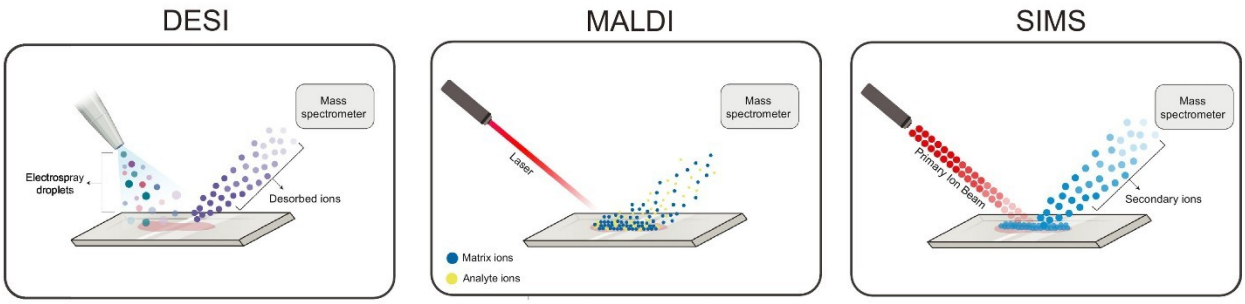


Figure 1-10. Mass spectrometry imaging techniques: DESI, MALDI, and SIMS.

Schematic representative of DESI, MALDI, and SIMS ionisation processes. DESI is an ambient ionisation method that employs charged solvent sprays on a sample surface to facilitate the desorption and ionisation of surface molecules. MALDI ionises molecules using a pulsed laser beam that raster-scans a matrix-coated, crystallised tissue surface. SIMS uses a high-energy primary ion beam to sputter material from the surface. The secondary ions generated are accelerated, resolved, and measured by the mass spectrometer. DESI = desorption electrospray ionisation; MALDI = matrix-assisted laser desorption ionisation; SIMS = secondary ionisation mass spectrometry. Adapted, with permission, from Rajbhandari et al.[132]

Beyond MALDI-MSI, a growing range of mass spectrometry imaging techniques reflects the field's rapid expansion. Desorption electrospray ionisation (DESI) enables ambient ionisation with minimal sample preparation and is well suited to detecting small molecules, lipids, and metabolites. This makes it attractive for rapid or clinically oriented applications; however, DESI typically offers lower spatial resolution and sensitivity compared with MALDI [135] . In contrast, secondary ion mass spectrometry (SIMS) employs a focused ion beam to achieve submicron spatial resolution, enabling highly detailed molecular localisation at the expense of reduced molecular coverage [133].

Together, these approaches highlight inherent trade-offs between spatial resolution and molecular breadth, while ongoing innovations such as post-ionisation strategies, including MALDI-2, continue to improve sensitivity and expand detectable molecular classes [133,136].

1.3.4 Suspension-based cytometry profiling

While spatial technologies enable *in situ* exploration of cellular organisation, many biological questions benefit from, or explicitly require, profiling in suspension. These

approaches are especially valuable for functional assays, drug screening, and studies of rare populations, where statistical power and sampling depth are critical.

Single-cell RNA sequencing remains the standard for transcriptome-wide analysis, enabling the characterisation of thousands of cells in a single experiment. However, sequencing techniques are constrained by high costs, protocol complexity, available sequencing depths, and computationally intensive data analysis [137]. A complementary strategy is to quantify a smaller set of transcripts while substantially increasing the number of cells analysed. Flow cytometry facilitates this by measuring multiple parameters in hundreds to thousands of cells per second, enabling experiments that encompass millions of cells. Moreover, flow cytometry platforms are widely available in research settings and can be employed repeatedly at comparatively lower cost [138] .

Early adaptations of fluorescence *in situ* hybridisation (FISH) protocols enabled quantification of gene expression on flow cytometry platforms [139] . These approaches required exceptionally bright FISH signals with high signal-to-noise ratios because flow cytometry lacks the subcellular resolution needed to distinguish individual RNA molecules from diffuse background. To increase signal strength, DNA padlock probes, combined with rolling circle amplification (RCA) and branched DNA technologies, have been used in microscopy and flow cytometry [140,141]. However, multiplexing in flow cytometry is constrained by spectral overlap among fluorophores. Rapidly developing spectral flow cytometry instruments address this by capturing complete emission spectra of fluorochromes, creating a spectral signature for each. Many fluorochromes can be used, then computationally unmixed to distinguish overlapping signals, enabling simultaneous measurement of up to 40 markers [138].

Alternatively, mass cytometry combines flow cytometry with mass spectrometry to determine the chemical composition of analytes by measuring the mass-to-charge (m/z) ratios of ions. In mass cytometry, cells are stained with antibodies conjugated to transition metals, enabling simultaneous measurement of up to 60 parameters, with a theoretical capacity for up to 120. Suspension Cytometry by Time of Flight (CyTOF) distinguishes isotope-labelled antibodies with minimal channel overlap, thereby elucidating protein expression [138,142]. A workflow has also been developed to detect RNA transcripts on these platforms. The proximity ligation assay for RNA (PLAYR) enables highly multiplexed RNA quantification alongside antibody staining using mass

cytometry, supporting integrated analysis of transcriptional and protein-level features in suspension [141]. The PLAYR protocol provides padlock probes to detect at least 27 specific RNAs via signal amplification by RCA, thereby overcoming interference issues observed in branched-DNA methods [137].

Suspension-based cytometry platforms offer exceptional scalability and high-dimensional phenotyping, making them well-suited for population-level analyses, functional screening, and the study of rare cell states [142]. However, because measurements are performed on dissociated cells, spatial relationships and tissue-level context are not retained. This distinction highlights how suspension-based profiling and spatial approaches address complementary biological questions, motivating the development of experimental frameworks that can bridge these analytical regimes where appropriate.

1.3.5 Integrated multimodal workflows and platforms

1.3.5.1 Multiomic phenotyping approaches

The rise of diverse measurement modalities has catalysed a rapidly advancing landscape toward multiomic profiling that integrates genomic, epigenomic, transcriptomic, proteomic, and metabolomic analyses, increasingly with spatial resolution.

Over the past decade, DNA and RNA sequencing workflows have expanded to incorporate epigenomic layers such as DNA methylation and chromatin accessibility, enabling deeper insights into gene regulation and epigenetic variation [143]. Early examples include scM&T-seq for parallel single-cell genome-wide methylome and transcriptional sequencing [144] and sci-CAR-seq for joint profiling of chromatin accessibility and mRNA in single cells [145].

Further workflows integrate simultaneous single-cell RNA and protein detection, enabling direct investigation of how transcript abundance and protein activity drive essential biochemical processes [143]. CITE-seq uses DNA-barcoded antibodies to measure cell-surface proteins and mRNA within the same scRNA-seq experiment [146]. A significant orthogonal advance achieved by the PLAYR assay, as discussed in Section

1.3.4, enables simultaneous detection of more than 40 targeted proteins and RNAs via CyTOF [141].

Recent years have also seen an exponential rise in the number of spatial measurement technologies which aim to profile cells and tissues within their natural context. This has enabled the development of spatial multiomics techniques that aim to map intact tissues across multiple molecular levels. This is achieved by analysing fresh-frozen samples or FFPE sections using a combination of single-omic methods. Currently, profiling techniques are often applied to adjacent sections or sequentially to the same section, provided analyte integrity is maintained. However, simultaneous measurement of multiple analytes on the same tissue section may also be possible, depending on the technique (**Figure 1-11**A) [104]. As these imaging- and sequencing-based spatial-omics expand, the number of analytes measurable at the molecular level is increasing. Additionally, workflows are often complemented with morphological context through histological staining, such as haematoxylin and eosin (H&E), on the same or adjacent sections [104].

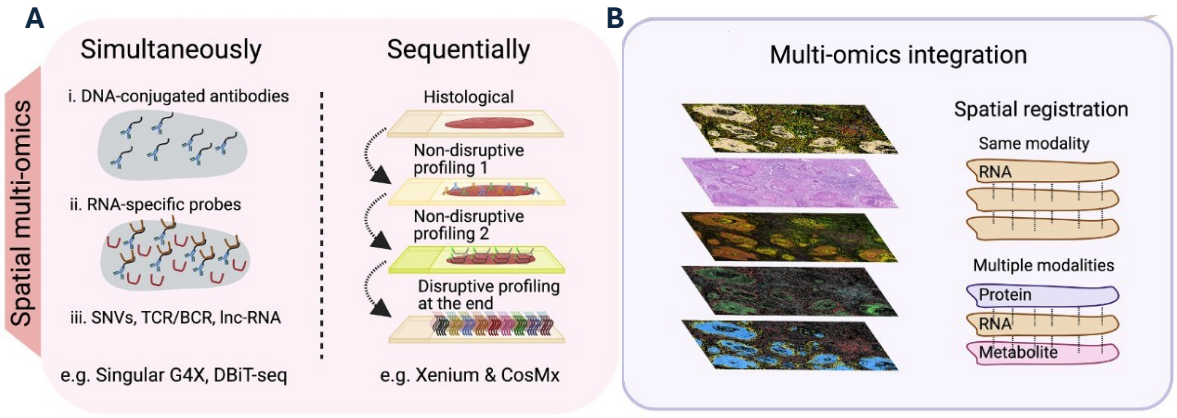


Figure 1-11. Integrative spatial multiomic workflows.

A) *Advancing spatial multiomics can be conducted simultaneously, targeting multiple phenotypic layers on the same sample, or sequentially.*

B) *Integration of multiomic datasets must then be performed using spatial registration techniques. Adapted from Liu et al.* [106]

DBiT-seq achieved a technical advance by enabling transcriptome-wide spatial RNA profiling, together with tens of proteins [122]. Existing RNA detection platforms have extended protein co-detection capabilities, with varying numbers of measured molecules, including STARmap PLUS [147] and spatial CITE-seq [148]. Academic work underpinning GeoMx [149] and CosMx [114] commercial platforms demonstrated workflows capable of measuring tens to hundreds of RNAs and proteins. Other marketed

platforms, including MERSCOPE and Visium, enable transcriptome-wide coverage with tens to hundreds of proteins. Additional methods of RNA–protein co-detection demonstrated spatial context through imaging-based approaches. Although subject to lower throughput, accessibility of imaging technologies and comparatively lower costs enable scalable experimental applications. For example, adaptations of RNAscope, a low-plex fluorescence *in situ* hybridisation method, have been combined with imaging mass cytometry for dual RNA and protein profiling [150]. Furthermore, RNAscope was integrated with CODEX, enabling high-plex antibody staining with RNA measurement [151].

Another axis of multimodal integration involves combining unbiased mass spectrometry imaging (MSI) with transcript-level analysis. For example, MALDI-MSI enables spatial metabolomic profiling at near single-cell resolution. However, integrating MSI with additional molecular analyses remains challenging due to specialised sample-preparation requirements and the comparatively lower durability of metabolites [104]. Despite these challenges, several new workflows have begun to bridge MSI with transcriptomic analyses. For example, the metaFISH pipeline applies low-plex fluorescence *in situ* hybridisation after MALDI-MSI, leveraging improved MALDI techniques that preserve RNA integrity for subsequent hybridisation [152]. Moreover, workflows have emerged for transcriptome-wide spatial transcriptomics post-MSI, namely DESI and MALDI, followed by Visium [153–155]. Alternatively, the OpenFISH workflow utilises padlock probes to detect hundreds of transcripts on tissue sections after performing MALDI-MSI, without the need for proprietary hardware [156].

Once multimodal datasets have been successfully acquired, an additional layer of complexity arises in their analysis, particularly in spatial settings (**Figure 1-11**B). Spatial multiomic data vary widely in resolution, feature type, and tissue coverage, complicating efforts to derive unified biological insights. In particular, integrating data across different spatial platforms introduces additional challenges, as differences in measurement scales, feature parameters, and sampling geometries must be reconciled. Current approaches typically integrate modalities by aligning shared molecular features, matching measurements from corresponding spatial locations or tissue sections, or applying computational models that learn joint representations of molecular profiles and spatial organisation [104,106]. Despite rapid methodological progress, robust integration across platforms remains difficult, particularly at scale. Addressing these challenges is

increasingly recognised as essential for constructing coherent, interpretable multiomic atlases of complex diseases such as cancer.

Together, these advances demonstrate the growing capacity to profile cellular states across molecular and spatial dimensions. However, in the absence of lineage information, these workflows remain inherently limited in their ability to resolve clonal relationships and evolutionary dynamics.

1.3.5.2 Multimodal lineage tracing approaches

Recent years have seen cellular barcoding expand beyond single-readout modalities, with several platforms aiming to develop barcodes interpretable across genomic, transcriptomic, and spatial technologies.

The ClonMapper, for example, is widely regarded as a benchmark static barcoding system, offering robust clonal identification through genomic DNA sequencing, single-cell RNA sequencing (scRNA-seq), and live clone isolation for downstream functional analyses [76]. Its highly diverse lentiviral library (10^7 barcodes) enables linking clonal identity to transcriptional profiles and subsequent phenotypic states, as assessed through additional functional analyses following clonal retrieval. A further study employs two orthogonal barcoding strategies: a Perturb-seq barcode to encode pooled genetic perturbations captured at the transcriptomic level, and a genomically integrated GALILEO barcode to enable stable clonal lineage tracing compatible with epigenetic and genetic readouts [157]. While these systems generate significant multimodal data, they rely on sequencing-based readouts, which can limit the scalability of experiments due to cost and intensive workflows. Furthermore, barcode designs exhibit limited inherent compatibility with spatial readout adaptation.

MEMOIR is an early spatially detailed lineage-tracing strategy for recording dynamic event histories. The method uses single-molecule fluorescence *in situ* hybridisation (FISH) to read out static and edited sequences, allowing reconstruction of cellular lineage over multiple cell divisions [58]. MEMOIR employs a CRISPR-based molecular recording strategy in which lineage information is encoded through irreversible edits at predefined genomic loci, resulting in finite recording capacity and limited temporal resolution. Barcode diversity and scalability are therefore constrained, particularly in complex or long-term biological processes that involve extensive clonal expansion.

Finally, MEMOIR links lineage recording and readout to a single, sequential *in situ* imaging strategy, limiting the ability to integrate lineage information across different spatial or multimodal measurement technologies.

The Zombie strategy reads CRISPR base edited barcodes *in situ*, also using cycles of FISH and fluorescence imaging [72]. The system demonstrated detection of evolving sequences using competing matched and mismatched probes *in vitro*. It employs phage RNA polymerases to transcribe barcodes in fixed cells, thereby overcoming challenges such as stochastic silencing that can limit the detection of expressed barcodes. However, compatibility with other workflows is limited as the barcodes are not inherently transcribed. Although 81 barcodes were demonstrated, increasing diversity through additional editable arrays is possible, but it would require further rounds of time-consuming imaging cycles. Additionally, the reliance on a highly specialised, custom-built microscopy platform substantially limits the method's accessibility and restricts its adoption by the wider research community.

Recently, dynamic barcode systems have been adapted for readout with advancing spatial transcriptomics techniques. The KP-Tracer is a genetically engineered mouse model of lung cancer for evolvable lineage tracing with single-cell RNA-seq readout [158]. The KP-tracer was adapted to obtain spatially resolved datasets by employing Slide-seq, which uses beads to capture transcripts for scRNA-seq and Slide-tags to profile individual nuclei, enabling spatially resolved lineage tracing and unbiased transcript measurement at 10 μm resolution [159]. Thus, spatial resolution is constrained by bead size and capture density, preventing single-cell reconstruction of phylogenies in smaller spatial neighbourhoods. In addition, lineage readout remains sequencing-centric, limiting flexibility for integration with alternative imaging or cytometric modalities, while the substantial costs of these platforms further constrain the number of samples and experimental conditions that can be analysed in practice.

The recently reported PEtracer measures lineage information and spatial organisation at single-cell resolution *in vivo*. It is compatible with scRNA-seq and imaging-based MERFISH spatial transcriptomics via T7 *in situ* transcription, using an adapted Zombie protocol [160]. This enables high-throughput, simultaneous *in situ* detection of barcodes and transcripts. However, T7 amplification involves several morphologically destructive steps to generate RNA and preserve its position, limiting potential for downstream

phenotyping, such as immunofluorescence and histological staining. Additional considerations include the elevated costs of commercial spatial transcriptomic platforms, which limit the number of samples that can be feasibly analysed. Furthermore, the complexity of the evolvable lineage recording and decoding frameworks necessitates specialised, computationally intensive analysis pipelines. This limits broader adoption in experimental settings that primarily aim to track clonal lineages across multiple measurement scales.

Collectively, these approaches highlight substantial progress in integrating lineage tracing with multimodal readouts but also underscore persistent, often unavoidable trade-offs among scalability, spatial resolution, phenotypic breadth, and technical accessibility.

1.3.6 Persistent constraints in multimodal clonal tracing

Despite significant advancements in lineage tracing and multimodal phenotyping technologies, a common set of constraints persists across current approaches. Static barcode libraries provide diversity and stability for long-term lineage tracking, whereas evolving libraries enable temporal resolution but introduce additional technical complexity and finite recording capacity. For both barcode classes, integrating spatial resolution and broad phenotypic readout remains challenging, as lineage-detection and molecular-profiling workflows must be compatible and co-registered *in situ*. Although advances in RNA, protein, and molecular detection have expanded the scope of multimodal measurements, robust integration with lineage tracing across multiple phenotyping levels remains lacking. Furthermore, decoding highly diverse barcode libraries within their native biological context continues to require specialised, high-plex workflows that are difficult to generalise across platforms.

Across existing systems, these limitations manifest as persistent trade-offs between lineage resolution, phenotypic depth, spatial context, scalability, and accessibility. These constraints largely arise because barcode architectures are typically designed around a specific readout modality, thereby limiting compatibility with alternative measurement technologies for which they were not originally intended. Sequencing-based barcode readout enables deployment of high-complexity libraries with transcriptome-level profiling but disrupts tissue architecture. Spatial sequencing and imaging-based

methods preserve barcode localisation yet often at the expense of barcode diversity, limited multimodal compatibility, or reliance on resource-intensive detection platforms. Suspension- and imaging-based mass cytometry has demonstrated epitope-level barcoding for protein phenotyping, but clonal-level tracing is not compatible due to inherent limits on library complexity.

In summary, persistent limitations in current multimodal lineage-tracing platforms restrict the ability to comprehensively resolve clonal dynamics within their native biological context. These constraints underscore a gap between existing technological capabilities and the integrated, high-resolution datasets required to interrogate complex systems such as cancer. This unmet need defines the rationale and scope of this thesis, as outlined in the following section.

1.4 Rationale and scope of this thesis

1.4.1 Perspectives and outlook

Advances in high-parameter profiling techniques and multiomic workflows have enabled comprehensive characterisation of cells, including their microenvironmental context. These approaches are particularly powerful for studying cancer, where evolutionary trajectories, clonal dynamics, and ecosystem interactions shape disease behaviour. Integrating clonal identity with multiomic phenotyping, including transcriptomic, proteomic, and metabolic measurements, offers a unified framework to understand how clones expand, interact, and adapt. Accordingly, meaningful investigation of cancer evolution increasingly depends on integrated, clonal-resolution datasets [89].

From this perspective, such integration is essential for a mechanistic understanding of intratumoural heterogeneity, therapy resistance, and disease progression. Realising this potential requires experimental systems that support controlled perturbation, scalable model platforms that reflect clinically relevant heterogeneity, and analytical frameworks that integrate multimodal lineage and phenotypic data. Together, these considerations highlight a clear shift away from single-modality lineage tracing toward framework-based approaches that comprehensively interrogate clonal dynamics within their native biological context [37,48,51,89] .

Nonetheless, no existing framework offers a versatile, accessible barcode that is detectable across **multiple platforms**, enables **spatial readout**, and is compatible with both **sequencing** and **cytometry-based** assays. This gap limits the ability to connect clonal architecture with multiomic phenotypes in a scalable, generalisable manner.

1.4.2 X-CODE as a flexible, scalable barcoding framework

Addressing these constraints led to the development of **X-CODE**, a novel static RNA barcoding framework, designed to support flexible, cross-platform and spatially resolved lineage tracing with multimodal phenotyping. Generation of high-complexity barcode libraries from nucleic acid sequences enables detection of clonal identity in sequencing-based analyses. Yet, when appropriately designed, expressed barcode transcripts can also be detected using complementary DNA probes, offering an alternative to sequencing-based readout [56,70,160,161]. Probe-based detection decouples barcode identity from the detection modality, allowing the same RNA barcode to be visualised with fluorescent or isotope-conjugated probes, via imaging and cytometry platforms, while remaining compatible with phenotypic and spatial assays.

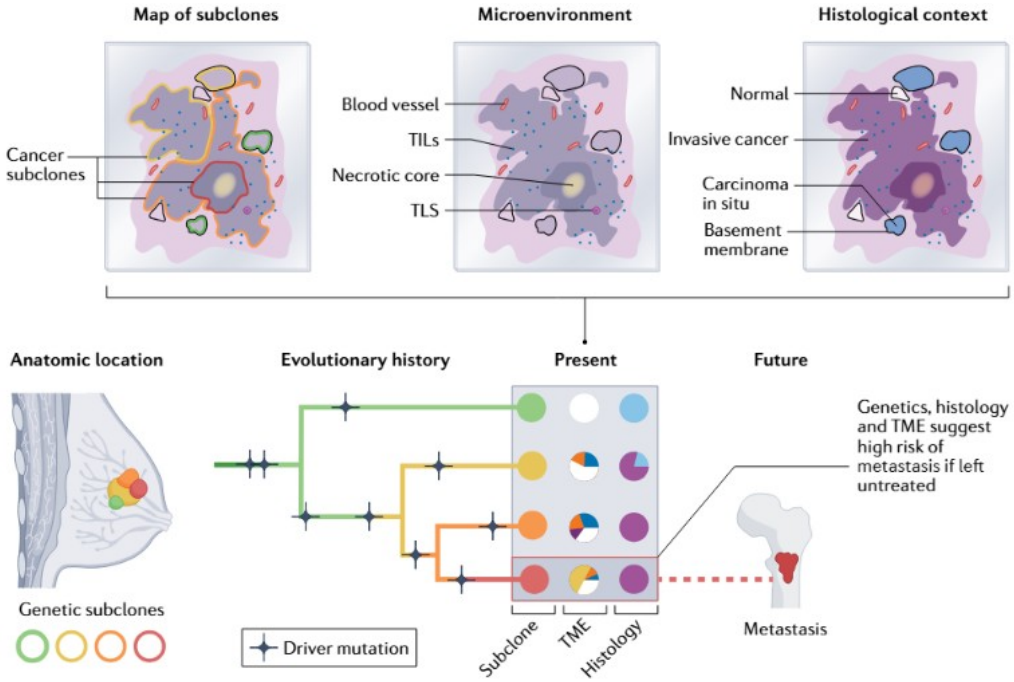


Figure 1-12. Generating an integrated, multimodal picture of tumour heterogeneity.

Integrating analyses of multimodal samples, such as clonal maps, tumour microenvironment (TME) composition, and histological context, may help identify cellular states that affect disease progression and are clinically relevant, which might be missed when using each component alone. TILs = tumour-infiltrating lymphocytes; TLS = tertiary lymphoid structure. Reproduced, with permission, from Seferbekova et al. [37]

This readout flexibility enables a unified barcoding framework that can be deployed across flow cytometry, mass cytometry, and spatial imaging, such as imaging mass cytometry and multiplexed fluorescence imaging. This facilitates a direct link between clonal structure and transcriptional states, protein phenotypes, and tissue architecture. As illustrated in **Figure 1-12**, X-CODE aims to provide an integrated workflow that overlays clonal maps with microenvironmental features and histological context, thereby building a multimodal picture of tumour heterogeneity. This multi-layered perspective has the potential to reveal biologically meaningful patterns with clinical relevance, such as the emergence of adverse lineages with microenvironment-dependent vulnerabilities, that would not be visible with any single modality.

Collectively, these insights establish the aim and main objectives of the thesis, as outlined below.

Chapter 2 Aim and objectives

The main aim of this thesis is to develop a versatile RNA barcoding system for scalable clonal tracing with cross-platform barcode retention, allowing spatial resolution and integrated multi-modal phenotyping (**Figure 2-1**).

This project engineers a highly diverse barcode library from combinatorial barcode units in Chapter 4. It establishes reliable detection protocols across various platforms, including high-mass cytometry by time-of-flight (CyTOF) and high-parameter imaging, as discussed in Chapters 5 and 6, respectively. Additional phenotyping is incorporated through unbiased mass spectrometry imaging in Chapter 7. Finally, it demonstrates biological and technical feasibility in a proof-of-principle system, specifically applied to study tumour heterogeneity in a syngeneic mouse model of prostate cancer evolution, as described in Chapter 8.

This comprehensive, multi-layered X-CODE framework aims to enable high-resolution mapping of clonal architectures and reveal the molecular features that influence tumour dynamics.

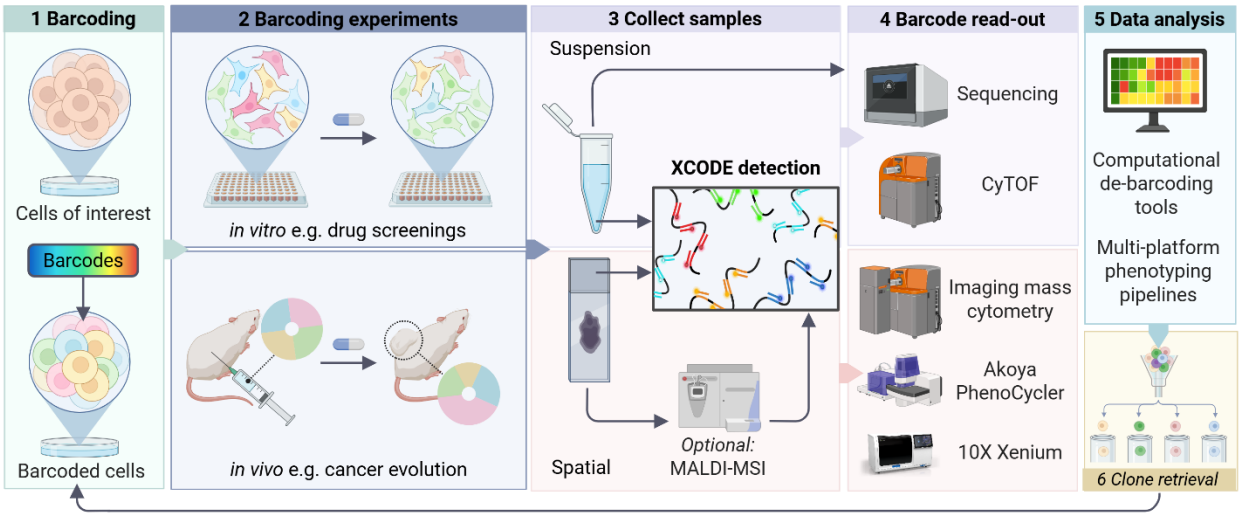


Figure 2-1. Representative workflow using the X-CODE framework.

1) *Barcodes are introduced into cells via lentiviral integration.* ***2)*** *The barcoded population is used for in vitro or in vivo investigations.* ***3)*** *Longitudinal sampling involves collecting cells in suspension or preserving intact samples for spatial techniques or dividing samples for both. XCODE detection protocols are performed as necessary, with matrix-assisted laser desorption/ionisation (MALDI) mass spectrometry imaging (MSI) performed prior to XCODE detection.* ***4)*** *Imaging mass cytometry and high-parameter fluorescence imaging (via Akoya PhenoCycler) detect barcodes and allow for antibody staining. Mass cytometry by time-of-flight (CyTOF) is used for in-suspension barcode readout and protein phenotyping. Bulk sequencing provides overall barcode proportions, while single-cell RNA sequencing offers transcriptomic data.* ***5)*** *Data analysis pipelines support de-barcoding and cellular phenotyping. Subsequently, the construct enables the retrieval of interesting clones from archival or intermediate samples. Created with BioRender.com.*

2.1 Objectives

Objective 1 *– To develop a scalable combinatorial RNA barcode library constructed from predefined barcode units.*

Objective 2 *– To establish robust, multi-platform barcode detection protocols and integrated phenotyping workflows.*

Objective 3 *– To demonstrate proof-of-principle application of the X-CODE system for investigating clonal dynamics during cancer evolution.*

Chapter 3 Materials and methods

3.1 Core reagents

Core reagents that are used across methodologies are listed below.

- UltraPure DNase/RNase-Free Distilled Water, hereafter referred to as RNase-free water (Thermo Fisher; #10977035)
- 10X Phosphate-Buffered Saline, pH 7.4 (PBS, Thermo Fisher; #AM9625)
- 16% (wt/vol) Paraformaldehyde (PFA, Thermo Fisher; #11400580)
- 2-methyl butane (isopentane) (Sigma-Aldrich; #M32631)
- Methanol (VWR; #20847.307)
- Ethanol (VWR; #20821.330)
- RNase*Zap* RNase Decontamination Solution (Thermo Fisher; #AM9780)

3.2 Oligonucleotides

All oligonucleotides were purchased from Integrated DNA Technologies (IDT) as 25 nmol or 100 nmol custom DNA Oligos or MiniGene constructs, depending on use, sequence length and required modifications as indicated below.

3.2.1 X-CODE barcode oligonucleotides

Barcode oligonucleotides for X-CODE 15K and 2M library generation were ordered as MiniGene constructs in the pUCIDT (ampicillin) vector.

3.2.2 Primers for X-CODE cloning

Primers were ordered as 25 nmol custom DNA Oligos. Gibson assembly primers for generating simple barcodes A, B, and C barcodes are detailed in **Supplementary Table 1** (see Section 3.6 for method details). Additionally, Gibson assembly primers for assembling the X-CODE library backbone are detailed in **Supplementary Table 4** (see Section 3.8.1 for method details). Polymerase chain reaction (PCR) primers for

generating the add-on sequences for the X-CODE library backbone are listed in **Supplementary Table 5** (see Section 3.7 for method details).

3.2.3 Hybridisation probes

X-CODE hybridisation probes (padlock probes, 57 bp), targeting the variable binding sequence, were ordered as 25- or 100-nmol custom DNA Oligos with a 5'-end phosphorylation modification (/5Phos/) to facilitate ligation and purified by standard desalting.

The X-CODE universal probe (31 bp), targeting the universal binding sequence, was ordered as a 100 nmol custom DNA oligo with a 3'-end phosphorothioate modification (*_) to prevent nuclease degradation and purified by standard desalting.

Specific sequences for probes targeting simple barcodes A, B, and C are detailed in **Supplementary Table 2**, and for probes targeting X-CODE library barcode units 1–27 are detailed in **Supplementary Table 6**. Additionally, the specific probe sequences targeting β-actin mRNA are listed in **Supplementary Table 3**.

3.2.4 Fluorophore detection probes

X-CODE detection fluorescent probes (21 bp) were ordered as 100 nmol custom DNA oligos with 5'-end fluorescent modifications, purified by standard desalting. Specific sequences, and their fluorescent labels, for probes targeting simple barcodes A, B, and C are detailed in **Supplementary Table 2**, and for probes targeting X-CODE library barcode units 1–27 are detailed in **Supplementary Table 10**. Additionally, fluorescent probes for β-actin mRNA detection are listed in **Supplementary Table 3**.

3.2.5 Thiol detection probes

X-CODE detection probes (21 bp) for isotope-conjugation, detailed in Section 3.15.1, were ordered as 100 nmol custom DNA Oligos with a 5'-end thiol modification (–SH), purified by standard desalting. Specific sequences for probes targeting simple barcodes A, B, and C are detailed in **Supplementary Table 2**, and for probes targeting X-CODE library barcode units 1–27 are detailed in **Supplementary Table 8** and **Supplementary**

Table 9. Additionally, thiol probes for β-actin mRNA detection are listed in **Supplementary Table 3**.

3.3 Antibodies

Antibodies for CyTOF mass cytometry were kindly supplied by the Tumour Functional Heterogeneity Lab at the Institute of Cancer Research.

Table 3-1. Antibodies for CyTOF mass cytometry.

Target	Clone	Isotope conjugation
pRB	J112-906	150Nd
Ki-67	B56	209Bi
pERK1/2 [Thr202/Tyr204]	D1314.4E	167Er
P4EBP1 [Thr37/46]	236B4	149Sm
CD24	M1/69	160Gd
CD49f	GoH3	141Pr

Table 3-2. Antibodies for fluorescence microscopy.

Target	Clone	Fluorophore	Supplier
Na/K-ATPase	EP1845Y	Alexa Fluor 488	Abcam; #ab197713
Pan cytokeratin	AE1/AE3	Alexa Fluor 488	Thermo Fisher; #53-9003-82

Antibodies for Akoya PhenoCycler-Fusion were kindly conjugated by the Gorrini Lab at the University of Leeds.

Table 3-3. Antibodies for Akoya PhenoCycler-Fusion.

Antibodies indicated by an asterisk (*) were custom-conjugated according to Akoya's guidelines.

Target	Akoya Barcode (BX)	Akoya Reporter (RX)	Supplier
CK5	BX049	RX049-ATTO550	Akoya; #5250019
CK8	BX070	RX070-Alexa Fluor 647	Akoya; #5550024
CD45	BX007	RX007-Alexa Fluor 488	Akoya; #4150002
Vimentin *	BX031	RX070-Alexa Fluor 488	Cell Signalling Technology; #46173
AR	BX052	RX052-ATTO550	Akoya; #5250012
Ki-67	BX047	RX047-ATTO550	Akoya; #4250019
Pan cytokeratin	BX019	RX019-Alexa Fluor 488	Akoya; #6150007
Na/K-ATPase *	BX032	RX032-ATTO550	Abcam; #ab167390
	BX017	RX017-ATTO550	Abcam; #ab167390

3.4 Tissue culture

3.4.1 Cell lines

TRAMP-C1 (ATCC; #CRL-2730), RM-1 (ATCC; #CRL-3310) and MyC-CaP (ATCC; #CRL-3255) cell lines are commercially available from the American Type Culture Collection (ATCC).

HEK-293T cell line was generously provided by the Preclinical Modelling of Paediatric Cancer Evolution team at the Institute of Cancer Research (ICR).

3.4.2 Cell culture

TRAMP-C1 were grown in Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher; #10313021) supplemented with 10% Fetal Bovine Serum (FBS; Sigma-Aldrich; #F9665), 4 mM L-glutamine (Thermo Fisher; #25030081), 1x Penicillin–Streptomycin (Thermo Fisher; #15140122), 0.005 mg/ml bovine insulin (Scientific Laboratory Supplies; #LZBE02-033E20) and 10 nM dehydroisoandrosterone (Sigma-Aldrich; #D4000) at 37 °C with 5% CO_2. HEK-293T, RM-1 and MyC-CaP were grown in DMEM supplemented with 10% FBS, 4 mM L-glutamine and 1x Penicillin–Streptomycin at 37 °C with 5% CO_2. All cell lines were tested negative for mycoplasma using the Mycoplasma PCR detection kit (Abcam; #ab289834). Cells were grown to 80-90% confluency before passaging every 3-5 days, depending on experimental need.

3.4.2.1 Seeding chamber slides

Chamber slides (Ibidi; #80841) were coated with 50 µg/mL poly-D-lysine (Sigma-Aldrich; #A003E) in PBS for 1 h to overnight at 37 °C. Slides were washed three times with RNase-free water, then allowed to dry fully, uncovered, in a biological safety cabinet for several hours. Cells were lifted from culture vessels, counted using the Rebel system (Echo; #REB-01-D), and seeded at a suitable density in the appropriate growth medium to achieve the desired confluence within 1–3 days.

3.5 Molecular and genetic methods

3.5.1 Plasmid transformation and amplification

3.5.1.1 Chemical transformation

Chemically competent cells were thawed on ice and transformed with either 5 µL of the assembly product or a ligated plasmid without the insert (negative control). After mixing, samples were incubated on ice for 30 min, heat-shocked at 42 °C for 30 sec, chilled for 2 min, and then recovered in 950 µL of SOC Outgrowth Medium (New England Biolabs; #B9020S) at 37 °C for 1 h with shaking. Cultures were spread on agar plates supplemented with ampicillin (100 µg/mL) and incubated overnight at 37°C. Individual colonies from the assembly plate were inoculated into ampicillin-supplemented Lysogeny broth (LB) and grown for 24 hours at 37 °C with shaking. Plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen; #27106) following the manufacturer's instructions. DNA concentration was measured, and clones were validated by Sanger sequencing.

3.5.1.2 Electroporation

Electroporation was performed using the GENE-Pulser electroporator (Bio-Rad #165278) with the Gene Pulser Controller (Bio-Rad #165-2098) at 2.5 kV, 25 µF, 200 Ω. SURE Electroporation-Competent Cells (Agilent; #200227) were thawed and kept on ice, with 2 mm cuvettes. For X-CODE library generation, the first pooled ligation used 50 µL of cells and 5 µL of plasmid; the second and third pooled ligations used 100 µL of cells and 10 µL of plasmid. A negative control used 50 µL of cells and 5 µL of backbone (lacking an insert).

Following electroporation, cells were resuspended in SOC medium (950 µL or 1900 µL, depending on original cell volume) and incubated at 37 °C at 250 rpm for 30 min. To calculate efficiency, the transformed cells were plated at different dilutions on agar plates containing 100 µg/mL of ampicillin, with controls plated at the highest dilution and grown overnight at 37 °C, inverted. The remaining cells were grown in 450 mL of LB-broth supplemented with 100 µg/mL of ampicillin and incubated overnight at 37 °C, shaking. Plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen;

#27106) following the manufacturer's instructions. DNA concentration was measured, and clones were validated by Sanger sequencing.

3.5.2 Lentiviral production and transduction

HEK293T cells were seeded at 20% density into 10 cm plates. The following day, cells were transfected with the expression plasmid and the VSVg, GAG, and REV packaging plasmids using Lipofectamine 3000 (Thermo Fisher; #L3000015) according to the manufacturer's instructions. The medium was replaced after 16 h, and the culture was collected after 48 h, filtered through a 0.45 µm filter, and concentrated by ultracentrifugation at 24,000 rpm for 2.5 h at 4 °C. The resulting pellet was resuspended in Hanks' Balanced Salt Solution (HBSS) (Thermo Fisher; #14170112) and stored at –80 °C.

For infection, 20,000 cells from the chosen cell line were seeded in 10 cm culture dishes and exposed to varying concentrations of lentiviral particles in the presence of 5 µg/mL polybrene (Thermo Fisher; #12965169). After 48 hours, cells were selected with 1 µg/mL puromycin (Thermo Fisher; #A1113803). Once the selection was complete, the MOI of the infections was calculated from the counts of surviving cells across varying lentivirus concentrations. A multiplicity of infection (MOI) of <0.2 to <0.3 was maintained.

3.5.3 Next-generation sequencing

Genomic DNA was extracted from barcoded cells or frozen tumour tissues using the Qiagen DNeasy Blood & Tissue Kit (Qiagen; #69504), following the manufacturer's instructions. For complete library characterisation, the X-CODE plasmid pool was used directly as the PCR template. Barcode sequences were amplified by PCR and subsequently gel-purified. The purified amplicons were submitted for amplicon sequencing using the Amplicon-EZ service (Azenta), according to the provider's recommended protocol.

3.6 Generation of simple A, B, and C barcodes

X-CODE simple-barcoded lines were generated with assistance from Matteo Tartiagni.

Primers for Gibson assembly were designed using the NEBuilder Assembly Tool [162]. Four specific sequences were combined into three unique pairs to generate barcodes A, B, and C (**Supplementary Table 1**). These barcode sequences were generated by PCR amplification from pre-existing plasmid templates using Q5 High-Fidelity DNA Polymerase (New England Biolabs; #E0555L). PCR products were gel-purified using the QIAquick Gel Extraction Kit (Qiagen; #28706). The CROPseq-Guide-puro plasmid (Addgene; #86708) was digested with MluI-HF (New England Biolabs; #R3198S) and gel-purified to the manufacturer's instructions. Gibson assembly was performed using the NEBuilder HiFi DNA Assembly Kit (New England Biolabs; #E5520S) with a 1:5 vector-to-insert ratio, followed by incubation at 50 °C for 15 min. The resulting plasmid was used to transform chemically competent cells as described in Section 3.5.1.1, and used for lentivirus production and transduction of TRAMP-C1 cells, as described in Section 3.5.2.

3.7 Design of the X-CODE construct

The design of the X-CODE libraries was carried out by members of the Tumour Functional Heterogeneity Lab at the Institute of Cancer Research, including Roberto Rota, Dr Marco Bezzi, and Matteo Tartagni.

Below is a methodology-focused overview of the X-CODE barcode's design, construction, and library generation. A complete explanation of the concept is provided in Section 4.3.

3.7.1 Design of X-CODE barcode units

The 27 X-CODE variable binding sequences (VBS) were derived from the insert probes of the proximity ligation assay for RNA (PLAYR), which is used to measure endogenous transcripts via a probe-based detection strategy [137,141]. Of the total 27 VBS, 23 were directly adapted from the PLAYR system, whereas 4 were generated in-house following PLAYR design principles (GC content, melting temperature, and length). The universal binding sequence (UBS) was designed by combining the two backbone-binding regions from the probes used in the PLAYR system. Combining a VBS with a UBS generates a 41-base-pair (bp) barcode unit (BU). The VBS (20 bp) and UBS (20 bp) are separated

by a single guanine nucleotide (G), providing spacing that enables the concurrent binding of the hybridisation probes: the padlock and universal probe (**Figure 3-1**).

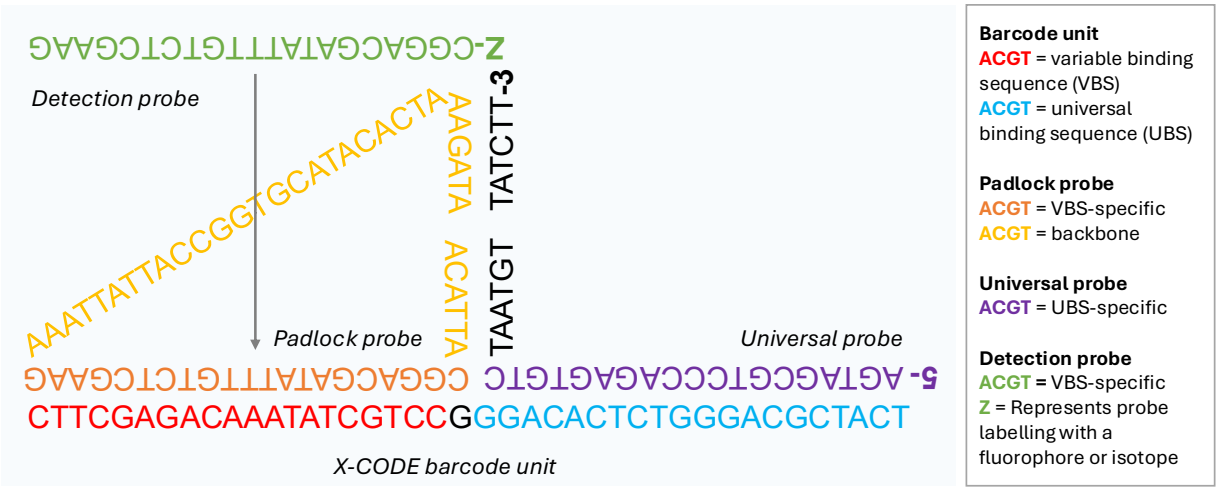


Figure 3-1. Sequence level design of an X-CODE barcode unit with padlock probes.

Representative sequence level schematic of X-CODE barcode unit (BU) 19. The variable binding sequence (VBS) can be one of 27 sequences and is placed upstream of the shared universal binding sequence (UBS). The BU is 40 base pairs (bp) in total, comprising the 20-bp VBS and UBS, with a single guanine nucleotide spacer between them. This creates space for the simultaneous binding of the hybridisation probes in the SNAIL configuration, comprising the padlock (shell) and universal (tail) probes.

3.7.2 Design of barcode unit sets 1–3

To generate the combinatorial complexity required for the X-CODE 15K and 2M libraries, barcode units (BU) were grouped into three groups of 9. Group 1 containing BU1 – 9, group 2 containing BU10 – 18, and group 3 containing BU19 – 27. Four BUs from each group were combined (9 choose 4) to combinatorially generate 126 unique BU-combinations (BU sets) per group. A linker region (ACACACAC) was inserted between the BU sequences within each set to facilitate assembly of the SNAIL complex. Within each set, a unique set identifier (USI) was added, forming the secondary short barcode. Each USI was generated by a 4 choose 4 combinatorial design, excluding homo-tetrameric and homo-trimeric motifs. The USI was positioned downstream of the BU combinations within the BU set, separated by a BamHI–linker–EcoRI cassette.

To facilitate integration of the BU-set oligonucleotides into the vector backbone, distinct restriction sites were added to the appropriate BU set. BU set-1 oligonucleotides carried MluI (5′) and SalI (3′) sites, with a BbsI–spacer–cut site motif before SalI for scaffold gRNA cloning downstream of USI-1. An additional BamHI–linker–EcoRI cassette between BU set-1 and USI-1 enabled the three-step clonal strategy to assemble the

final barcode. Both BU set-2 and -3 oligonucleotides share the same design, with an MfeI site at the 5′ end and a BglII site at the 3′ end. Upon digestion, the MfeI and EcoRI sites generate compatible cohesive ends, as do the BglII and BamHI sites.

3.8 Generation of the 15K and 2M X-CODE libraries

The generation of the X-CODE libraries was kindly undertaken by Roberto Rota of the Tumour Functional Heterogeneity Lab at the Institute of Cancer Research

3.8.1 Assembly of the X-CODE library backbone (CROPseq-derived)

The primers for Gibson assembly were designed using the NEBuilder Assembly Tool [162]. The Filler_Rev and LTR_Frwd primers (**Supplementary Table 4**) were designed to harbour two restriction sites for SalI and XbaI, separated by a linker region within the spacer regions. 1 µg of CROPseq-Guide-puro (Addgene; #86708) was digested with SnaBI (New England Biolabs; #R0130S) and SalI-HF (New England Biolabs; #R3138S) and gel-purified. The filler and the LTR fragments were generated through polymerase chain reaction (PCR) (**Supplementary Table 5**) from the CROPseq-Guide-puro plasmid (Addgene; #86708) and gel-purified. The Gibson assembly was performed using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs; #E5520S) in a 1:5 ratio according to the manufacturer's instructions. To increase reaction efficiency, the incubation at 50 °C was extended from 15 to 60 min. The ligation product (CROPseq-puro-filler-LTR-MluI-HF/SalI-HF) was used to transform chemically competent cells, as described in Section 3.5.1.

3.8.2 Assembly of X-CODE libraries

A pooled multi-step ligation-based strategy was employed for X-CODE library assembly based on the CombiGEM methodology [163]. X-CODE combinatorial set-1 oligonucleotide mix was digested with SalI-HF and MluI, while set-2 and set-3 were digested with MfeI (New England Biolabs; # R3589S) and BglII (New England Biolabs; #10159523). In the first ligation, gel-purified set1-MluI-SalI oligos were assembled into the CROPseq-puro-filler-LTR-MluI-HF/SalI-HF backbone (CROPX1). The product was transformed into competent cells using electroporation, amplified, and purified as described in Section 3.5.1.2. The purified CROP-set1 was digested with BamHI (New

England Biolabs; #R0136S) and EcoRI (New England Biolabs; #R3101S), gel-purified, and dephosphorylated with CIP (New England Biolabs; #M0525S) according to the manufacturer's instructions, yielding a digestion product termed CX1BamHI/EcoRI/CIP.

3.8.2.1 X-CODE 15K library generation

The second ligation assembly was prepared by combining CX1BamHI/EcoRI/CIP and gel-purified set2-MfeI-BglII (CROPX1X2). The ligation product was electroporated into competent cells, amplified and purified as described in Section 3.5.1.2. For add-ons integration, the scaffold-gRNA segment was amplified from CROPseq-puro-v2 (Addgene; #127458) using primers (**Supplementary Table 5**) containing BbsI (forward) and XbaI (reverse) sites with an ACAC linker, followed by PCR purification, digestion, and gel purification.

The digested amplicons were ligated into the BbsI/XbaI–digested CROPX1X2 backbone overnight at 16 °C, purified, and electroporated into competent cells, as described in Section 3.5.1.2, to generate CROPX1X2.1. Subsequently, the WPRE-SplitLTR-TSO-U6 fragment was PCR-amplified with primers carrying BamHI/EcoRI sites and ACAC linkers from the ClonMapper vector (Addgene; #137993), digested, and ligated into BamHI/EcoRI–digested CROPX1X2.1. The final ligation product was purified and electroporated into competent cells, as described in Section 3.5.1.2, to generate the final CROPX1X2.2 15K library.

Cloning was validated by Sanger sequencing of 5 randomly selected colonies. 3 of the 5 colonies selected for validation represented a complete set of the 18 BUs covering the 15K library. Thus, they were transduced into TRAMP-C1 cells, as described in Section 3.5.2, to generate the F, I, and H 15K clonal lines. The final CROPX1X2.2 15K library was also transduced into TRAMP-C1, RM-1, and MyC-CaP cell lines, as described in Section 3.5.2, to generate 15K barcoded populations.

3.8.2.2 X-CODE 2M library generation

The third ligation assembly was prepared by combining CX1X2BamHI/EcoRI/CIP and X3-MfeI-BglII. The ligation product was electroporated into competent cells, amplified and purified. For add-ons integration, the same strategy used for the 15K library was implemented. Cloning was validated by Sanger sequencing of randomly selected colonies. 4 colonies selected for validation represented a complete set of the 27 BUs of

the 2M library. Thus, they were transduced into RM-1 cells, as described in Section 3.5.2, to generate the W, Z, Y, and J 2M clonal lines.

Sanger sequencing revealed a high level of recombination, which is incompatible with library-scale up. Accordingly, a different strategy was adopted for assembling the 2M library. The ClonMapper vector (Addgene; #137993) was mutated through mutational PCR and Gibson assembly to remove the BglII site from within the WPRE-SplitLTR-TSO-U6 sequence. The mutated WPRE-SplitLTR-TSO-U6 fragment was PCR-amplified with primers carrying BamHI/EcoRI sites and ACAC linkers, digested, and ligated into BamHI/EcoRI–digested X3 minigenes.

The X3-mutWPRE-SplitLTR-TSO-U6 was then extracted from the minigenes using MfeI and BglII. The third ligation assembly was prepared by combining CX1X2.1 BamHI/EcoRI/CIP and X3-mutWPRE-SplitLTR-TSO-U6 -MfeI-BglII. The ligation product was PCR-purified and used directly for lentiviral production. All intermediate cloning steps were validated by Sanger sequencing.

3.9 X-CODE clonal retrieval

The implementation and optimisation of the clonal retrieval strategy were carried out by Roberto Rota with assistance from Matteo Tartagni of the Tumour Functional Heterogeneity lab at the Institute of Cancer Research (ICR).

Clonal retrieval was performed following the previously established ClonMapper methodology [76]. Briefly, barcode arrays were generated by annealing three pairs of overlapping oligonucleotides, each encoding the desired barcode sequence and flanked by overhangs. Complementary oligonucleotides were annealed by heating to 80 °C and then cooling to room temperature, yielding three dsDNA blocks. These blocks were phosphorylated and ligated sequentially to assemble the final barcode array. The ligation product (~170 bp) was selected by electrophoresis on a 2% agarose gel and gel-purified. The purified array was inserted into the Recall-miniCMV-sfGFP plasmid backbone (Addgene; #137995), which had been previously digested with BbsI (New England Biolabs; #R3539L), using T4 ligase at a 10:1 insert-to-vector ratio at 37°C for 1 h.

The assembly was transformed into chemically competent cells. Correct insertion of the barcode array was confirmed through Sanger sequencing. For functional recall, cells were transfected with the assembled recall plasmid and dCas9-VPR plasmid (Addgene; #63798) using Lipofectamine 3000 following the manufacturer's instructions. After 16–18 h, the medium was replaced, and after 48h post-transfection, the cells were harvested, washed, and filtered through a 35 µm strainer. GFP^+-cells were gated as recalled barcoded lineages and sorted for downstream expansion or analysis.

3.10 Animal work

Animal work was carried out by Roberto Rota and Nicole Pandell of the Tumour Functional Heterogeneity lab at the Institute of Cancer Research (ICR).

All animal experiments were approved and monitored by the ICR Animal Welfare and Ethical Review Body (protocol PP9736772), in compliance with the UK Home Office Animals (Scientific Procedures) Act 1986, the United Kingdom National Cancer Research Institute guidelines for the welfare of animals in cancer research, and the ARRIVE guidelines.

Mice were maintained under standard conditions with ad libitum access to food and water. Tumour growth was measured twice weekly with callipers, and volume was calculated as (length × width²) / 2. Mice were monitored regularly and euthanised by CO_2 asphyxiation followed by cervical dislocation when tumour volume reached approximately 850 mm² or when any single tumour dimension exceeded 15 mm (endpoint).

3.10.1 TRAMP-C1 and RM-1 *in vivo* experiments

1 × 10^6 W, Z, Y, J (1:1:1:1) RM-1 or 1 × 10^6 RM-1-15K or TRAMP-C1-15K were injected to generate tumours used for spatial analyses and MALDI. Cells were detached and filtered to obtain a single-cell suspension, mixed in 200 µL 1:1 PBS and Matrigel (Corning; #356231), and injected subcutaneously into a single depilated and sterilised flank of C57BL/6J recipient mice (RRID: IMSR_JAX:000664). Mice were monitored and culled in accordance with the limits described above.

3.10.2 MyC-CaP *in vivo* experiments

30 FVB-NJ mice (RRID: IMSR_JAX:001800), 6-8 weeks old, were subcutaneously injected with 1 × 10^6 barcoded MyC-CaP cells in 200 µL of 1:1 PBS and Matrigel solution. All experimental groups, excluding the uncastrated and pre-castration control groups, received chemical castration via Degarelix (0.315 mg, 100 µL, every 14 days) after 21 days or when tumours exceeded 400 mm^2.

Control groups included eight uncastrated (intact) mice sacrificed at the endpoint (EP-Ctrl) and five mice sacrificed prior to castration (Pre-ADT). Post-castration groups were collected at early (4 days: ADT-D4, n = 5), late (10 days: ADT-D10, n = 6), and endpoint (EP-ADT, n = 6) time points. Mice were monitored and culled in accordance with the limits described above.

3.10.3 *In vivo* sample collection and processing

3.10.3.1 Fresh-frozen samples

Tumours destined for cryosectioning were snap-frozen immediately after resection. Tumours were loosely wrapped in aluminium foil and then floated on the surface of dry-ice cooled isopentane for ~1 min per 1 cm^2 of tissue. Frozen samples were transferred to sealed containers, such as microcentrifuge tubes, and stored at −80 °C long-term (several months to a year).

Samples collected from the MyC-CaP experiment were sectioned into four parts: two central sections were snap-frozen as described above, and the two outer sections were allocated either for freezing for DNA extraction (as detailed in 3.5.3) or for mechanical dissociation to generate primary cell cultures.

3.10.3.2 Formalin-fixed samples

Tumours designated for paraffin-embedding were submerged in 4% PFA (in 1X PBS) and fixed overnight at 4 °C with gentle rolling (<24 h), after which samples were transferred to 1× PBS and stored at 4 °C until further processing.

3.11 Tissue block generation

3.11.1 OCT embedding

Tissue samples designated for cryosectioning, supported by OCT compound, were embedded using a rapid-freezing protocol. Prior to sample embedding, OCT compound (optimal cutting temperature; Sakura Tissue-Tek, #4583) was placed on ice, and forceps, cryomolds, and isopentane were prechilled for at least 30 minutes, with isopentane cooled on dry ice. Fresh-frozen tissue samples were retrieved from −80 °C storage on dry ice, and cryomolds were labelled to ensure unambiguous sample identification. Cryomolds were filled with prechilled OCT to prevent air bubbles and then placed on ice.

Using dry-ice prechilled forceps, tissue samples were placed into the OCT-filled moulds and oriented as required prior to freezing, with additional OCT added to fully cover the tissue if necessary. The cryomolds were immediately transferred to the surface of chilled isopentane to induce rapid freezing. Once the OCT had turned opaque (typically within 1–5 minutes), blocks were placed on dry ice for a further 15 minutes to ensure complete solidification and to allow evaporation of residual isopentane. Embedded OCT blocks were then stored in airtight containers at −80 °C until sectioning or long-term storage.

3.11.2 HPMC/PVP embedding

HPMC/PVP embedding medium was prepared and used for single and multi-sample tissue block generation as previously described [164,165].

Briefly, hydroxypropyl methylcellulose (HPMC; 15 g) and polyvinylpyrrolidone (PVP; 5 g) were combined in a glass laboratory bottle and gradually dissolved in 180 mL deionised water with intermittent mixing to ensure homogeneous dispersion and prevent aggregation. The embedding medium was stored at 4 °C until complete dissolution of the polymers, typically overnight, and was used within two weeks of preparation.

Prior to embedding, the medium was gently mixed and equilibrated on ice. Isopropanol and isopentane were chilled separately using dry ice to generate cooling slurries.

Embedding moulds were filled with HPMC/PVP medium and placed on wet ice to allow air bubbles to dissipate. Tissue samples were retrieved from −80 °C storage and maintained on dry ice prior to embedding. For multi-sample blocks, specimens were arranged and oriented on aluminium foil on dry ice. Then, samples were rapidly positioned into the prefilled moulds in a defined layout, ensuring alignment against the base of the mould to maintain a consistent sectioning plane.

Once fully covered with embedding medium, moulds were lowered into an isopropanol–dry ice slurry until the block was completely frozen; thereafter, the block was briefly immersed (~30 sec) in an isopentane–dry ice slurry to improve freezing homogeneity. Frozen blocks were placed on dry ice to allow evaporation of residual isopentane and were either stored at −80 °C for long-term preservation or until used for cryosectioning.

3.11.3 Formalin-fixed paraffin embedding

Formalin-fixed paraffin-embedded (FFPE) tissue blocks were generated by Oscar Williamson of the Pathology Facility at the Institute of Cancer Research according to standard histological processing protocols. Briefly, tumour samples previously fixed in 4% PFA, then stored in 1X PBS, were dehydrated through a graded ethanol series, cleared in xylene, and infiltrated with molten paraffin wax for embedding. Embedded blocks were allowed to solidify and were stored at 4 °C rather than at room temperature to better preserve RNA integrity until sectioning.

3.12 Tissue block sectioning

3.12.1 OCT and HPMC/PVP block sectioning

OCT- or HPMC/PVP-embedded tumour blocks were cryosectioned using a ThermoFisher HM525NX cryostat with the chamber precooled to −20 °C. For RNA-sensitive samples, the chamber and working surfaces were cleaned with RNase decontamination solution prior to sectioning. Tissue blocks were transferred from −80 °C storage on dry ice and allowed to equilibrate in the cryostat chamber for at least 20 minutes to minimise sectioning artefacts.

Blocks were removed from their moulds and mounted onto the specimen chuck using OCT compound or RNase-free water for OCT and HPMC/PVP blocks, respectively. The specimen was aligned parallel to the blade, and excess embedding medium was trimmed using an initial cutting thickness of 20 µm to expose the tissue surface. Final sections were cut at 10 µm, and initial sections were discarded until artefact-free sections were obtained. Sections were flattened, mounted on SuperFrost Plus, Adhesion Slides (VWR; #631-0448), briefly melted onto the slides, and immediately transferred to a cryobar. Sectioning was repeated as needed to obtain multiple sections per slide.

Completed slides were placed in prechilled slide boxes and stored on dry ice prior to transfer to long-term storage at −80 °C in sealed secondary containers. In addition, HPMC/PVP sections intended for MALDI mass spectrometry imaging were vacuum sealed prior to long-term storage at −80 °C to preserve metabolite integrity.

3.12.2 FFPE block sectioning

FFPE tissue blocks were sectioned by Oscar Williamson of the Pathology Facility at the Institute of Cancer Research according to standard histopathological procedures. Briefly, paraffin blocks were sectioned at 4 µm using a rotary microtome and mounted onto glass slides. Sections were dried prior to storage and were stored at 4 °C rather than at room temperature to better preserve RNA integrity until use.

3.13 Haematoxylin and eosin staining

Following multiplexed imaging on the PhenoCycler-Fusion platform, haematoxylin and eosin (H&E) staining was performed on the same tissue sections to assess tissue morphology. H&E staining was carried out according to a protocol optimised for post-PhenoCycler imaging workflows [166] . Briefly, the PhenoCycler flow cell was removed by incubation in xylene for 24 h, after which tissue sections were subjected to standard haematoxylin and eosin staining (Abcam; #ab245880). Stained sections were mounted using a xylene-based mounting medium, cover-slipped, and allowed to cure prior to imaging. Mounted slides were stored at room temperature and imaged using standard brightfield microscopy.

3.14 Core X-CODE assay buffers

The following tables detail the core X-CODE assay buffers shared across suspension and spatial detection protocols, regardless of read-out methodology.

Table 3-4. Core X-CODE assay buffer components.

All buffers, and their components if required, were diluted in RNase-free water.

Component	Final concentration	Supplier and product number
PBSTR+ wash buffer		
PBS	1×	Thermo Fisher; #AM9625
Tween-20	0.1% (v/v)	Sigma-Aldrich; #P2287
RNasin Plus	40 U/mL	Promega; #N2615
PBSTR wash buffer		
PBS	1×	Thermo Fisher; #AM9625
Tween-20	0.1% (v/v)	Sigma-Aldrich; #P2287
RNasin Plus	4 U/mL	Promega; #N2615
PBST wash buffer		
PBS	1×	Thermo Fisher; #AM9625
Tween-20	0.1% (v/v)	Sigma-Aldrich; #P2287
Hybridisation buffer		
Saline sodium citrate (SSC)	1×	Thermo Fisher; #10418353
Tween-20	1% (v/v)	Sigma-Aldrich; #P2287
Polyvinyl sulfonic acid	2.5%	Sigma-Aldrich; #278424-250mL
Salmon sperm DNA	100 µg/mL	Life Technologies; #15632011
Ribonucleoside vanadyl complex	20 mM	New England Biolabs; #S1402S
RNasin Plus	40 U/mL	Promega; #N2615
X-CODE shell probes	100 nM per probe	Integrated DNA Technologies
X-CODE universal tail probe	1 µM	Integrated DNA Technologies
Padlock probes (e.g. B-actin)	100 nM each per probe	Integrated DNA Technologies
Post-hybridisation buffer		
PBS	1× PBS	Thermo Fisher; #AM9625
SSC	4× SSC	Thermo Fisher; #10418353
Tween-20	0.1% (v/v)	Sigma-Aldrich; #P2287
RNasin Plus	40 U/mL	Promega; #N2615
Ligation buffer		
T4 DNA ligase buffer	1×	Thermo Fisher; #10723941
T4 DNA ligase	40 U/mL	Thermo Fisher; #10723941
RNasin Plus	5 U/mL	Promega; #N2615
Rolling circle amplification (RCA) buffer		
EquiPhi29 DNA polymerase	5 U/mL	Thermo Fisher; #15973292
EquiPhi29 reaction buffer	1×	Thermo Fisher; #B39
Dithiothreitol	1 mM	Thermo Fisher; #R0861
dNTP mix	250 µM	Thermo Fisher; #R0181

3.15 X-CODE detection in suspension

3.15.1 Isotope-oligonucleotide conjugation

Metal conjugation to detection oligonucleotides was performed following the PLAYR protocol [137]. Detection probe oligonucleotides carrying a 5′ thiol modifier (**Supplementary Table 2**, **Supplementary Table 8**, and **Supplementary Table 9**) were metal-labelled using a polymer-based chelation strategy. Briefly, oligonucleotides were reduced in Bond-Breaker TCEP solution (neutral pH; Thermo Fisher; #77720) to expose free thiol groups, followed by ethanol precipitation and resuspension in L-buffer supplied in the Maxpar X8 Antibody Labelling Kit (Standard Biotools; #201300).

In parallel, lanthanide metals were chelated to polymer reagents supplied in the Maxpar X8 Antibody Labelling Kit and incubated to generate metal-loaded polymers. Reduced oligonucleotides were subsequently conjugated to the metal-loaded polymers under controlled conditions and purified by centrifugal filtration to remove unbound metal and reaction components. Labelled detection probes were quantified by UV absorbance at 260 nm, diluted to a final working concentration of 1 µM, aliquoted, and stored at −20 °C until use.

3.15.2 X-CODE CyTOF assay

The core X-CODE buffers used for X-CODE detection in suspension are detailed in **Table 3-4**.

3.15.2.1 Cell preparation

RNA detection by CyTOF was performed following an adapted version of the PLAYR protocol [137]. 1.0–1.5 × 10^6 cells were counted and collected per sample; a range optimised for TRAMP-C1 and RM-1 cells. Cells were pelleted (before fixation: 500 g, 3 min, 4 °C), washed once in 1X PBS, and resuspended in 200 µL of 1 µM Cisplatin-194Pt (Standard BioTools; #201194) in PBS for 5 min at room temperature. Cells were either processed for optional extracellular antibody staining (as described in Section 3.15.2.2) or were washed with 1X PBS (in RNase-free water).

Cells were resuspended in ice-cold 1.6% PFA in RNase-free PBS (~1 mL per 1 × 10^6 cells) for 10 min on ice. After centrifugation, the pellet was loosened, and ice-cold methanol (100 µL per 1 × 10^6 cells) was added dropwise while vortexing to prevent clumping. Cells were incubated on ice for 10 min, then stained as described below. Alternatively, cells were stored in methanol at –80 °C for several months without significant RNA degradation.

3.15.2.2 Optional extracellular antibody staining

Cells were washed once in ice-cold Cell Staining Buffer (CSB; Standard BioTools; #201068), resuspended in 100 µL of antibody (**Table 3-1**) cocktail in CSB and incubated on ice for 30 min with frequent agitation. Cells were washed with CSB, then with 1X PBS (in RNase-free water), and resuspended in 100 µL of freshly prepared 5 mM BS3 (Thermo Fisher; #10249963) in PBS and incubated on ice with agitation for 30 min.

Cells were resuspended in ice-cold 1.6% PFA in RNase-free PBS (~1 mL per 1 × 10^6 cells) for 10 min on ice. After centrifugation, the pellet was loosened, and ice-cold methanol (100 µL per 1 × 10^6 cells) was added dropwise while vortexing to prevent clumping. Cells were incubated on ice for 10 min, then stained as described below. Alternatively, cells were stored in methanol at –80 °C for several months without significant RNA degradation.

3.15.2.3 Hybridisation, ligation, and rolling circle amplification

Upon removal from –80 °C storage, cells were thawed for 5 min on ice then pelleted (after permeabilisation: 800g, 3 min, 4 °C), washed once with 500 µL PBSTR+ wash buffer and immediately resuspended in 500 µL of Hybridisation buffer containing X-CODE padlock probes and the universal probes and β-actin probes (**Supplementary Table 2**, **Supplementary Table 3**, and **Supplementary Table 6**). The padlock probe mix was incubated at 90 °C for 5 min, then cooled on ice before addition to the final Hybridisation buffer. Cells in the Hybridisation buffer were incubated at 40 °C for 2 h in a thermal mixer with agitation at 1,400 rpm. Cells were then washed twice with 500 µL ice-cold PBSTR wash buffer and incubated in 500 µL of Post-hybridisation buffer at 40 °C for 20 min.

Cells were washed twice with PBSTR wash buffer, resuspended in 500 µL of Ligase buffer, and incubated at 21 °C for 1 h at 1,100 rpm. Cells were washed twice with

PBSTR, then resuspended in 500 µL of RCA buffer and incubated at 37 °C for 5 h to overnight with agitation at 1,100 rpm. Cells were washed twice with CSB and incubated in a CSB solution containing isotope-labelled detection probes (**Table 3-1**) for 1 h at 37 °C with agitation at 1,100 rpm. Cells were either processed for optional extracellular antibody staining (as described in Section 3.15.2.4) or washed twice with CSB. Then incubated in 500 µL of 62.5 nM Ir 191-193 (Standard BioTools; #201192B) in 1.6% PFA, diluted in PBS, for 1 h at room temperature or for up to 3 days at 4 °C, with minimal signal loss.

3.15.2.4 Optional intracellular antibody staining

Cells were washed twice in CSB, then resuspended in 100 µL of intracellular antibody (**Table 3-1**) cocktail in CSB on ice for 1 h with frequent agitation. Cells were washed twice with CSB and incubated in 500 µL of 62.5 nM Ir 191-193 (Standard BioTools; #201192B) in 1.6% PFA, diluted in PBS, for 1 h at room temperature or for up to 3 days at 4 °C, with minimal signal loss.

3.15.3 CyTOF sample acquisition

For CyTOF sample acquisition, the Helios mass cytometer (Standard Biotools) was calibrated according to the manufacturer's protocols. Cells were then washed in 2 mM EDTA (Thermo Fisher; #15575020) in CSB, then CSB, followed by Cell Acquisition Solution (CAS; Standard Biotools; #201240). Immediately before acquisition, cells were diluted to 0.5-1 × 10^6 cells/mL in a 10% EQ Four Element Calibration Beads (Standard BioTools; #201078) solution with 2 mM EDTA in CAS. Cells were filtered through a 35 µL filter and acquired at 100-400 events per second.

3.16 Spatial X-CODE detection

The X-CODE spatial assay can be conducted on cells plated or STAMPed on microscope or chamber slides, on fresh-frozen tissue sections embedded in optimal cutting temperature (OCT) or hydroxypropyl methylcellulose/polyvinylpyrrolidone (HPMC/PVP) hydrogel, and on formalin-fixed paraffin-embedded (FFPE) samples. The following details outline the preparation required for each sample type to undergo the X-CODE spatial assay; these steps

remain consistent once the samples are prepared. After this, the assay is tailored to the specific readout technology.

3.16.1 Cell or tissue sample preparations

3.16.1.1 Chamber slide preparation

The growth medium was aspirated from the chamber slide wells, and the cells were washed twice with PBS, then fixed with 4% PFA in PBS for 10 min at room temperature. The fixative was removed, and cells were permeabilised with ice-cold 100% methanol, dropwise, and incubated for 15 min at –80 °C before immediately commencing the X-CODE spatial assay. Alternatively, cells on chamber slides can be stored at –80 °C in sealed containers for several weeks to months without significant RNA degradation. To perform the X-CODE spatial assay after long-term storage, chamber slides were thawed on ice for ~5 min, then the methanol was removed.

3.16.1.2 Cell preparation for STAMPing

Cells were attached to microscope slides post-fixation by using a modified STAMPing protocol [167], applying minor methodology adaptations to complete workflows with readily available laboratory equipment.

Chamber slides (Ibidi; #80841) were coated with 50 µg/mL poly-D-lysine (Sigma-Aldrich; #A003E) in PBS for 1 h to overnight at 37 °C. Slides were washed three times with RNase-free water, then allowed to dry fully, uncovered, in a biological safety cabinet for several hours. Cells were lifted from culture vessels, counted using the Rebel system (Echo; #REB-01-D), fixed with 4% PFA in PBS (1 mL per 1 × 10^6 cells) for 30 min at room temperature, then washed twice with RNase-free water (800 *g*, 3 min, 4 °C), and resuspended in RNase-free water containing 0.01% Triton-X-100 (Sigma-Aldrich; #T8787).

The cell number and resuspension volume were estimated using ChatGPT 4–5 to ensure adequate coverage of the well area without overcrowding. The cell suspension was applied evenly to the slide, and slides were dried sequentially at 4 °C (30 min), room temperature (5 min), and on a 42 °C hot plate (1.5 h). Finally, the adhered cells were permeabilised dropwise with 100% methanol at –80 °C overnight within the cassettes.

3.16.1.3 OCT and HPMC/PVP section preparation

Sectioned slides were removed from –80 °C storage, then removed from their vacuum packaging and thawed at 40 °C for 1 min in a Hybridisation oven (Advanced Cell Diagnostics HybEZ Oven; #241000ACD-2). Slides were then immediately fixed in 4% PFA for 30 min at room temperature. Slides were washed in PBS (2 min), then submerged in 1% SDS (Promega; #V6551) in RNase-free water for 2 min, followed by two PBS washes (2 min). Slides were incubated in pre-chilled 70% methanol in RNase-free water on ice for 1 h, then washed twice in PBS (2 min) before immediately commencing the X-CODE spatial assay.

3.16.1.4 FFPE section preparation.

Slides were baked at 60 °C for 1 h on a slide baker, deparaffinized using xylene (Thermo Fisher; #10467270) and rehydrated through an ethanol series (100%, 90%, 70%) in RNase-free water. To permeabilize the tissue, a freshly prepared pepsin solution (Sigma-Aldrich; #P7012) at 0.5 mg/mL in cold 0.1 M HCl (VWR; #31955.293P) was applied to cover each section entirely (~200 µL per 1 cm^2 section). Slides were incubated at room temperature for 30 min. Then, slides were washed twice in PBS (2 min) and air-dried for 5 min before commencing X-CODE spatial assay.

3.16.2 X-CODE spatial assay

The core X-CODE buffers used for spatial X-CODE spatial detection are detailed in **Table 3-4**.

The volume of buffer used depends on the size of the barrier drawn or chamber slide; for example, a 2 cm^2 barrier holds approximately 200 uL of buffer.

3.16.2.1 Barrier placement and cell washing

For tissue sections. OCT, HPMC/PVP, or FFPE-prepared slides were wiped using a lint-free tissue to remove excess moisture. Then, a hydrophobic barrier was drawn around the tissue sections using a PAP pen (Vector Laboratories; #H-4000), and the pen was allowed to dry completely (~5 min). Then, PBSTR+ was immediately added to the tissue within the PAP pen barrier.

For plated or STAMPed cells. PBSTR+ was added gently to the chamber wells to avoid disturbing the cells.

3.16.2.2 Hybridisation, ligation, and rolling circle amplification

Following a 2 min incubation, the PBSTR+ was carefully removed from inside the barrier or wells, then Hybridisation buffer was added. Slides were incubated for 3 h at 40 °C in a humidified Hybridisation oven (Advanced Cell Diagnostics HybEZ Oven; #241000ACD-2). Following hybridisation, slides were washed twice with PBSTR for 2 min. They were then incubated in Post-Hybridisation buffer for 20 min at 40 °C, followed by two washes in PBSTR (2 min).

Then, slides were incubated for 2 h at room temperature in Ligase buffer, washed twice with PBSTR for 2 min, and RCA buffer was added. Slides were incubated overnight at 30 °C in a humidified oven. Following rolling-circle amplification, the protocol stages vary depending on the detection technology used.

3.16.3 X-CODE spatial detection

3.16.3.1 Detection by fluorescence microscopy

Slides were washed twice with PBST for 2 min, then incubated with 200 nM X-CODE fluorescent detection probes (**Supplementary Table 10**) in PBST for 2 h at 37 °C in a humidified oven. Additionally, antibodies could be incubated with the fluorescent detection probes simultaneously (**Table 3-2**). Following two PBST washes, chamber wells were removed, if applicable, and slides were mounted in mounting medium containing DAPI (Thermo Fisher; #00-4959-52), sealed with transparent nail polish, and dried in the dark. Images were acquired by fluorescence microscopy using confocal instruments (Zeiss LSM700 or Zeiss LSM980) or the Zeiss Axio Scan Z1 slide scanner.

The Zeiss LSM700 and Zeiss Axio Scan Z1 enabled simultaneous visualisation of four fluorophores, and the Zeiss LSM980 enabled simultaneous visualisation of five markers (including DAPI nuclear staining). Slide scanning set-up was performed according to the manufacturer's procedures for the Zeiss Axio Scan Z1. For confocal microscopy, areas of interest within the sample were identified using the Zeiss LSM700 eyepiece for DAPI staining or the Zeiss LSM980 acquisition interface. At 20X magnification, the Zeiss ZEN microscopy software was used to optimise laser power and detector gain for each

channel to maximise signal while avoiding saturation. These presets were saved for subsequent confocal microscopy to ensure consistency across imaging rounds, depending on magnification and tissue-embedding type.

3.16.3.2 Detection by Akoya PhenoCycler-Fusion

Plated cells and OCT- and HPMC/PVP hydrogel-embedded tissue sections were funnelled into Akoya's antibody staining workflow for fresh-frozen samples, with minor adaptations to the manufacturer's instructions, using Akoya Staining (#7000008) and Sample Kits (#7000017) for PhenoCycler-Fusion. Briefly, the slides were washed twice in PBS for 2 min, chamber wells were removed if applicable, and the slides were incubated in Staining Buffer for 20–30 min at room temperature. Next, slides were incubated with an antibody cocktail (**Table 3-3**) where each antibody is conjugated to a unique Akoya PhenoCycler Barcode, for 3 hours at room temperature in a humidity chamber.

Post-staining, slides were washed with Staining Buffer twice for 2 min each, then fixed in 1.6% PFA for 10 min at room temperature. Slides were washed three times with PBS before a 5-min incubation in 100% methanol at 4 °C. Slides were washed three times in PBS, then incubated with Fixative Reagent in PBS for 20 minutes at room temperature in a humidity chamber. Following three additional PBS washes, the slide underwent PhenoCycler-Fusion flow cell assembly and was loaded into the instrument according to the manufacturer's instructions. If applicable, PAP pen barriers were promptly removed prior to flow cell assembly using a xylene-soaked cotton swab in a fume cupboard, without touching the tissue section, and the slides were then washed a further three times in PBS.

The PhenoCycler-Fusion was calibrated, and the Reporter Plate was prepared and loaded in accordance with the manufacturer's instructions. The Reporter Plate included the appropriate Akoya PhenoCycler fluorescent Reporters for visualisation of antibody staining (**Table 3-3**). In addition, 5 µL of 10 µM X-CODE fluorescent detection probes (**Supplementary Table 10**) were incorporated into the Reporter Plate configuration, analogous to the inclusion of Akoya Reporters. X-CODE detection probes and Akoya Reporters can be used within the same imaging cycle, provided that three distinct fluorophores are applied per cycle, consistent with the instrument's capacity. Specific

cycle combinations used for detecting the 15K and 2M X-CODE libraries are detailed in **Supplementary Table 11** and **Supplementary Table 12**.

3.16.3.3 Detection by Imaging Mass Cytometry (IMC)

For IMC, samples were washed twice with Cell Staining Buffer (CSB; Standard BioTools; #201068) for 2 min. Isotope-conjugated X-CODE detection probes (**Supplementary Table 8** and **Supplementary Table 9**) were then added and incubated at 37 °C for 1 h. Samples were subsequently washed twice with CSB for 2 min.

For nuclear staining, samples were incubated with Cell-ID Intercalator-Ir (Standard Biotools; #201192A) at a 1:400 dilution in PBS (300–500 µl per ~20 mm² section) for 30 min at room temperature in a humidified chamber. Samples were rinsed twice with RNase-free water for 5 min, air-dried at room temperature, and, if applicable, the chamber wells were removed.

Slides were imaged using a Hyperion imaging mass cytometer (Standard Biotools), calibrated according to the manufacturer's protocols. An overview microscopy image at low magnification was obtained using the Hyperion software to guide the selection of regions of interest. For each tissue sample, at least 3 regions of interest were acquired across the tissue.

3.17 MALDI-mass spectrometry imaging

10 µM sections of fresh-frozen HPMC/PVP-embedded X-CODE barcoded tissues, prepared according to 3.12.1, were adhered to SuperFrost Plus, Adhesion Slides (VWR; #631-0448). Slides were vacuum-packed to prevent metabolite degradation and shipped on dry ice to the Rosalind Franklin Institute (RFI) in Harwell, Didcot.

Sample preparation, including matrix application and subsequent matrix-assisted desorption/ionisation mass spectrometry imaging (MALDI-MS), was kindly undertaken by Anthony Devlin and Alexander Wilkes at the RFI.

3.17.1 DHB matrix application

A 40 mg/mL 2,5-dihydroxybenzoic acid (DHB, Tokyo Chemical Industry; #KN-490-79-9) in 70% (v/v) methanol (14262; Thermo Fisher, Hampton, NH) using liquid chromatography (LC) grade water. The matrix was applied using a TM HTX3 sprayer (nozzle temperature = 75°C, gas pressure = 10 psi, flow rate = 100 µL/min, velocity = 1200 mm/min, track spacing = 2.5 mm, number of passes = 8 in a criss-cross pattern with 2 s drying time).

3.17.2 DHAP matrix application

A 10 mg/mL 2,5-dihydroxyacetophenone (DHAP, Thermo Fisher; #A12185.06) matrix preparation was used for MALDI-MSI. The DHAP matrix also contained 69.95% (v/v) acetonitrile (Thermo Fisher; #A955500), and 29.95% 200 mM ammonium acetate (Merck; #631-61-9) in LC grade water (Thermo Fisher; #7732-18-5), and 0.1% formic acid (Sigma-Aldrich; #Z0797502 21). The matrix was applied using a TM HTX3 sprayer (nozzle temperature = 50 °C, gas pressure = 10 psi, flow rate = 120 µL/min, velocity = 1200 mm/min, track spacing = 2.5 mm, number of passes = 8 in a criss-cross pattern with 2 s drying time).

3.17.3 Matrix-assisted laser desorption/ionisation

The instrument was mass- and mobility-calibrated by direct infusion electrospray ionisation (ESI) of Agilent Tune Mix (Agilent Technologies; #G2431). MSI analysis was undertaken using a time-TOF fleX (Bruker Daltonics, Bremen, DE) instrument in positive ion mode without TIMS. The ion source parameters were set as follows for MALDI; MALDI offset = 50 V and shots = 150, rate = 10 kHz. For direct infusion ESI, an end plate offset of 500 V and capillary voltage of 3000 V were used, with the nebulizer set to 0.3 bar, the dry gas at 3.5 L/min, and the dry temperature at 200 °C.

The instrument parameters were as follows (for a mass range of 300–1350 m/z, using a quadrupole mass filter at 300 m/z with 850 Vpp collision RF): deflection 1 delta = −70 V, funnel 1 RF = 350 Vpp, funnel 2 RF = 350 Vpp, multipole RF = 300 Vpp, isCID = 0 v, ion energy = 5 V, CID = 5 V. The TOF transfer time was 85 µs, and the pre-pulse storage was 10 µs. Imaging parameters such as pixel size and x,y coordinates were set using

Flex Imaging (Bruker Daltonics) software. The majority of images were acquired at a 10 µm pixel size.

Promptly after the end of MALDI experiments, the slides were vacuum-packed and stored at –80 °C and shipped on dry ice to the Institute of Cancer Research to undergo the X-CODE spatial assay for barcode visualisation.

3.17.4 Post-MALDI sample preparation for X-CODE spatial assay

Slides were thawed at room temperature for 2–5 min within the vacuum packaging, then removed from the packaging, and immediately submerged in 2% PFA in PBS for 1 h at room temperature. Following fixation, the slides were washed in PBS for 10 min, then incubated in pre-chilled 30% methanol on ice for 10 min, or until residual matrix material was fully dissolved. Slides were washed in PBS for 10 min, after which the X-CODE spatial assay was performed as described (3.16.2), beginning with application of the hydrophobic barrier as indicated for tissue sections.

3.18 Data analysis and visualisation

X-CODE data analysis was performed with assistance from members of the Tumour Functional Heterogeneity Laboratory at the Institute of Cancer Research (ICR). Suspension X-CODE data analysis was supported by Marco Bezzi. X-CODE sequencing data analysis was performed by Roberto Rota.

Development and optimisation of the X-CODE spatial debarcoding pipeline were carried out in collaboration with Emma Westlund at the Microscopy Facility, ICR and Roberto Rota.

MALDI–MSI analysis was kindly performed by James McKenzie at Imperial College London.

3.18.1 Next-generation sequencing analysis

Raw barcode sequencing data were processed to remove adapter sequences using Cutadapt v1.18 [168]:

- 5′: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
- 3′: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT.

The trimmed R1 and R2 reads were decompressed to .fastq format and aligned using the NGmerge tool v0.5 [169] . Aligned reads were then compared to a predefined barcode whitelist, retaining only perfect matches. A list of the 15,876 unique barcodes, along with their corresponding counts, was generated and used for downstream barcode distribution analysis and plotting.

3.18.2 MALDI-MSI analysis

Matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) analysis was kindly performed by James McKenzie of the Takats Lab at Imperial College London.

All mass spectrometry imaging files were converted from .d format to .mz5 format using ProteoWizard/msConvert (version 3.0.24066-8d99481) [170] and imported into MATLAB 2024b using in-house-developed code.

3.18.2.1 Comparison of MALDI data obtained across laser powers

For each of the 4 samples analysed with different laser powers (30, 40, 50, 60%), a tissue mask was determined to exclude the background pixels, and the mean spectrum across the tissue pixels was determined across the full m/z range at 0.001 intervals. The 4 mean spectra were rescaled and averaged, and peak picking was performed, yielding a total of 1613 peaks. Principal component analysis was performed on total-intensity normalised (multiplied by 1000) and log10-transformed data, with an offset of 1. The non-zero pixel count for each of the 1613 variables was determined as a proportion of the total number of tissue pixels. Variables were sorted into centiles according to their mean intensity, and the 25-50-75th percentiles of peak frequency were determined.

3.18.2.2 MyC-CaP MALDI data analysis

12 MyC-CaP tumour samples were analysed by MALDI-MSI across experimental timepoints: pre-castration samples named 277, 281, and 304; day 4 post-castration samples 295 and 393; day 10 post-castration samples 291 and 300; endpoint castration

samples 285, 292, and 296; and endpoint control (intact) samples 288 and 306. A tissue mask was constructed to separate each of the 12 sections from the background, and the mean spectrum was computed across all pixels for each section. Peak picking yielded 2296 features detected in at least two of the 12 tissue sections. To account for pixel sparsity across the acquisition, image smoothing was performed using a 15 x 15 pixel Gaussian filter (sigma = 3). The dataset was normalised and transformed as described above prior to multivariate analysis using principal components analysis (PCA) and uniform manifold approximation and projection (UMAP). For univariate analysis to determine variable specificity, the data were normalised but not log10 transformed.

3.18.2.3 Co-registration with the consecutive MyC-CaP fluorescence image

Clone information was exported from QuPath into geoJSON format, along with a down-sampled RGB fluorescence image. An image for each clone was generated from the geoJSON coordinates at the same size as the fluorescence image (12672 x 5184 pixels). The UMAP image of components 1-3 for each suitable section was generated and co-registered to its matching fluorescence image. Characteristic regions in each image pair were identified and used to define affine transformations that maximised the overlap of these regions. The clone image for each section was similarly warped, yielding a final image size of 4407 x 2520, matching the dimensions of the mass spectrometry imaging dataset. The final outputs are a fluorescence image and clone images that map directly to the mass spectrometry image. Any tissue sections containing clones with fewer than 20 pixels were discounted. The smallest group was 25 pixels for clone 8466 in sample 303; the largest was 14339 pixels for clone 8958 in sample 285.

3.18.2.4 Variable specificity

Each of the 2296 variables was analysed to estimate its specificity relative to a particular clone across all sections, and to a particular clone in a single section. Only pixels with a clone annotation following co-registration were analysed. Due to the imbalance in pixel counts across clones and samples, a repeated under-sampling approach was employed. A subsample of observations was randomly selected, with which the ROC curve was generated for the group with the highest mean rank (from

Kruskal Wallis test) versus all other groups. From this, the AUC and optimal FPR and TPR rates were determined.

This procedure was repeated 100 times. The sample size (per group) for each of the 6 clones was 1000 observations, and for each of the 21 unique clone/sample combinations was 20 observations. The mean and standard deviation were computed across 100 iterations. Variables were ranked by mean true positive rate but were discounted if there was no agreement between the clones across both comparisons; that is, the group with the highest mean rank in the clone-only analysis must match that in the clone/sample analysis. Tentative annotation of variables was performed using the bulk structure search function of LIPID MAPS structural database (LMSD) with a tolerance of ±0.05 for all $[M-H]^-$ ions [171,172].

3.18.3 X-CODE spatial analysis and debarcoding

3.18.3.1 Cell segmentation

To quantify X-CODE RNA amplicons (fluorescent spots) in images acquired by confocal microscopy, slide scanning, and PhenoCycler-Fusion, raw image files were visualised using QuPath 0.5 and 0.6 [173]. Cellular segmentation was performed with StarDist 2D (model: dsb2018_heavy_augment; channel: DAPI) in QuPath, applying percentile normalisation (1–99%), a detection threshold of 0.5, a pixel size of 0.5 µm, and a cell expansion of 5 pixels when segmenting tissues, and an expansion of 10 pixels when segmenting plated cells [174]. Shape and intensity measurements were recorded for each segmented cell.

3.18.3.2 Spot detection (Fiji-based)

To optimise spot detection parameters, a representative test image was selected, and individual cells were visually inspected. For each barcode unit channel, a single segmented cell was isolated and exported to FIJI/ImageJ for manual threshold assessment [175] . Image contrast was adjusted to identify well-separated, high-intensity puncta with minimal background signal, after which spot detection was performed using the Find Maxima function. The prominence threshold yielding optimal discrimination of true spots was recorded for each barcode sequence. Final spot quantification was performed using a custom QuPath script that incorporated the optimised cell expansion value and per-channel prominence thresholds, with careful attention to QuPath's zero-

indexed channel numbering. Spot counts were generated either per image or across entire projects. Measurement tables were exported to include image identifiers, cell IDs, and per-channel spot counts, enabling downstream aggregation of spot counts per cell across images.

3.18.3.3 Spot detection (QuPath-based)

Spot detection was performed independently for each image channel using QuPath's built-in Subcellular Detection tool. Detection parameters were optimised to ensure inclusion of visually validated true signals while minimising false-positive spot detection from background fluorescence. As standard, the minimum spot size was set to >4 pixels, and per-channel intensity thresholds were manually adjusted based on visual inspection. Additional detection parameters, including smoothing, shape- and intensity-based spot splitting, and cluster handling, were applied. QuPath's scripting interface was used to export the total estimated number of spots and the mean intensity of each spot for each channel (representing a barcode unit) per cell. Subsequent calculations of per-cell mean fluorescence intensity for each channel were performed in Python.

3.18.3.4 15K debarcoding

Cellular barcode unit (BU) measurements and coordinates (centroid X and Y positions) were exported as .csv files for downstream analysis following spot detection using the QuPath-based pipeline. Barcode decoding was performed in Python (v3.11). To reduce inaccuracies in barcode assignment, the 11 most frequently detected BU in each cell were considered for debarcoding. Each cell was then assigned an 18-digit string of ones (1) and zeroes (0) corresponding to the presence or absence of each BU in the library. Each string was compared against a whitelist of 15,876 valid combinations derived from the 18 BUs in the 15K library.

Cells matching a valid combination were assigned the corresponding barcode. Additionally, a further classifier extension is applied to increase the efficiency of spatial debarcoding. Any unassigned cell that is located within 50 µm of an assigned cell, whose sequence differs by at most 2 BUs from the valid barcode, will also be assigned that same barcode. Subsequently, each cell's original sequence and the top 11 adjusted (extended) sequences are exported as a .csv with the assigned barcode. The assigned barcodes were then reimported into QuPath v0.6, and each cell was classified by its

barcode using a custom script that referenced the cell's ID and coordinates. This enables cells to be visualised and categorised by barcode within QuPath's interface.

3.18.4 Data visualisation and figure generation

Flow cytometry and mass cytometry (CyTOF) data (.fcs) were visualised and analysed using FCS Express 7 (v7.28.0035, De Novo Software), including t-SNE, histograms, scatter, contour, and density plots.

Bar plots and heatmaps were generated in GraphPad Prism (v10.1.0) or Microsoft Excel for statistical summarisation and aesthetic customisation.

DNA sequencing data were visualised with SnapGene (v8.2.0) for sequence mapping and annotation, while quantitative DNAseq results and growth curves were plotted using ggplot2 [176] in RStudio (v4.3.1).

Alluvial plots were constructed in Plotly (v5.16.0) for Python to display cell-state transitions or lineage relationships.

Schematic diagrams and cartoon-style illustrations were composed using BioRender (version as of 2026).

Chapter 4 Design and validation of the X-CODE framework for probe-based detection of combinatorial barcodes

4.1 Research in context

Probe-based RNA detection offers an alternative to sequencing-based barcode readout [56,70,160,161]. However, technical challenges that may impede conceptual implementation must be considered. Probe-based RNA detection methods, beyond conventional fluorescence *in situ* hybridisation [177], often rely on signal amplification strategies. This is to achieve sufficient sensitivity and signal-to-noise ratios, particularly in complex cellular or tissue environments. Several amplification methods have been developed, reflecting trade-offs between sensitivity, specificity, multiplexing capacity, and experimental complexity [178] (**Figure 4-1**A, top). Upon transcript amplification, target readout can be achieved by *in situ* sequencing via ligation or synthesis. In contrast, cyclic imaging approaches are used, in which signals can be binary, using on-off patterns, or multi-colour, where unique signal combinations enable identification of the transcript from predetermined patterns (**Figure 4-1**A, bottom).

Among available amplification approaches, rolling circle amplification (RCA) enables sensitive and specific detection with relatively simple probe architectures [111,156,179–184]. RCA-based approaches can vary in probe designs and enzymatic needs. For instance, this is exemplified by comparing several well-established, distinct RCA-based RNA detection methods (**Figure 4-1**B). FISSEQ uses reverse transcription of RNA into cDNA prior to amplification [117], introducing additional enzymatic steps that can reduce efficiency and increase variability. Single-probe approaches, such as BOLORAMIS, simplify probe design by employing minimal padlock structures that directly target RNA [185]. In contrast, the PLAYR uses a layered probe design for transcript targeting, enabling target flexibility and multiplexing [137,141]. Whereas the STARmap hybridises two probes in an SNAIL configuration directly to RNA targets [118].

Following probe hybridisation, RCA-based strategies require circularisation of the DNA template for amplification. This is achieved by a ligase enzyme, which catalyses the formation of a circular DNA product. Then, a primer, within the probe architecture, will be used by a DNA polymerase (e.g., φ29) to produce a concatemeric complementary

amplicon containing multiple copies of the target sequence. Compared with alternative amplification strategies, RCA offers a favourable balance of sensitivity, specificity, and design simplicity [118], producing DNA amplicons that can be readily read out with detection probes tagged with fluorophores or isotopes. These properties make RCA-based RNA detection particularly well-suited for high-complexity barcode detection across multiple readout platforms.

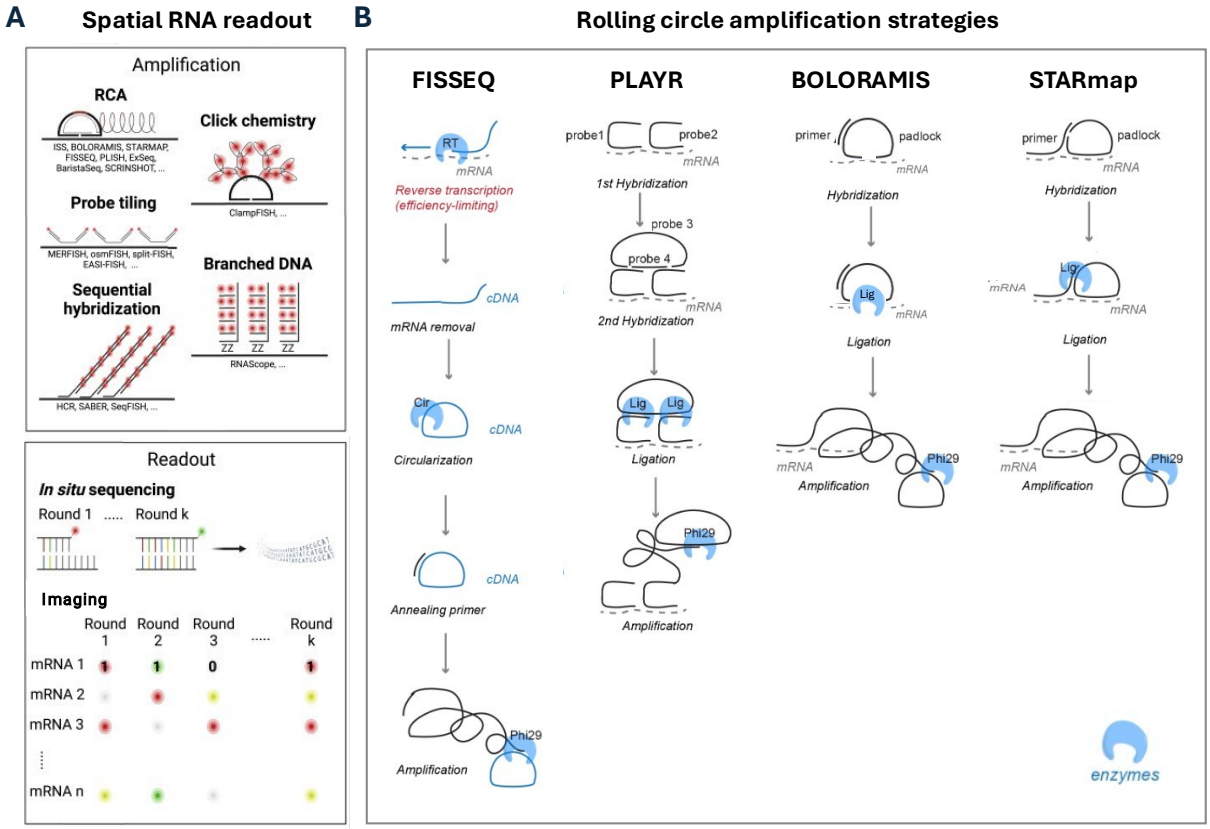


Figure 4-1. Probe-based strategies, using amplification, for spatial profiling of RNA.

A) *Top: Probe-based strategies for spatial RNA profiling and examples of workflows which employ these methods. (RCA) = Rolling circle amplification. Bottom: Readouts include in situ sequencing via ligation or synthesis, or imaging by hybridising fluorescently labelled probes to each target. Signals can be binary or combinatorial, using on-off patterns or colourimetric patterns, or both.*

B) *Examples of RNA detection workflows encompass rolling circle amplification strategies, via differing probe architectures, for increased signal generation. Adapted, with permission, from Ofer et al. [178] and Wang et al. [118].*

This chapter details the design and validation of the X-CODE framework, encompassing the optimisation of probe-based detection, the construction of high-complexity combinatorial libraries, and the incorporation of features that enable downstream functionalisation and clonal retrieval.

4.1.1 Chapter aims

- To design and validate modular barcode units for specific and efficient detection in suspension and spatially resolved readouts.
- To generate high-complexity barcode libraries by combinatorial assembly of barcode units.
- To engineer a multi-capacity barcode construct that supports expanded detection modalities and enables downstream functionalisation.

4.2 Probe-based detection of X-CODE barcode units

Achieving sufficient barcode diversity for clonal lineage tracing with probe-based RNA detection requires generating complexity through combinatorial design rather than using large numbers of unique probes, given the practical limits on multiplexing imposed by probe chemistry and detection platforms. In the X-CODE framework, individual barcode identities are defined by combinations of modular barcode units (BUs). A BU contains a known sequence from a specific, finite set of short RNA sequence elements. Complexity arises when multiple BUs are combined in unique arrangements, generating a library of barcodes. As an initial proof-of-concept, a simple barcode construct, denoted barcode A, was designed to contain two BUs, BU1 and BU2 (**Figure 4-2**A, left) (**Supplementary Table 1**). Therefore, conceptually, when barcode A is introduced into a cell, accurate simultaneous detection of BU1 and BU2 allows that cell to be identified as clone A.

To enable specific and robust detection of individual BUs with a minimal probe set, an RNA detection strategy was developed that integrates established principles of probe-based RNA detection and CRISPR screening readouts [118,141,184]. X-CODE employs a dual-probe architecture in a “snail” configuration, comprising a BU-specific padlock probe (the “snail shell”) and a linear primer probe (the “snail tail”) that must bind in close proximity on the target RNA (**Figure 4-2**A, right). Each BU comprises a 20-base-pair (bp) variable-binding sequence (VBS) targeted by a specific BU-based padlock probe, followed by a 20 bp universal binding sequence (UBS) shared across all BUs. This shared UBS enables detection with a single universal probe, reducing probe complexity while maintaining multiplexing capacity and signal specificity. When both probes bind, a downstream amplification reaction can occur to enhance the barcode signal, generating multiple copies of the barcode sequence for probe-based detection (**Figure 4-2**B).

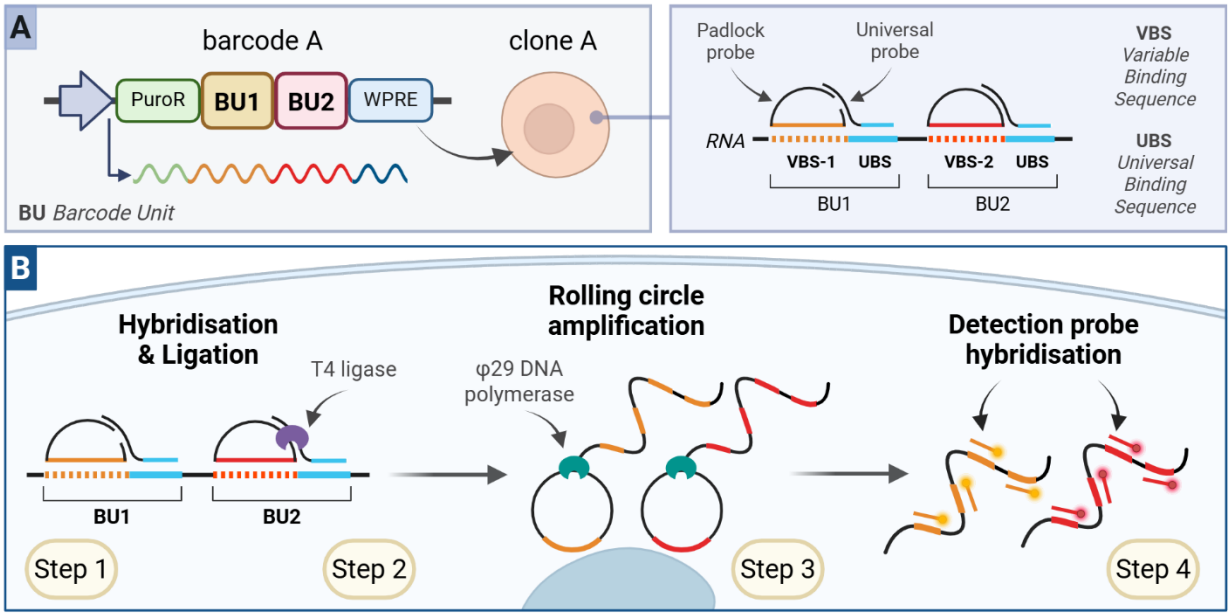


Figure 4-2. Design of a simple barcode for probe-based detection using a limited set of sequences.

A) *A simple barcode composed of two barcode units (BUs), e.g., barcode A (BU1 and BU2), is placed downstream of a constitutive promoter in a construct with a selection marker (e.g. puromycin resistance, PuroR). Cells are stably barcoded by lentiviral infection, selected, and the transcribed barcode RNA is detected via complementary probe binding. Each BU is comprised of a variable binding sequence (VBS), bound by the specific padlock probe, and a universal binding sequence (UBS) present across all BUs, which is bound by the universal probe.*

B) *Schematic of the X-CODE probe-based rolling circle amplification workflow for barcode detection. Padlock and universal probe hybridisation enables circularisation of the padlock probe by a ligase, followed by RCA-based amplicon generation. Fluorophore- or isotope-labelled probes are hybridised for detection, and predetermined signal combinations enable barcode assignment to individual cells. Created with BioRender.com.*

In practice, barcodes are introduced into the population of interest via stable lentiviral integration downstream of a promoter, enabling constitutive expression of barcode transcripts. To detect the barcodes, the cells are fixed and permeabilised to prepare them for padlock-probe-based rolling-circle amplification of BUs. Concurrent hybridisation of the shell and tail probes enables ligation by T4 DNA ligase, closing the padlock probe and using the tail as a primer to generate a circular DNA molecule. The subsequent addition of φ29 DNA polymerase drives rolling-circle amplification, producing localised concatemeric DNA amplicons containing multiple copies of the BU sequence. These amplified products provide a stable substrate for downstream detection with complementary fluorescent or metal-labelled barcode detection probes (**Figure 4-2**B).

4.2.1 Detection of barcode units by flow cytometry and fluorescence imaging

To validate the dual-probe X-CODE detection strategy in suspension, TRAMP-C1 [186] , a murine prostate cancer line, was barcoded to express either BU1 alone, BU2 alone, or

the combined BU1+BU2 configuration (barcode A). Barcoded lines were processed using the X-CODE flow cytometry detection workflow. This protocol integrated established principles from reports detailing the identification of RNA using suspension cytometry platforms, such as the Proximity Ligation Assay for RNA (PLAYR) system, which enables detection of endogenous RNA transcripts by time-of-flight mass cytometry (CyTOF) [137,141]. In brief, cells were fixed with paraformaldehyde (PFA), permeabilised with methanol, and incubated with BU-specific padlock probes and universal tail probes, followed by ligation and rolling circle amplification of the barcode sequences.

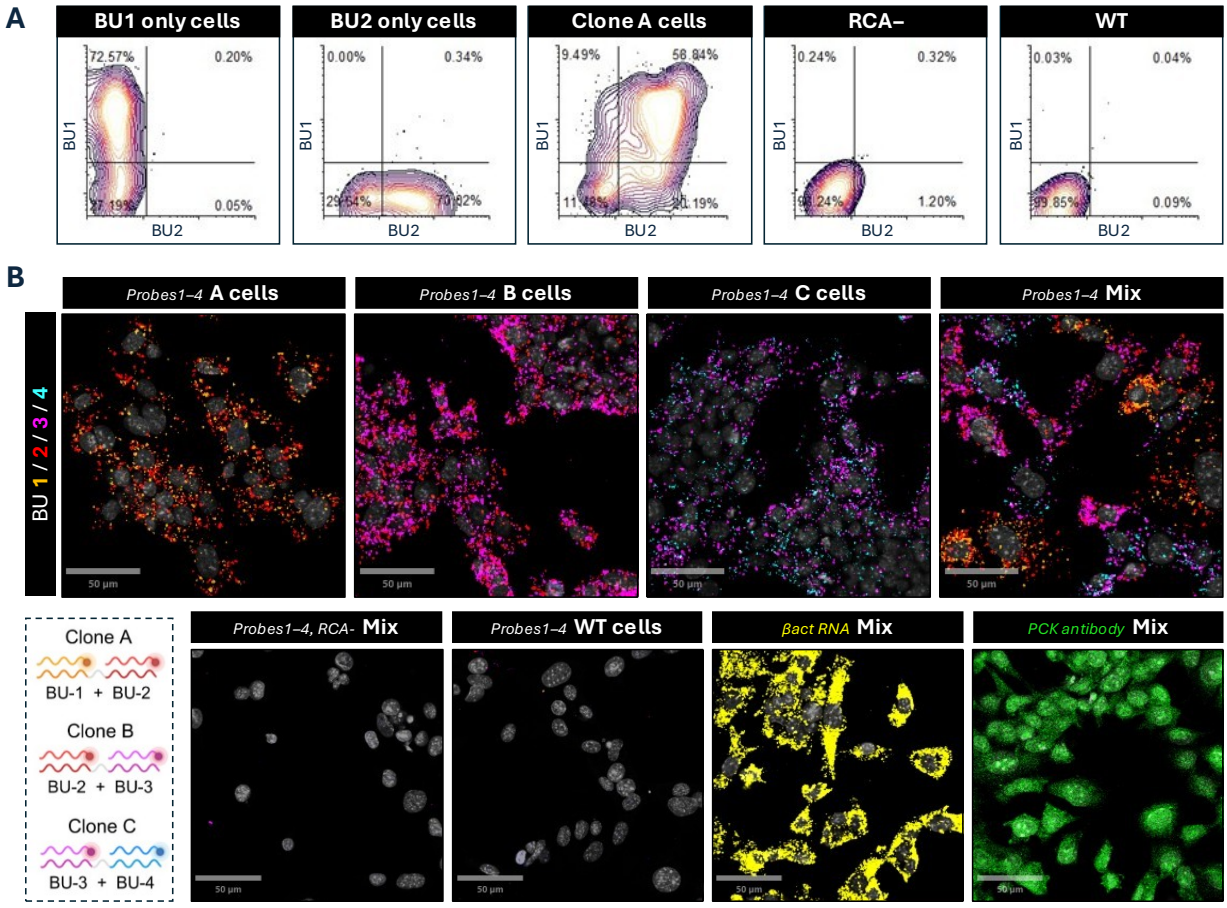


Figure 4-3. Probe-based detection of modular X-CODE barcode units by flow cytometry and fluorescence imaging.

A) *TRAMP-C1 cell lines expressing barcode units (BUs) and wild-type (WT) cells underwent staining using the X-CODE flow cytometry protocol. Contour plots showing the barcoded lines expressing BU1, BU2, or BU1+BU2 (barcode A). Rolling circle amplification negative (RCA–) and WT controls lack the BU fluorescent signal.*

B) *TRAMP-C1 lines expressing barcodes A, B, C, and WT controls were stained using the X-CODE fluorescence imaging protocol. Representative clones for A, B, and C and mixed populations are shown, visualised by BU sequences. A = BU1 (orange) and BU2 (red). B = BU2 (red) and BU3 (pink). C = BU3 (pink) and BU4 (blue). DAPI staining for nuclei (white) and β-actin RNA staining as a protocol control (yellow). Rolling circle amplification negative (RCA–) and WT controls lack the BU fluorescent signal. Antibody co-detection of pan-cytokeratin (PCK, green). Scale bars, 50 µM. Key created with BioRender.com.*

In contrast to the PLAYR assay, which employs isotope-conjugated detection probes for CyTOF readout, barcode amplicons were detected here using BU-specific probes conjugated to distinct fluorophores to enable flow cytometry analysis (**Supplementary Table 2**). Flow cytometric analysis of the barcoded TRAMP-C1 lines demonstrated robust and specific detection of individual BUs. Clone A cells were readily resolved by dual-positive BU1 and BU2 probe signals (**Figure 4-3**A). Also, single-BU-expressing lines showed clear probe signals for BU1 or BU2, but not both. In contrast, omission of the φ29 DNA polymerase, which catalyses rolling-circle amplification (RCA), resulted in no signal in this sample (RCA–), illustrating the dependence of signal generation on RCA. Additionally, wild-type (WT) TRAMP-C1 cells lacking barcode constructs were processed and stained identically to barcoded cells. Their analysis shows minimal background signal, confirming barcode specificity of the detection strategy.

Having established reliable BU detection in suspension, the X-CODE protocol was next adapted for spatial barcode detection by fluorescence imaging. Two additional TRAMP-C1 barcoded cell lines were generated, each harbouring a small-scale barcode composed of a unique combination of two of four barcode units: clone B, containing BU2 and BU3, and clone C, containing BU3 and BU4 (**Supplementary Table 1**). Clones A, B, and C were seeded either individually or as equal-proportion mixtures onto multi-chamber slides, along with WT TRAMP-C1 cells, and processed using the X-CODE imaging workflow for plated cells. This protocol includes minor adaptations to the X-CODE flow cytometry-based detection protocol; however, the core stages of signal generation remain, and a shared set of reagents is used. Cells were fixed with PFA, permeabilised with methanol, and hybridised with the universal probe and padlock probes targeting BU1–BU4. The downstream ligation and rolling circle amplification reactions were performed, and amplified BUs were detected using four BU-specific probes, each conjugated to a distinct fluorophore (**Supplementary Table 2**).

Fluorescence imaging of stained cells visualised BU signals as discrete puncta (spots) distributed throughout the cytoplasm (**Figure 4-3**B). This is consistent with reports describing RNA detection by padlock-probe-mediated rolling-circle amplification [118,156,185,187]. Specifically, clone A cells showed clear positivity for BU1 and BU2; clone B cells for BU2 and BU3; and clone C cells for BU3 and BU4, enabling unambiguous discrimination of all three clones within mixed populations. Detection of β-actin RNA using probes with the same snail-shell-and-tail architecture yielded a strong, uniform

signal in the mixed barcoded cells. Given β-actin's universal expression, its staining serves as an internal protocol control for overall RNA integrity, reagent performance, and imaging quality. Additionally, a limited background signal was observed in WT cells for barcode probes, indicating high specificity. Furthermore, barcode-protein co-detection was assessed by simultaneous antibody staining and imaging with pan-cytokeratin (PCK), a common epithelial marker often used to aid cellular segmentation [188].

Together, these results demonstrate that combinatorial assembly of a limited pool of short, probe-detectable RNA elements in modular barcode units enables robust barcode identification by both flow cytometry and spatial fluorescence imaging.

4.3 Construction of high-complexity combinatorial X-CODE libraries

Following validation of an efficient, specific detection protocol for modular barcode units using suspension and spatial modalities, a strategy for assembling high-complexity X-CODE libraries was developed. This required consideration of three significant constraints. Firstly, the generation of high-complexity barcodes that remain compatible with sequencing-free detection platforms that use labelled probes. In practise, this means detecting clonal resolution libraries with only tens of parameters by imaging or flow cytometry, while maintaining capacity for simultaneous phenotyping alongside barcode detection. For example, high-parameter technologies like mass cytometry, which currently use panels of ≥40 isotopes [138]. Fluorescence-based modalities, such as Akoya PhenoCycler-Fusion for imaging, which has a ≥100-marker capacity per experiment [106]. As well as the rapidly developing full-spectrum flow cytometry, which can detect up to 40 fluorescent markers simultaneously [138].

Secondly, the design must allow for scalable cloning without the cost-prohibitive synthesis of thousands of unique long-barcode oligonucleotides. Clonal resolution static libraries are currently considered to contain between 10^4 and 10^7 unique barcodes [56,70,76]. However, producing thousands of individual sequences for barcode cloning is practically infeasible and prohibitively expensive at current synthesis capacities. Thirdly, it was important to consider additional capabilities of the X-CODE construct. Notably, the incorporation of a compact secondary barcode compatible with sequencing-based readouts enables comprehensive cross-platform detection. Nonetheless, this "short

barcode" would be ineffective without remaining intrinsically linked to the probe-detectable barcode identity. Additionally, the functionalization of this short barcode must be considered to enable barcode-specific clonal retrieval for downstream analysis of clones of interest.

With these factors in mind, 27 variable barcode sequences (VBS) were selected for barcode unit construction. 23 of these unique sequences were adapted from the insert probes previously validated for specific and efficient detection of endogenous RNA in the PLAYR system [137,141] (**Figure 4-1**B, labelled *probe 4*). An additional 4 sequences were developed in accordance with the PLAYR criteria [137] (**Supplementary Table 7**). The single universal binding sequence (UBS) was designed by combining two backbone-binding regions described in the same work (**Figure 4-1**B, labelled *probes 1 and 2*). The UBS was paired downstream of each VBS to generate 27 BUs, each 41 base pairs (bp) in length (**Figure 4-4**A). All sequences were evaluated for GC content, predicted secondary structure, and annealing temperature.

4.3.1 A combinational cloning strategy for library diversification

The construction of the X-CODE libraries, as well as the generation of sufficient diversity, was designed to be achieved through a combinatorial, pooled ligation strategy, inspired by the combiGEM methodology [163] . The 27 BUs were divided into three groups of nine: group 1 containing BUs 1–9, group 2 containing BUs 9–18, and group 3 containing BUs 18–27 (**Figure 4-4**B). All possible unique BUs arrangements using four BUs within each group of nine (C(9,4) = 126) were generated and synthesised. Each combination was linked to a 4-bp unique set identifier (USI) contributing to the variable region of the secondary short barcode. The number of sequences synthesised for library generation, therefore, amounted to 378 oligonucleotides (126 for each of the three groups) (**Figure 4-4**C). The maximum BU combination using groups 1 and 2 generates 15,876 unique arrangements of eight BUs (126 * 126). Subsequently, combining groups 1, 2, and 3 yields a total of 2,000,376 unique arrangements of twelve BUs (126 * 126 * 126) (**Figure 4-4**D).

The combiGEM methodology utilises pairs of restriction endonucleases that recognise different DNA sequences but generate compatible cohesive (sticky) ends after digestion [163] . This allows fragments cut by one enzyme to ligate to fragments cut by the other.

Group 1 BU-combinations were designed to contain an MluI restriction site at the 5' end and a SalI restriction site at the 3' end to facilitate cloning into the backbone vector. Additionally, a restriction cassette comprising a BamHI restriction site, an 8-bp linker, and an EcoRI restriction site was inserted between the BUs and the USI. Groups 2 and 3 shared a consistent design with a BglII restriction site at the 5' end and an MfeI restriction site at the 3' end. Also, the BamHI–EcoRI restriction cassette was incorporated between the BUs and USI. Upon digestion, the MfeI and EcoRI sites generate compatible cohesive ends, as do the BglII and BamHI sites, enabling directional pooled ligation of the oligonucleotide groups 1, 2, and 3 (**Figure 4-4**D).

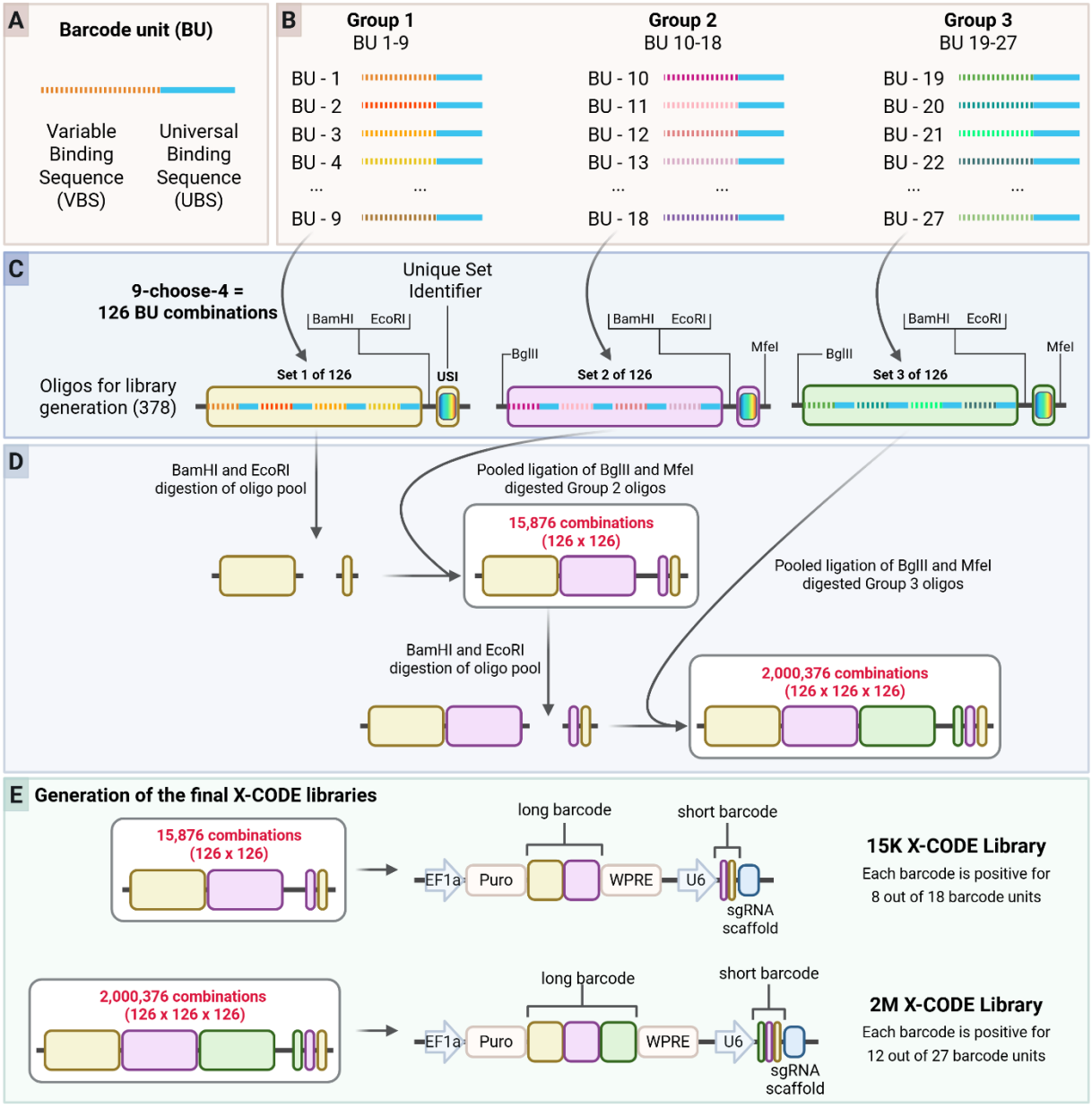


Figure 4-4. Construction of 15K and 2M combinatorial X-CODE libraries.

Schematic representation of the combinatorial strategy developed for the assembly of the 15K and 2M X-CODE barcode libraries. Created with BioRender.com.

A) *An X-CODE barcode unit (BU) is composed of a variable binding sequence (VBS) and a universal binding sequence (UBS).*

B) *27 BUs are grouped into three groups of nine: group 1 (BU1 – 9), group 2 (BU 10 – 18), group 3 (BU 19 – 27).*

C) *Combinatorial arrangement of four sequences (9-choose-4) from each group yields 126 unique combinations per group, for a total of 378 oligos for library generation (126 * 3).*

D) *Each combination of four BUs (BU sets) is linked to a unique set identifier (USI), which can be sequenced to retain barcode readout across modalities. Placing BamHI and EcoRI restriction sites between the BU sets and their USIs allows pooled ligation of groups 1–3 by digesting the incoming group with BglII and MfeI.*

E) *Combining groups 1 and 2 generates a 15K X-CODE library of 15,876 combinations (126*126) where each barcode is detected by 8-of-18 specific sequences. Combining groups 1, 2, and 3 generates a 2M X-CODE library of 2,000,376 combinations (126*126*126), where each barcode is detected by 12-of-27 specific sequences.*

Therefore, the three groups of 126 BUs combinations can be sequentially combined through iterative cycles of pooled digestion and ligation (**Figure 4-4**D), expanding library complexity from 126 to 15,876 and ultimately to 2,000,376 unique barcodes. Specifically, digestion of the plasmid backbone and group 1 oligonucleotides with MluI and SalI enabled pooled ligation of the first 126 combinations. Subsequently, the plasmid was digested with BamHI and EcoRI between the group 1's BUs and USI. Digestion of the group 2 oligonucleotides using BglII and MfeI enables pooled ligation of the second group between group 1's BUs and USI. This digestion reaction is repeated to integrate group 3 oligonucleotides between group 2's BUs and USI (**Figure 4-4**D). The resultant construct contains the correct sequential alignment of the BUs comprising the long barcode, followed by the sequential alignment of the USIs comprising the linked short barcode.

After which, the BamHI–EcoRI restriction cassette, between group 3's BUs and USI, enables the insertion of plasmid elements into the vector for the proper integration and expression of the barcode transcripts. Therefore, in the final design, long and short barcodes are inherently matched, enabling direct cross-referencing across sequencing- and probe-based datasets. Long barcodes for probe-based detection are encoded by unique combinations of 8-of-18 or 12-of-27 BUs, respectively, while short barcodes consist of two or three 4-bp blocks separated by fixed 6-bp linker sequences (**Figure 4-4**E). The short barcode is expressed downstream of a U6 promoter and upstream of a scaffold sequence, enabling barcode-specific clonal retrieval and providing a compact solution for DNA sequencing and scRNA-seq–based barcode recovery.

Therefore, this conceptual, pooled ligation strategy enables the creation of two high complexity barcode libraries with unique combinations: a 15,876-barcode (15K) library and a 2,000,376-barcode (2M) library.

4.4 Assessment of X-CODE library diversity and balance

The combinatorial pooled ligation strategy was implemented by Roberto Rota from the Tumour Functional Heterogeneity (TFH) Lab at the Institute of Cancer Research (ICR) to generate both the 15K- and 2M-plex libraries. For the construction of the 15K library, oligonucleotide groups 1 and 2 were ligated in a pooled reaction to generate combinatorial plasmid assemblies. The resulting library was purified, bacteria-amplified, and re-purified prior to transfection into packaging cells to produce the X-CODE 15K lentiviral preparation for downstream cellular transduction. The generation of the higher-complexity 2M library was achieved using a similar strategy, omitting bacterial amplification because propagation of the highly diverse plasmid pools was prone to recombination (**Figure 4-5**A).

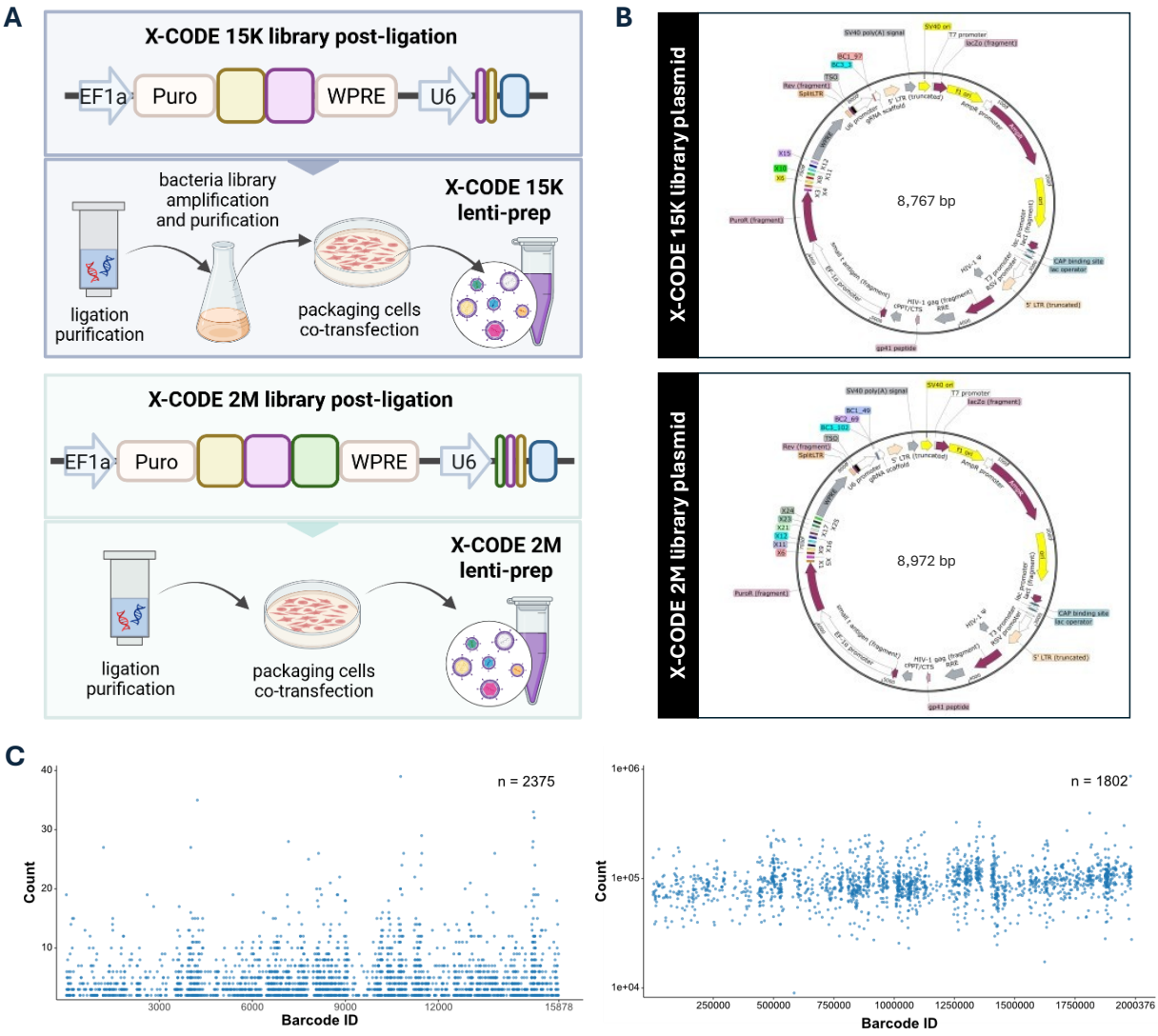


Figure 4-5. Combinatorial cloning generates diverse and balanced X-CODE libraries.

A) *Schematic representation of the experimental workflow followed for the generation of the X-CODE 15K and 2M barcode libraries. Created with BioRender.com.*

B) *The expected barcode architecture was confirmed by whole-plasmid sequencing for the X-CODE 15K and 2M barcode libraries.*

C) *Next-generation amplicon sequencing analysis of barcode representation from RM-1 cells transduced at low MOI with X-CODE 15K and 2M libraries. Short barcodes were PCR-amplified, gel-purified, and sequenced. Scatter plots show individual barcode IDs on the x-axis and read counts on the y-axis. A total of 2,375 unique barcodes were detected in RM-1 15K cells, with plots showing barcodes with counts > 1. A total of 1,802 unique barcodes were detected in RM-1 2M cells, with plots filtered for barcodes with counts > 8,000.*

Whole-plasmid sequencing was performed to validate the correct assembly of both libraries and to confirm the expected pairing between long and short barcode elements. Sequencing data demonstrated successful construction of the intended barcode architectures in both the 15K and 2M libraries (**Figure 4-4**B). To assess the diversity and balance of the generated X-CODE barcode libraries following cellular delivery, RM-1 murine prostate cancer cells were transduced with the X-CODE 15K or 2M lentiviral preparations at low multiplicity of infection, a condition chosen to minimise the probability of multiple barcode integrations per cell. Genomic DNA was extracted from infected populations, and barcode regions were amplified by polymerase chain reaction (PCR) and analysed by amplicon-based next-generation sequencing. Sequencing revealed broad barcode representation and relatively even barcode frequencies across both libraries, consistent with balanced library composition following lentiviral delivery (**Figure 4-4**C).

Together, these results demonstrate that the combinatorial pooled ligation strategy enables the generation of diverse and balanced X-CODE barcode libraries. Validation by whole-plasmid sequencing and next-generation sequencing following cellular infection confirms preservation of barcode complexity and representation in both the 15K and 2M libraries, supporting their suitability for downstream lineage-tracing applications.

4.5 Functionalisation of the X-CODE construct for live clone retrieval

To enable barcode-specific retrieval of clones of interest from viable cell populations, the X-CODE construct was designed to be compatible with established CRISPR-based recall strategies such as the CaTCH and ClonMapper [68,76]. This was achieved using a CROPseq–derived vector architecture [189], in which the X-CODE short barcode is expressed from a U6 promoter as the variable spacer within a complete single-guide RNA (sgRNA) scaffold (**Figure 4-6**A, left). Following lentiviral integration, this cassette

serves a dual function, supporting both sequencing-based barcode identification and CRISPR-mediated live clonal retrieval. Similar architectures have been used in recent systems to enable barcode-specific clonal retrieval via CRISPR activation, typically by recruiting transcriptional activators to barcode-matched reporter constructs [68,76]. Within the X-CODE framework, this design allows short barcodes embedded within high-complexity combinatorial libraries to be directly repurposed as functional guides for clone-specific recall.

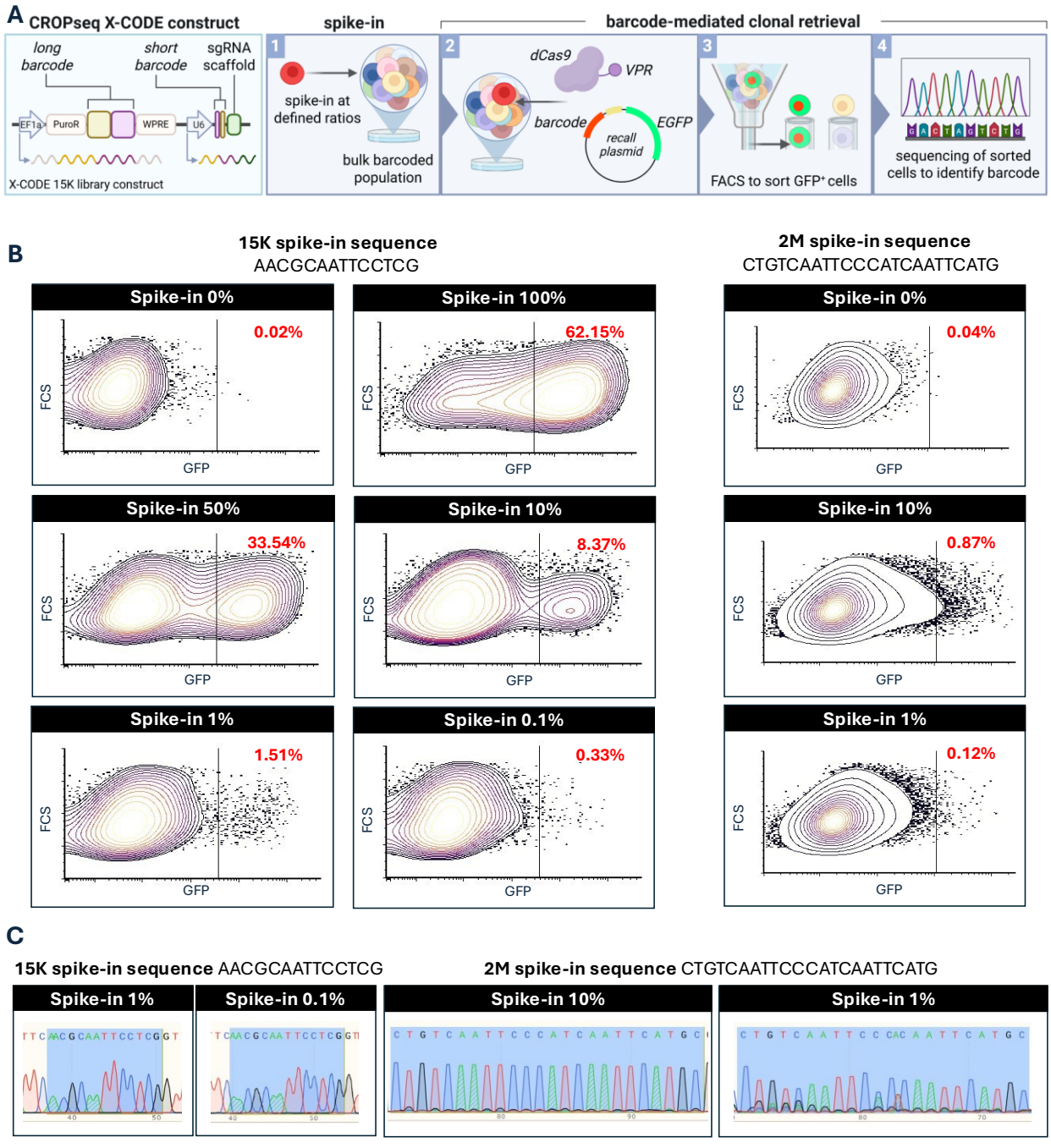


Figure 4-6. Live-cell isolation of barcode-defined clones using the X-CODE framework.

A) *Schematic representation of the X-CODE CROPseq-based construct (left) and the experimental workflow for the clonal spike-in retrieval sensitivity assay using X-CODE recall plasmid activation. 1: A population of barcoded cells*

carrying a specific barcode was spiked into a bulk X-CODE–tagged cell population at defined ratios. 2: Cells were transfected with dCas9-VPR and a barcode-specific recall plasmid, inducing enhanced green fluorescent protein (EGFP) expression only in cells harbouring the cognate barcode. 3: After 48 h, GFP⁺ cells were quantified by flow cytometry, sorted, and 4: sequenced for validation of the barcode sequence. Created with BioRender.com.

***B)** Flow cytometry contour plots showing EGFP activation across spike-in dilution conditions for the 15K and 2M barcodes and their respective bulk populations. Plots show forward scatter (FCS) on the y-axis and GFP⁺ signal intensity on the x-axis, and the percentage of spike-in cells detected in each condition is indicated in red. A 15K barcode (AACGCAATTCCTCG) was spiked in at 0%, 100%, 50%, 10%, 1%, and 0.1%. A 2M barcode (CTGTCAATTCCCATCAATTCATGC) was spiked in at 0%, 1%, and 0.1%.*

***C)** Sorted GFP⁺ populations were subjected to Sanger sequencing, where sequencing chromatograms show the recovered barcode sequenced highlight in blue for the 1% and 0.1% spike-in conditions for the 15K clone, and the 10% and 1% spike-in conditions for the 2M clone.*

To implement barcode-directed retrieval, cells are transiently transfected with a recall reporter plasmid containing barcode-matched target sites upstream of an enhanced green fluorescent protein (EGFP) reporter, together with a nuclease-dead Cas9 fused to a transcriptional activation domain (dCas9–VPR). In this configuration, binding of the sgRNA barcode to its cognate target sequence recruits dCas9–VPR to the reporter, resulting in barcode-specific EGFP expression. Therefore, this modular system requires three components for recall: the sgRNA-encoded barcode present in the integrated X-CODE construct, dCas9–VPR to mediate transcriptional activation, and a barcode-specific reporter plasmid (**Figure 4-6**A, right).

To validate the compatibility of the X-CODE construct with barcode-specific clonal retrieval using short barcodes of varying lengths, two clonal cell lines were generated, each carrying a unique barcode derived from either the 15K or 2M library. Roberto Rota from the TFH Lab at the ICR spiked these clones into bulk populations at defined frequencies, each labelled with the corresponding library (**Figure 4-6**B). The specific 15K clone with the sequence AACGCAATTCCTCG was spiked into the bulk 15K library-expressing cell population at 0.1%, 1%, 10%, 50%, and 100%. The specific 2M clone with the sequence CTGTCAATTCCCATCAATTCATGC was spiked into the bulk 2M library population at 1% and 10%. Mixed populations were then transfected with the appropriate barcode-specific recall reporter, designed as previously described [76], together with dCas9–VPR.

Forty-eight hours post-transfection, EGFP-positive cells were isolated by fluorescence-activated cell sorting (FACS), and barcode identity was confirmed by Sanger sequencing (**Figure 4-6**C). Barcode recall scaled with spike-in frequency across multiple orders of magnitude, whereas recovery efficiency decreased at higher clone frequencies, consistent with previously reported CRISPR-based recall systems [76] . For

example, the 15K 0.1% and 1% spike-in samples yielded 0.33% and 1.51% cells recovered by FACS analysis, respectively, compared with 33.54% at 50% spike-in. Lower recovery rates were observed with the specific 2M barcode spike-in in the bulk 2M population: 0.12% and 0.87% of cells were recovered after 0.1% and 10% clone spike-ins, respectively.

Together, these results demonstrate that the X-CODE construct can be functionally extended beyond cellular tracking to enable barcode-specific retrieval of viable clones of interest. The CROP-seq–derived sgRNA expression module directly couples combinatorial barcoding with CRISPR-mediated recall, enabling isolation and functional interrogation of selected clones from complex populations.

4.6 Discussion

Together, the results presented in this chapter establish the feasibility of the X-CODE framework as a probe-based, combinatorial barcoding strategy for clonal tracing applications. By encoding barcode identity through the modular assembly of short, probe-detectable RNA barcode units, X-CODE enables robust and specific barcode readout using both flow cytometry and spatial fluorescence imaging. Notably, X-CODE detection across suspension and spatial platforms is based on a shared core reagent set, reducing additional reagent costs and enabling flexible workflow adaptation across different readout technologies. A pooled ligation strategy, designed following established concepts [163], was implemented to generate high-complexity barcode libraries composed of unique combinations of barcode units. This design decouples barcode diversity from the number of probes, allowing high-complexity barcodes to be resolved using a constrained, reusable probe repertoire.

Using this combinatorial ligation approach, two libraries were constructed, comprising 15,876 and 2,000,376 distinct barcode identities, respectively. Whole-plasmid sequencing and next-generation sequencing following cellular infection demonstrated that the combinatorial pooled ligation strategy enables the generation of highly diverse and balanced X-CODE barcode libraries. Barcode representation was maintained across both the 15K and 2M libraries, indicating efficient library delivery and minimal compositional bias. In addition, incorporating elements into the X-CODE construct enabled the selective retrieval of individual clones from complex populations, consistent

with established recall strategies [76]. Barcode recall was scalable across several orders of magnitude and exhibited recovery dynamics comparable to those of previously reported systems [68,76], supporting the compatibility of X-CODE with targeted clonal isolation workflows.

Existing spatially resolved barcoding strategies encompass a diverse range of implementations, each optimised for specific experimental objectives [72,92,160,183]. For example, protein-based barcoding approaches enable spatial detection through antibody-based platforms. Probe-based RNA barcodes have been successfully applied to optical pooled CRISPR screens and to the detection of evolvable barcodes for dynamic lineage recording. Although these methods have provided important insights, they frequently impose practical trade-offs between achievable clonal resolution, phenotypic depth, spatial information content, and experimental scalability. As a result, resolving lineage identity alongside molecular state and spatial context within the same system remains a persistent technical challenge. This work represents the first demonstration of a high-complexity static RNA barcode system that can be visualised with a limited set of probes, with the potential to provide platform-agnostic resolution of clonal dynamics and associated molecular states across diverse experimental systems.

The X-CODE framework addresses several of these limitations through a dual-expressed RNA barcoding architecture, in which a long barcode is optimised for probe-based amplification and spatial detection, while a matched short barcode supports sequencing-based readout and barcode-specific clonal retrieval. This design enables continuous multi-platform clonal tracing across probe- and sequencing-based workflows. Even at the highest library complexity generated, approximately two million unique barcodes, X-CODE detection requires a fixed set of probes across hybridisation and detection steps, comprising 27 padlock probes, a universal probe, and 27 labelled detection probes. In addition, using a predefined barcode whitelist further improves demultiplexing accuracy and analytical robustness compared with fully random barcode libraries. Together, these features position X-CODE as a scalable and flexible framework for integrating high-resolution clonal tracing with multimodal phenotypic and spatial analysis.

4.6.1 Limitations

While this chapter demonstrates the successful generation of high-complexity barcode libraries by sequencing, probe-based detection was validated only for simple two-unit barcodes. Suspension-based detection by flow cytometry and spatial detection using confocal imaging confirmed that the underlying barcode unit architecture can be visualised robustly and specifically. However, resolving the full combinatorial complexity of the X-CODE libraries requires higher-parameter technologies. In practice, detecting the 15K library requires at least 18 independent parameters, while detecting the 2M library requires at least 27. An additional experimental consideration is the incorporation of simultaneous phenotypic measurements, such as protein detection via tagged antibodies, alongside barcode detection, which further increases the required marker capacity.

Therefore, the following chapters focus on optimising the detection of highly multiplexed barcode units using high-parameter platforms, which provide the necessary detection capacity. These include suspension- and imaging-based mass cytometry, which leverage isotope-conjugated probes to overcome the spectral overlap inherent to fluorescence-based assays. They currently provide approximately 40-marker capacity, and a theoretical capacity of approximately 120 parameters is reported, dependent on the stable isolation of isotopes that can be used for labelling [138] . In addition, the Akoya PhenoCycler Fusion system, which automates highly multiplexed, iterative fluorescence imaging, can visualise more than 100 markers in a single experiment. Furthermore, ongoing advances in other high-parameter technologies, such as spectral flow cytometry, will help mitigate capacity constraints and enable increased utilisation of the X-CODE framework.

Although the X-CODE construct includes a short, sequence-recoverable barcode, single-cell RNA sequencing (scRNA-seq) was not performed in this chapter. Nevertheless, the X-CODE library cloning strategy incorporated a template-switch sequence into the construct, which has previously been reported to enhance transcript capture efficiency in scRNA-seq workflows [76] . As a result, robust scRNA-seq–based detection of X-CODE barcodes should be feasible using established protocols [76] .

4.6.2 Conclusions

In conclusion, this chapter establishes X-CODE as a modular, combinatorial RNA barcoding framework that supports scalable clonal tracing across analytical platforms. By validating the barcode unit architecture and probe-based amplification strategies, robust, specific detection of barcode units was achieved in both suspension-based and spatially resolved readouts. Combinatorial assembly of these units enabled the generation of high-complexity barcode libraries with balanced representation. Importantly, integration of additional design features extended the capabilities of the X-CODE construct, supporting sequencing-based validation and barcode-specific retrieval of viable clones for downstream functional interrogation. Together, these findings demonstrate that combinatorial barcode design, coupled with probe-based RNA amplification, provides a robust foundation for flexible, high-resolution clonal analysis across analytical techniques.

Chapter 5 High-parameter suspension detection of X-CODE barcodes by mass cytometry

5.1 Research in context

Mass cytometry by time-of-flight (CyTOF) is a single-cell analysis technology that combines antibody-based detection with inductively coupled plasma time-of-flight mass spectrometry, enabling highly multiplexed measurements. In CyTOF, antibodies are conjugated to stable heavy-metal isotopes rather than fluorophores, eliminating signal overlap and enabling simultaneous detection of dozens of molecular features per cell. Experimentally, cells are stained with panels of metal-tagged antibodies, nebulised, atomised, and ionised within a plasma torch. The isotope signals are then separated by a time-of-flight mass analyser. These features enable precise, low-background detection over a wide dynamic range and support high-parameter single-cell profiling [142]. Currently, CyTOF supports panels with more than 40 parameters per cell, allowing many measurements to be integrated into a single experiment [190,191].

A defining strength of mass cytometry by time-of-flight (CyTOF) is its scalability in both cellular throughput and multiplexing capacity. Modern instruments routinely acquire millions of single cells per day, enabling statistically robust interrogation of complex, heterogeneous populations [142]. Sample multiplexing strategies, such as palladium-based mass-tag barcoding, further increase throughput by enabling multiple samples to be stained, pooled, and analysed simultaneously while remaining computationally separable post-acquisition [192]. More recently, the thiol-reactive organoid barcoding *in situ* (TOB*is*) workflow was implemented, enabling robust, scalable workflows for multiplexed single-cell analysis of organoid cultures using mass cytometry [193,194]. These multiplexing approaches reduce technical variability and improve experimental efficiency, particularly in large-scale studies.

Additional methodological developments have enhanced the experimental capabilities of mass cytometry. The use of probe-based RNA detection by mass cytometry has broadened its applications beyond protein-level analysis. The Proximity Ligation Assay for RNA (PLAYR) workflow demonstrated that oligonucleotide probes conjugated to metal isotopes can detect targeted endogenous RNA transcripts alongside protein

markers at single-cell resolution. This approach employs hybridisation-based probe strategies analogous to those used in fluorescence *in situ* hybridisation but leverages isotope-encoded readout to increase target capacity [137,141]. **Figure 5-1** provides an overview of the bench steps required to perform the PLAYR assay for co-detection of transcripts and proteins. The PLAYR workflow extends CyTOF measurements beyond protein phenotyping, providing a foundation for multimodal single-cell analysis using suspension mass cytometry.

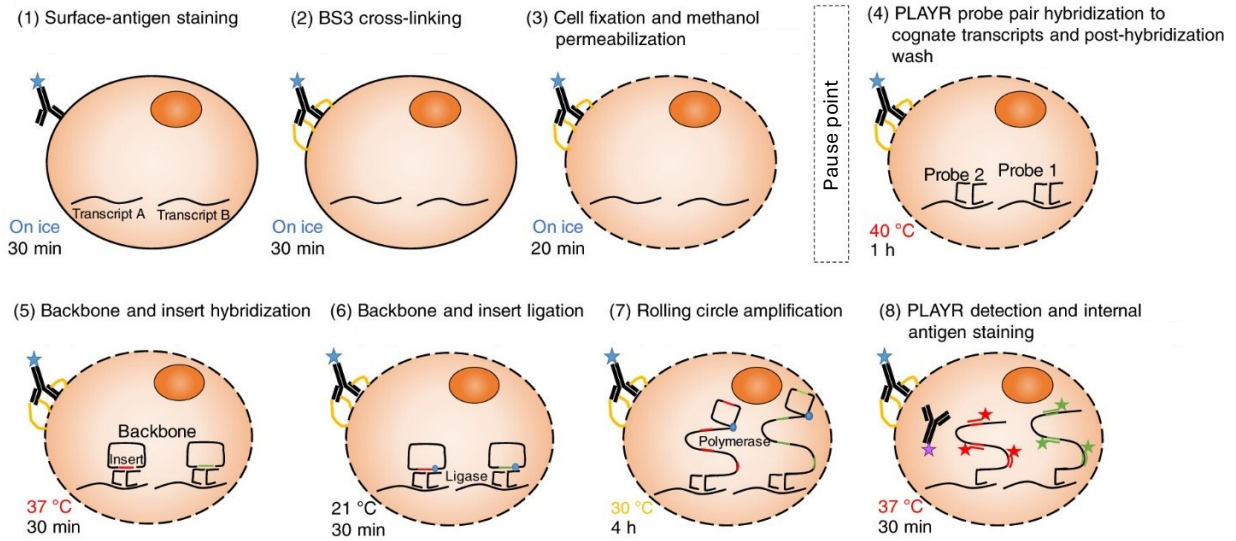


Figure 5-1. Schematic representation of the Proximity Ligation Assay for RNA (PLAYR).

Schematic overview of PLAYR for simultaneous staining of targeted mRNAs with surface and intracellular proteins. Surface antigen staining is performed in unfixed cells on ice (1), then bound antibodies are cross-linked to the cells using BS3 (bis(sulfosuccinimidyl)suberate) (2). Cells are then fixed with paraformaldehyde and permeabilised using methanol (3). Cells can then be stored at −80 °C in methanol for several months (Pause point). To perform RNA detection using PLAYR, the probe pair is hybridised to targeted transcripts, followed by a post-hybridisation wash (4). Then the backbone and insert probes are hybridised (5), ligated (6), and then rolling circle amplification is performed (7). For CyTOF acquisition, the cells are incubated with isotope-labelled detection probes and intracellular antibodies (8). Adapted, with permission, from Duckworth et al. [137]

Collectively, these technological advances position CyTOF as a suitable platform for probe-based readout of engineered RNA barcodes. Its high parameter capacity enables simultaneous detection of multiple barcode units alongside extensive protein phenotyping. Additionally, its throughput supports analysis of large clonal populations at single-cell resolution. By adapting X-CODE detection workflows for isotope-based readout, this chapter explores the feasibility of barcode detection using suspension mass cytometry. It evaluates the detection of combinatorial X-CODE barcode units and examines the co-detection of RNA barcodes with antibody-based protein phenotyping. Finally, it assesses the detection of high-complexity barcoded cell lines from the 15K and 2M X-CODE libraries.

5.1.1 Chapter aims

- To validate isotope-based detection of modular X-CODE barcode units by CyTOF.
- To establish compatibility of X-CODE detection with antibody-based phenotypic staining workflows.
- To demonstrate efficient detection of high-complexity X-CODE barcoded line cells by CyTOF.

5.2 Detection of modular X-CODE barcode units by CyTOF

To assess whether X-CODE barcode units (BUs) could be detected using mass cytometry by time-of-flight (CyTOF), X-CODE detection was adapted for isotope-based readout. BU-detection probes were conjugated to distinct metal isotopes according to previously established protocols [137] , enabling barcode readout to be integrated directly into CyTOF panels for acquisition (**Figure 5-2**A). Simple-barcoded TRAMP-C1 A, B, and C cell lines, previously validated by fluorescence imaging, were expanded in culture, mixed at equal proportions, and processed using the X-CODE CyTOF protocol. This workflow was influenced by the PLAYR methodology (**Figure 5-1**) [137], substituting hybridisation of endogenous RNA transcripts with a four-probe strategy for the detection of X-CODE barcode units with the validated SNAIL configuration. Cells were stained with isotope-conjugated X-CODE detection probes for BUs 1–4 (**Supplementary Table 2**). The stained mixed TRAMP-C1 population and unbarcoded TRAMP-C1 wild-type (WT) controls were acquired on a Helios mass cytometer (Standard Biotools).

CyTOF acquisition revealed robust signals across all four isotope channels, indicating successful detection of BU1–BU4 (**Figure 5-2**B). TRAMP-C1 A cells harbour barcode A composed of BU1 and BU2; similarly, TRAMP-C1 B cells express BU2 and BU3, and TRAMP-C1 C cells express BU3 and BU4. Density plots showing bivariate BU signals on the x- and y-axes enabled clear discrimination of A, B, and C cells within the mixed population. For example, plotting BU1 versus BU2 identified a distinct double-positive $BU1^{+}BU2^{+}$ population corresponding to TRAMP-C1 A cells. Within the same plot, a $BU1^{-}BU2^{+}$ population represented TRAMP-C1 B cells, while the double-negative $BU1^{-}BU2^{-}$ population corresponded to TRAMP-C1 C cells, consistent with their lack of BU1 and BU2 expression. In contrast, density plots generated from unbarcoded

TRAMP-C1 WT cells showed no signal across all four BU detection channels, demonstrating the specificity of X-CODE barcode detection by CyTOF.

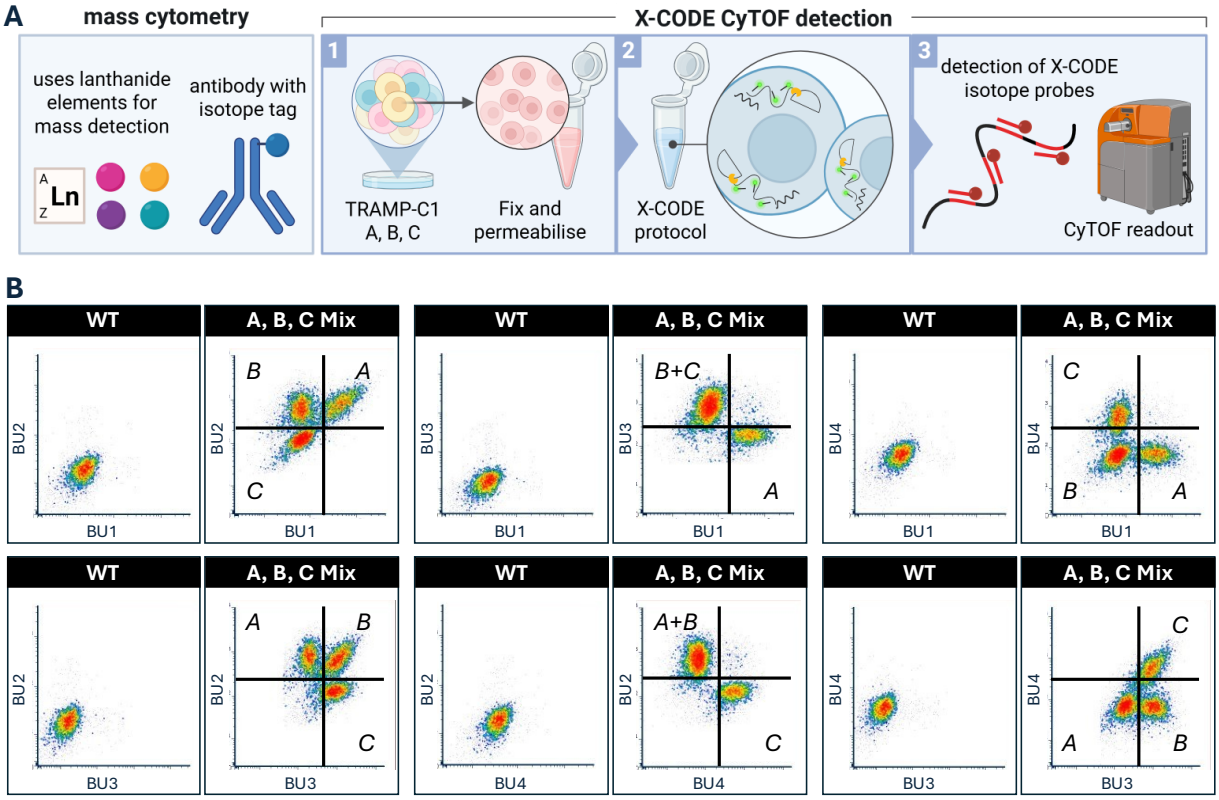


Figure 5-2. Detection of X-CODE barcode units by CyTOF using simple-barcoded cells.

A) *Schematic of the X-CODE CyTOF protocol for detection of barcode units (BUs) compared to mass cytometry detection of isotope-conjugated antibodies. Simple-barcoded TRAMP-C1 lines harbouring barcode A (BU1+BU2), barcode B (BU2+BU3), and barcode C (BU3+BU4) were mixed, stained by the X-CODE CyTOF protocol, and acquired by Helios mass cytometer. Created with BioRender.com.*

B) *Bivariate density spots showing signal from distinct isotope-conjugated probes targeting BU1 – BU4 sequences in TRAMP-C1 A, B, C mixed population. Unbarcoded wild-type (WT) TRAMP-C1 cells lack BU signal.*

In summary, these results demonstrate that X-CODE barcode units can be robustly and specifically detected by CyTOF at single-cell resolution, enabling subsequent determination of clonal identity using small-scale barcodes in suspension cells.

5.3 Assessment of X-CODE co-detection compatibility with antibody staining

The compatibility of X-CODE barcode co-detection with conventional CyTOF immunophenotyping workflows was then evaluated. TRAMP-C1 A, B, and C cells were expanded in culture, mixed in equal proportions, and processed using the X-CODE CyTOF protocol with minor adaptations to accommodate extra- and intracellular

antibody staining. In accordance with standard CyTOF practice [137], extracellular markers were stained first, followed by intracellular and phospho-specific targets after fixation and permeabilisation. A preliminary antibody panel was selected to span the surface and intracellular compartments and included extracellular CD24 and CD49f, intracellular Ki67, and the phospho-specific proteins pRB1, pERK, and p4EBP1. In parallel, cells were stained with isotope-conjugated X-CODE detection probes targeting barcode units (BUs) 1–4.

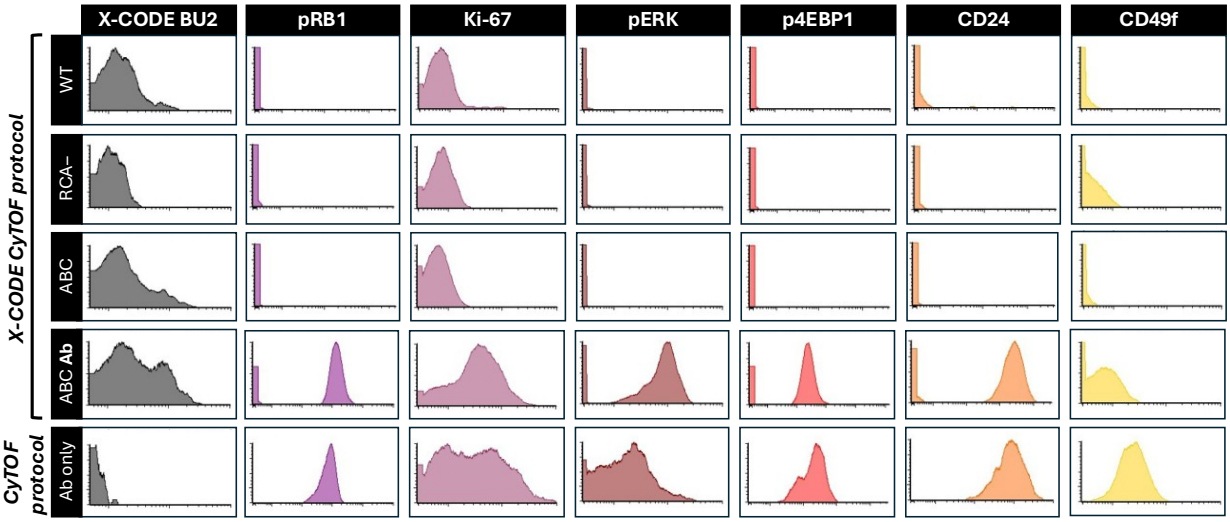


Figure 5-3. Validation of X-CODE detection compatibility with immune-phenotyping by CyTOF.

TRAMP-C1 A, B, and C cells and unbarcoded wild-type (WT) cells were processed using the X-CODE CyTOF protocol to detect RNA barcodes by staining for barcode units (BUs) 1 – 4 (ABC). A rolling circle amplification-negative (RCA–) sample was also included, in which φ29 DNA polymerase is omitted. Cells also underwent a variant of the X-CODE CyTOF protocol that incorporated simultaneous antibody staining (ABC Ab). TRAMP-C1 ABC cells underwent the standard CyTOF protocol using the same antibodies for comparison (Ab only), these included: pRB1, Ki67, pERK, p4EBP1, CD24, and CD49f. Histograms showing signal intensity distributions for the indicated markers under the different processing conditions. The first panel represents a randomly selected BU target (BU-2); all remaining panels display antibody-derived signals.

Comparison of signal intensity distributions showed that antibody-derived signals were broadly comparable whether acquired using the standard CyTOF immunostaining protocol or incorporated into the X-CODE workflow (**Figure 5-3**). Although some variation in signal intensity was observed between conditions for individual markers, overall antibody detection was maintained. Incorporation of antibody staining likewise did not adversely affect X-CODE barcode detection, as exemplified by BU2 signal intensity (**Figure 5-3**). Barcode signal depended on the rolling circle amplification step, with no detectable signal in samples processed without φ29 polymerase (RCA–) and none in wild-type (WT) control cells.

These data demonstrate that X-CODE barcode readout is compatible with antibody staining and standard CyTOF acquisition conditions, enabling concurrent detection of barcode units and protein markers within the same experiment.

5.4 Validation of high-complexity X-CODE detection by CyTOF

5.4.1 Confirmation of isotope conjugation to X-CODE detection probes

To enable detection of high-complexity X-CODE barcodes in the 15K- and 2M-plex libraries, detection probes targeting barcode units (BUs) 1–27 were conjugated individually to unique metal isotopes. The PLAYR assay, as previously described, uses oligonucleotide–isotope coupling with a size-exclusion spin-filter–based conjugation and purification workflow, repurposing reagents and equipment distributed by Standard BioTools for custom antibody conjugation in CyTOF workflows [137] . Using this strategy, isotope assignments were selected for X-CODE detection probes, and custom probe panels were generated for the 15K- and 2M-plex libraries, targeting BUs 1–18 and BUs 1–27, respectively (**Supplementary Table 8** and **Supplementary Table 9**). In addition, detection probes targeting β-actin RNA were conjugated to metal isotopes for use as a rolling-circle amplification control.

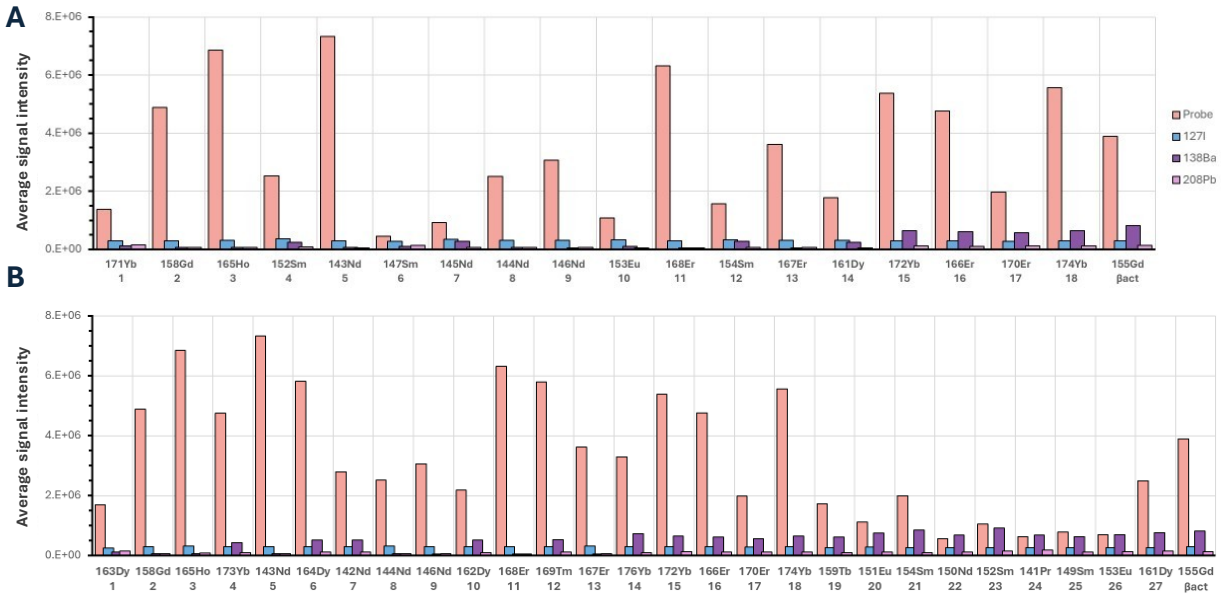


Figure 5-4. Assessment of isotope-conjugation to X-CODE detection probes.

*Isotope-tagged probes complementary to barcode units (BUs) for detection of the X-CODE barcode libraries were acquired without cellular staining to assess successful isotope conjugation. Isotope detection probes were diluted to 100 pM and acquired in solution-mode on a Helios mass cytometer (Standard Biotools) for 60 seconds. Bar graphs showing the average signal intensity per detection probe over the acquisition period for probes detecting BUs 1 – 18 in the 15K library (**A**), and for probes detecting BUs 1 – 27 in the 2M library (**B**). BU probe tags are given by their*

element and isotope. The average signal intensity for common background elements, iodine (^{127}I), barium (^{138}Ba), and lead (^{208}Pb), is also shown.

To verify successful isotope conjugation prior to staining X-CODE library cells, newly conjugated detection probes were analysed by CyTOF in solution mode to assess isotope identity and relative signal intensity independently of cellular background. Probes were diluted to 100 pM in double-distilled water, and signal intensity was acquired for 60 seconds per probe. All probes exhibited detectable signal in their expected isotope channels, with signal levels mostly exceeding intensities observed in standard background isotopes (^{127}I, ^{138}Ba, and ^{208}Pb) (**Figure 5-4**). Average signal intensity varied across probes, reflecting intrinsic differences in isotope sensitivity, characteristic of CyTOF detection [142].

These data confirmed accurate isotope conjugation and supported their subsequent use for staining cells carrying 15K- and 2M-plex X-CODE barcodes.

5.4.2 Detection of the X-CODE 15K library by barcode units 1–18

To assess CyTOF-based detection of barcode units (BUs) 1–18, clonal TRAMP-C1 cell lines expressing defined 15K X-CODE barcodes were analysed. Three clonal lines, denoted F, H, and I, were isolated from a TRAMP-C1 population infected with the 15K X-CODE library. Each clone expresses a distinct barcode composed of a subset of BUs from the 15K library, and together the three clones span the full BU1–18 sequences. These defined barcodes provided a controlled system for evaluating the detection of individual BUs with known ground truth. TRAMP-C1 F, H, and I cells were stained using the X-CODE CyTOF protocol with detection probes targeting BUs 1–18 and acquired by CyTOF (**Figure 5-5**A).

Stained clonal samples were acquired alongside a rolling circle amplification–negative (RCA–) sample and TRAMP-C1 wild-type (WT) control cells. The barcode unit-associated signal was absent in both RCA– and WT controls, confirming the dependence on the amplification step and the specificity of barcode detection. Concatenation of acquisition files from clonal lines F, H, and I enabled direct comparison of signal distributions across all BUs (**Figure 5-5**B). In the merged dataset, both positive and negative populations were observed for each BU, consistent with the absence of specific barcode sequences in individual clonal barcodes. Dimensionality reduction

using t-SNE on the concatenated data resolved the three clonal populations into distinct clusters (**Figure 5-5**B). Projection of individual BU signals onto the t-SNE embedding revealed clone-specific combinatorial patterns, with each cluster defined by its expected 8-of-18 X-CODE sequences composition. For example, the blue island shown in **Figure 5-5**A represents TRAMP-C1 F cells.

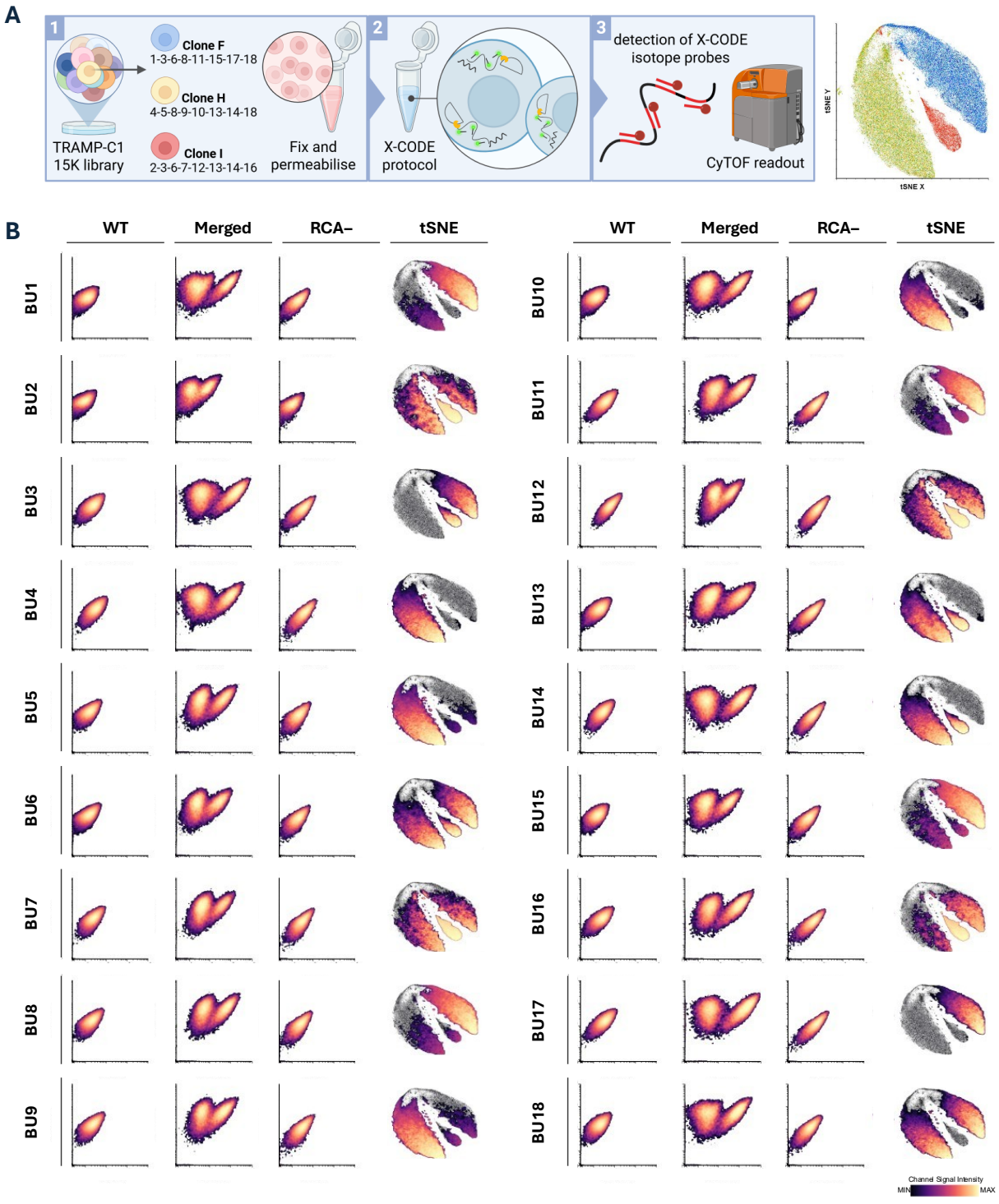


Figure 5-5. Detection of barcode units 1–18 in 15K X-CODE library barcoded cell lines.

A) *Schematic representation of the X-CODE CyTOF protocol for detection of barcode units (BUs) in three barcoded TRAMP-C1 clones F, H, and I, which collectively represent the complete set of 18 BUs in the 15K library, known barcodes as depicted. Cells were stained for BUs 1–18 using the X-CODE CyTOF protocol, and the three clones were acquired separately on a Helios mass cytometer (Standard Biotools). The resulting CyTOF datasets were merged and visualised using t-SNE, revealing three distinct clusters corresponding to the three clones, each coloured by clone. Blue = F cells, yellow = H cells, red = I cells. Created with BioRender.com.*

B) *Density plots and t-SNE visualisations summarising BU-specific signal patterns for each of the 18 BUs in the 15K library. For every BU, panels show: density plot of signal from wild-type (WT) cells, density plot from (merged) FHI clone data, density plot from control samples processed without the RCA step (RCA–), and a t-SNE of merged data colour-coded for BU signal intensity.*

To extend this analysis beyond clonal cell lines, a heterogeneous TRAMP-C1 population harbouring the 15K X-CODE library was analysed. This unbiased library population was stained for BUs 1–18 using the X-CODE CyTOF protocol and acquired by CyTOF. Across the library, both positive and negative populations were observed for each BU, consistent with the expectation that individual cells express only a subset of BUs (**Figure 5-6**). The separation between positive and negative populations, as well as the observed signal intensity, varied across detected barcode sequences. Such variation likely reflects a combination of barcode sequence–dependent hybridisation efficiency and intrinsic differences in CyTOF isotope detection sensitivity, as observed in cell-independent measurements of isotope-oligonucleotide conjugations.

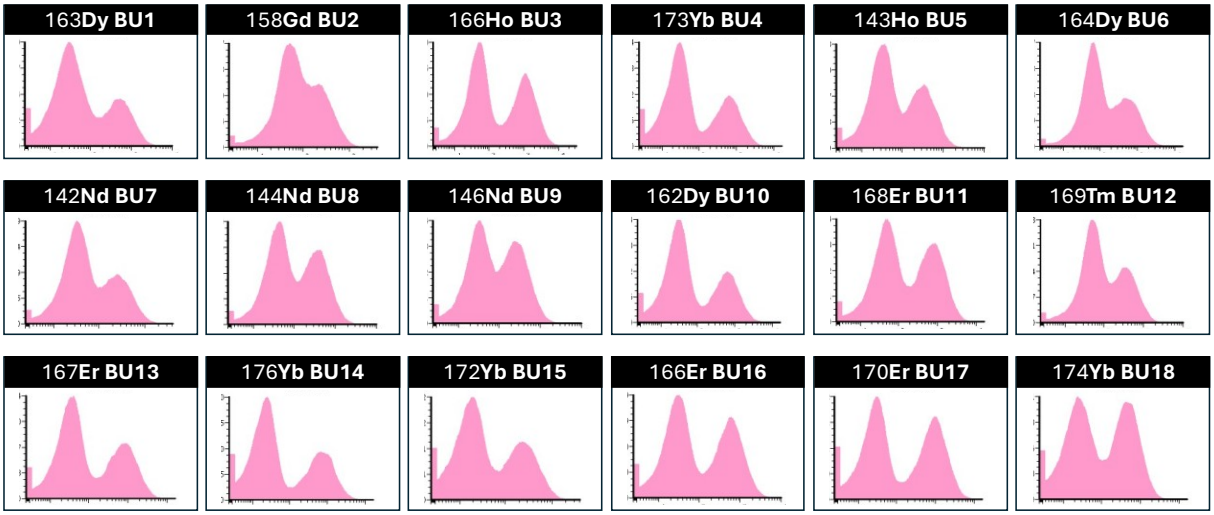


Figure 5-6. Signal intensities for barcode units 1–18 in a heterogeneous barcoded TRAMP-C1 15K population.

Histograms summarising signal intensity for barcode unit (BU)-specific positive and negative populations within a heterogeneous TRAMP-C1 population barcoded with the X-CODE 15K library. Cells were stained for BUs 1–18 using the X-CODE CyTOF protocol and acquired by a Standard Biotools Helios mass cytometer.

Together, these data demonstrate that barcode units 1–18 from the 15K X-CODE library can be specifically and robustly detected by CyTOF in both clonal control lines and heterogeneous library populations.

5.4.3 Detection of the X-CODE 2M library lines by barcode units 1–27

Following validation of barcode unit detection in the 15K X-CODE library, barcode detection was extended to cover the full design space of the 2M X-CODE library. This analysis assessed whether higher-complexity combinatorial barcodes spanning all 27 barcode units could be resolved in a pooled, single-run CyTOF acquisition. Four RM-1 clonal cell lines derived from the X-CODE 2M library, denoted J, W, Y, and Z, were selected for analysis because their defined barcode identities collectively span barcode units (BUs) 1–27. The four clonal populations were mixed in equal proportions and stained with isotope-conjugated detection probes targeting BUs 1–27 (**Supplementary Table 9**) according to the X-CODE CyTOF protocol, then acquired simultaneously by CyTOF (**Figure 5-7**A).

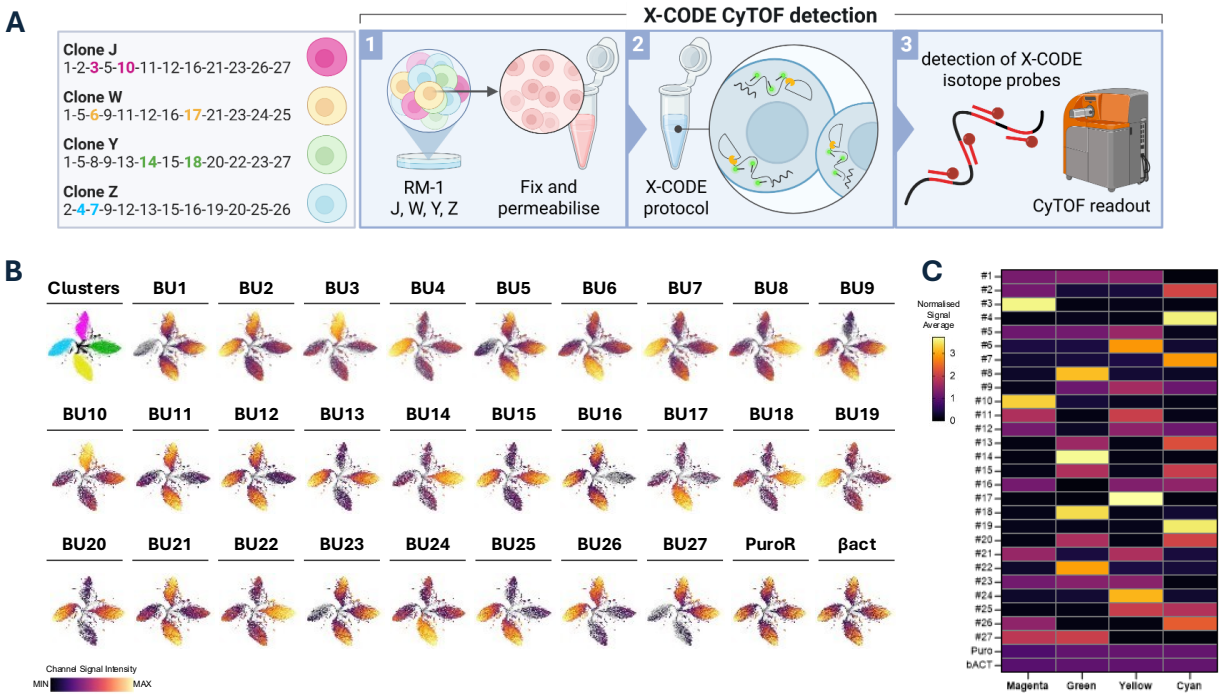


Figure 5-7. Barcode unit detection across known barcode populations from the 2M X-CODE library.

A) *Schematic representation of the X-CODE CyTOF protocol for detection of barcode units (BUs) in four barcoded RM-1 clones J, W, Y, Z, which collectively represent the complete set of 27 BUs in the 2M library, known barcodes as depicted. Cells were stained for BUs 1–27 using the X-CODE CyTOF protocol, and the four clones were mixed in equal proportions and acquired on a Helios mass cytometer (Standard Biotools). Created with BioRender.com.*

B) *CyTOF JWYZ mix acquisition data were analysed and represented as a t-SNE visualisation with cluster-specific heatmaps. t-SNE analysis revealed four distinct cell clusters: magenta (top, J cells), cyan (left, Z cells), yellow (bottom, W cells), and green (right, J cells). Each panel displays the same t-SNE, with the colour scale representing the signal intensity of a specific BU detection probe across the clusters.*

C) *Heatmap displaying normalised average signal intensity for each specific BU detection probe across the four clusters.*

Dimensionality reduction of the mixed RM-1 J, W, Y, and Z population using t-SNE resolved four distinct clusters corresponding to the four clonal populations (**Figure 5-7**B). Projection of individual BU signals onto the t-SNE embedding revealed clone-specific combinatorial patterns, with each cluster defined by the expected set of 12 of 27 X-CODE sequences. For example, the pink cluster corresponds to clone J cells, characterised by signal enrichment for the 12 barcode units comprising the J barcode. A heatmap of normalised average signal intensity for each detection probe across the four clusters further showed that the highest signal intensities were associated with BUs expressed by the corresponding clonal populations (**Figure 5-7**C).

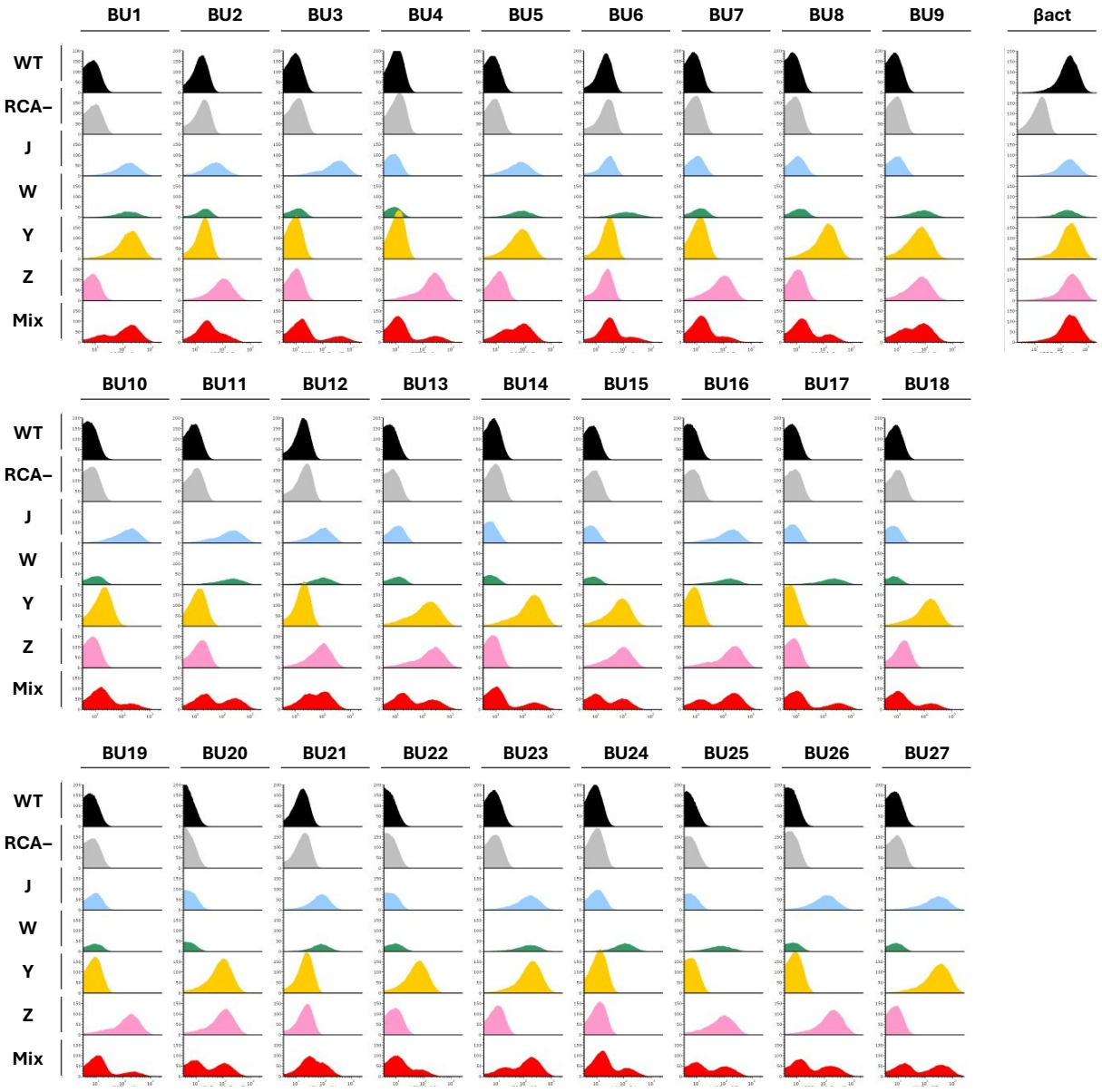


Figure 5-8. Detection of barcode units 1–27 across X-CODE 2M library barcoded cell lines.

Histograms summarising signal intensity for barcode unit (BU)-specific positive and negative populations within RM-1 populations barcoded with known X-CODE 2M library barcodes (J, W, Y, and Z). Single lines, an equally mixed population, and controls omitting the φ29 polymerase (RCA–) and RM-1 unbarcoded wild-type (WT) samples were

stained for BUs 1–27 using the X-CODE CyTOF protocol. Samples were acquired by a Standard Biotools Helios mass cytometer.

To further assess BU detection specificity and efficiency, separate J, W, Y, and Z populations were processed and acquired under identical conditions. Along with rolling circle amplification–negative (RCA–) controls, and RM-1 wild-type (WT) cells. Barcode-associated signal was detected in clonal and mixed barcoded populations but was absent in both RCA– and WT controls, confirming amplification dependence and detection specificity (**Figure 5-8**). In contrast, β-actin RNA signal was consistently detected across RM-1 WT cells and barcoded populations, with no signal observed in RCA– controls. Within the mixed population, signal distributions for individual BUs were separated into positive and negative populations, as hypothesised, because each cell expresses only a subset of X-CODE sequences. As observed in the 15K library analysis, variation in signal intensity was evident across BUs, likely reflecting a combination of sequence-dependent hybridisation efficiency and intrinsic isotope-dependent sensitivity in CyTOF detection.

Collectively, these data demonstrate that barcode units 1–27 can be specifically and concurrently detected by CyTOF within the 2M X-CODE library, enabling resolution of high-complexity combinatorial barcodes in pooled cell populations.

5.5 Discussion

This chapter establishes X-CODE as a barcode system compatible with sequencing-free readout by suspension mass cytometry for single-cell analysis. Using CyTOF-based detection, barcode units were robustly and specifically detected at single-cell resolution, enabling determination of clonal identity within suspension cell populations. Initial experiments demonstrated reliable detection of small-scale barcodes in mixed populations, providing a foundation for evaluating barcode readout using CyTOF. Successful isotope-conjugation of X-CODE detection probes enabled the extension of barcode detection to library-scale using high-parameter panels, supporting analysis of both 15K- and 2M-X-CODE libraries.

Across these experiments, barcode units were detected in both clonal control lines and heterogeneous populations, with combinatorial barcode identities resolved in pooled, single-run CyTOF acquisitions. Moreover, integration of X-CODE detection with

antibody-based CyTOF immunophenotyping was demonstrated, showing that barcode readout is compatible with standard staining workflows and acquisition conditions. Antibody-derived signals were preserved when incorporated into the X-CODE protocol, and barcode unit detection remained efficient. Thus, the X-CODE framework enables scalable identification of clonal populations and simultaneous immune phenotyping using mass cytometry.

Mass cytometry has emerged as a fundamental technique for high-dimensional immunophenotyping and signalling analysis, enabling the analysis of dozens of molecular features per cell across millions of cells in a single experiment. Its high throughput and multiplexing capacity enable detailed interrogation of heterogeneous populations and facilitate comparison across many experimental conditions [142] . Probe-based RNA detection workflows such as PLAYR have further extended CyTOF beyond protein-centric measurements, providing a foundation for multimodal single-cell analysis [137,141] . Building on these advances, the work presented here adapts probe-based CyTOF detection to enable readout of engineered, combinatorial RNA barcodes, supporting lineage-aware analysis of heterogeneous suspension cell populations at scale.

By enabling sequencing-free, clonal-level barcode detection alongside simultaneous immunophenotyping, X-CODE extends the utility of mass cytometry for investigating clonal heterogeneity at scale. This capability is particularly relevant for applications requiring analysis of large populations and rare cell states, such as cancer, where phenotypic diversity and clonal structure underpin disease progression and treatment response [5] . In particular, this analytical context is well-suited to investigating liquid malignancies, which have historically been studied using flow-based approaches. Given the high parameter capacity enabled by emerging spectral flow cytometry platforms, the principles described here may also be extensible beyond mass cytometry [138] . More broadly, high-throughput, sequencing-independent clonal analysis provides a complementary approach to existing lineage-tracing strategies, enabling interrogation of clonal dynamics across millions of cells while retaining direct linkage to single-cell phenotype.

5.5.1 Limitations

While this chapter demonstrates compatibility between X-CODE barcode detection and CyTOF-based antibody staining, with selected markers spanning surface, intracellular, and phospho-specific targets, only a limited antibody panel was evaluated. As noted in the established CyTOF-based RNA and protein co-detection workflow, PLAYR, panel-specific optimisation may be required to ensure robust performance of antibody markers [137]. Accordingly, future applications of X-CODE in mass cytometry would involve systematic optimisation of expanded antibody panels tailored to specific biological questions, consistent with standard CyTOF best practices [142]. An additional consideration is that detection of high-complexity X-CODE libraries can occupy up to 27 CyTOF measurement channels, necessitating careful selection of antibody markers within the remaining parameter space.

Although the full 15K X-CODE library was analysed experimentally in this chapter, computational debarcoding of library-scale barcode identities was not presented. Resolution of combinatorial measurements at this scale typically requires dedicated computational pipelines capable of classifying cells based on multi-dimensional signal patterns. Published packages designed for decoding multiplexed sample-level barcodes, such as the R package Premessa [192], may be repurposed for X-CODE deconvolution. Pipelines built on such work have been developed and applied to X-CODE data by computational scientists. Nonetheless, detailed implementation and optimisation of debarcoding algorithms are beyond the scope of this chapter, which focuses on experimental validation of barcode detection by CyTOF. As a result, this work does not address computational debarcoding in detail; instead, it establishes the experimental feasibility required to generate data for downstream computational analysis.

A further limitation relates to variability in detection probe performance across barcode units. Differences in signal intensity were observed between barcode units, reflecting both sequence-dependent hybridisation efficiency and intrinsic isotope-dependent sensitivity in CyTOF detection. While all barcode units tested in this work were detectable, this variability necessitates consideration in downstream analysis. In practice, debarcoding pipelines can be designed to accommodate signal-specific differences via adaptive thresholding or weighting strategies, thereby mitigating the

effects of unequal measurement distributions. Nonetheless, probe-level variability remains an important technical consideration when extending X-CODE detection to increasingly complex barcode libraries.

5.5.2 Conclusions

In conclusion, this chapter validated isotope-based detection of modular X-CODE barcode units by CyTOF and established compatibility between RNA barcode detection and antibody-based phenotypic staining workflows. Detection was extended from defined, low-complexity barcodes to high-complexity 15K- and 2M-X-CODE libraries, enabling the resolution of combinatorial barcode identities in pooled cell populations. Together, this work demonstrates that the X-CODE framework supports high-throughput, combinatorial barcode detection by suspension mass cytometry across multiple library scales while remaining compatible with simultaneous, high-dimensional protein phenotyping. This establishes a sequencing-independent strategy for clonal analysis that is directly compatible with mass cytometry workflows, thereby extending barcode-based lineage tracing to experimental settings traditionally centred on flow and suspension-based phenotyping, including applications in liquid malignancies.

Chapter 6 High-parameter spatial detection of X-CODE barcodes by multiplexed imaging

6.1 Research in context

Highly multiplexed imaging platforms offer a valuable opportunity to detect expressed barcodes *in situ* while retaining capacity for further sample phenotyping. Such systems include the Akoya PhenoCycler-Fusion, which was developed for high-plex protein detection at subcellular resolution through automated cyclic fluorescence imaging of antibody staining. Recently, it has been repurposed for RNA detection using the RNAscope assay in a combined workflow with protein staining [151], demonstrating that nucleic acid targets can be integrated into PhenoCycler imaging pipelines.

Additionally, imaging mass cytometry (IMC) enables high-parameter protein detection via isotope-tagged antibodies, allowing simultaneous measurement of dozens of targets at subcellular resolution. Furthermore, IMC-based co-detection of RNA and proteins has been demonstrated in human tissue [150]. Although high-complexity RNA barcode libraries have not yet been evaluated on these platforms, their ability to image multiple parameters in a single experiment makes them well-suited for *in situ* barcode detection. Consequently, both technologies offer viable routes for spatial barcode readout while maintaining compatibility with robust immunophenotyping workflows.

The success of *in situ* barcode detection in intact tissues is critically influenced by the sample-preservation method and its effects on RNA integrity. Formalin-fixed, paraffin-embedded (FFPE) tissue is the clinical standard for preserving morphology and protein markers, but formaldehyde cross-linking and paraffin infiltration introduce fragmentation and chemical modifications that can reduce RNA integrity [195] . Across multiple studies, RNA extracted from FFPE tissue exhibits significantly lower RNA integrity numbers (RIN) and greater fragmentation compared to fresh-frozen samples [196,197], reflecting substantial degradation and reduced transcript accessibility for hybridisation-based assays.

Although *in situ* hybridisation can detect transcripts in degraded samples, hybridisation signal intensity and transcript counts are often reduced relative to frozen tissue,

reflecting differences in RNA integrity and accessibility [182]. Fresh-frozen tissue better preserves RNA integrity and is often preferred for spatial transcriptomics applications due to high RIN scores and reduced molecular damage [198]. Embedding fresh-frozen samples in optimal cutting temperature (OCT) compound facilitates cryosectioning for several spatial applications.

The advantages of incorporating mass spectrometry imaging (MSI) into multi-modal spatial workflows for unbiased mapping of metabolites, peptides, and lipids are increasingly recognised [104,106]. However, OCT-embedded tissue is considered unsuitable for MSI, such as matrix-assisted laser desorption ionisation (MALDI) [199,200]. Therefore, efforts to develop multimodal-compatible embedding compounds have increased. A hydrogel composed of hydroxypropyl methylcellulose (HPMC) and polyvinyl-pyrrolidone (PVP) has been designed to maintain sample structural integrity while enhancing compatibility with several molecular techniques, including MSI [164,165]. This offers a preservation strategy that balances morphology with molecular accessibility across imaging and MSI workflows for barcoded tissue samples, as explored further in Chapter 7.

Taken together, these technological developments and sample considerations establish a rationale for optimising workflows to resolve X-CODE barcodes spatially using high-parameter detection platforms. This chapter focuses on optimising padlock probe hybridisation conditions, expanding fluorescence detection capacity, and adapting X-CODE detection for imaging mass cytometry. Furthermore, barcode detection is assessed across OCT, HPMC/PVP hydrogel, and FFPE tissue preparations to enable robust, broadly applicable spatial barcode readout.

6.1.1 Chapter aims

- To evaluate the spatial detection of X-CODE barcodes using highly multiplexed fluorescence imaging platforms.
- To adapt X-CODE barcode detection for high-parameter imaging mass cytometry.
- To validate robust X-CODE detection in intact samples across multiple tissue preparation and embedding methods.

6.2 Optimisation of hybridisation probe conditions for spatial X-CODE detection

Efficient, specific probe-based spatial detection of four modular X-CODE units using confocal microscopy was demonstrated in Chapter 4. To successfully utilise and decode the X-CODE 15K and 2M libraries *in situ*, efficient detection of all 18 or 27 sequences, respectively, is required. The X-CODE barcode unit (BU) comprises a variable binding site (VBS) that can adopt the identity of any of the 18 or 27 sequences, followed by a common universal binding site (UBS). Detection of the BUs is achieved by binding a padlock probe specific to the VBS sequence and a universal probe that recognises the UBS across all modular units (**Figure 6-1**A). To ensure the simultaneous, efficacious detection of a larger number of barcode units, padlock and universal probe concentrations were optimised using fluorescence imaging.

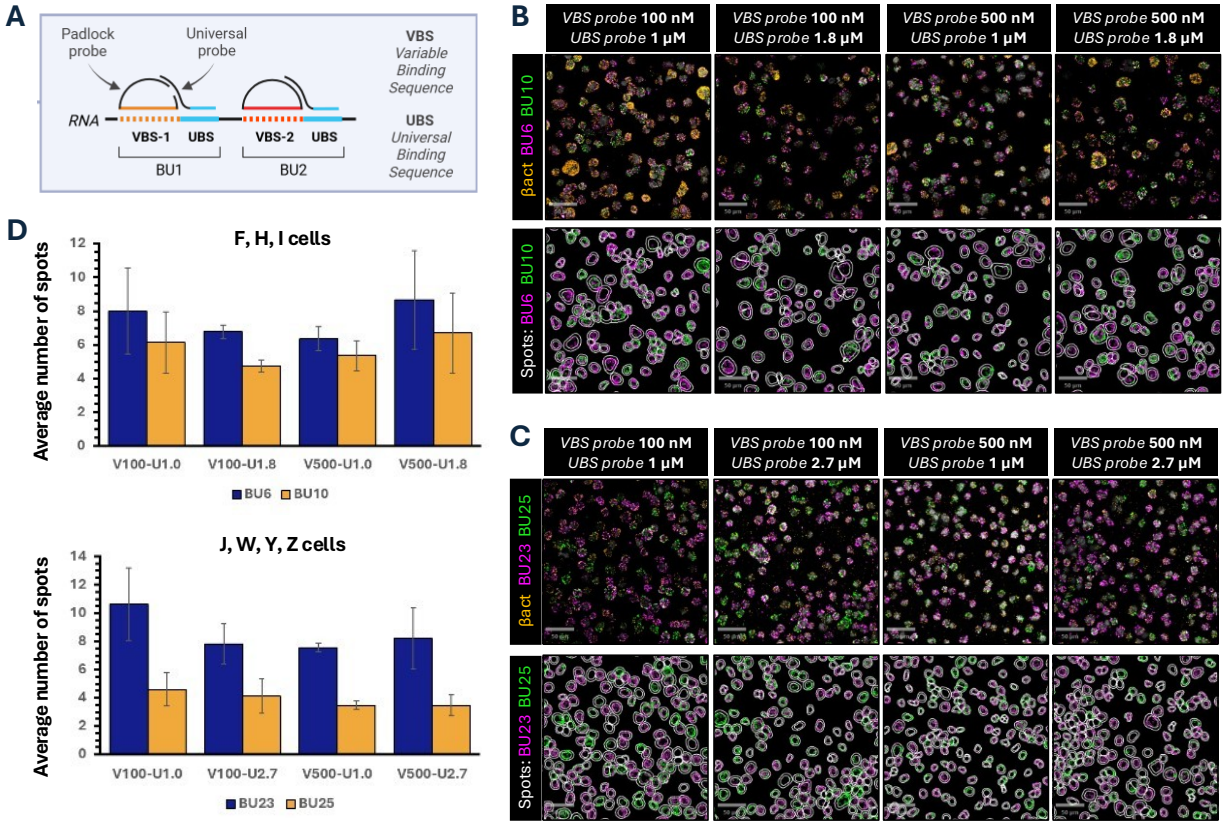


Figure 6-1. Optimisation of padlock probe and universal probe concentrations for spatial X-CODE readout.

A) *Schematic representation of the X-CODE barcode unit (BU) comprising the variable binding sequence (VBS), which can adopt X-CODE sequences 1 – 18 (15K library), or 1 – 27 (2M library) for the padlock probe to bind. The universal binding sequence (UBS) allows the universal probe to bind for downstream ligation, followed by rolling-circle amplification of the VBS. Created with BioRender.com.*

B) *Equally mixed TRAMP-C1 15K clonal lines (F, H, I) were adhered to microscope slides using a modified STAMP protocol, then underwent X-CODE detection. X-CODE padlock probes 1–18 were hybridised at a final concentration of 100 or 500 nM, and the universal probe at 1 or 1.8 µM. Fluorescence imaging readout visualised BU6 (pink) and*

BU10 (green) across hybridisation conditions, with DAPI staining visualising nuclei (white) and β-actin RNA staining as a protocol control (orange). Scale bars, 50 µm.

C) *Equally mixed RM-1 2M clonal lines (J, W, Y, Z) were adhered with the modified STAMP protocol, then underwent X-CODE detection. X-CODE padlock probes 1–27 were hybridised at a final concentration of 100 or 500 nM, and the universal probe at 1 or 2.7 µM. Fluorescence imaging readout visualised BU23 (pink) and BU25 (green) across hybridisation conditions, with DAPI staining visualising nuclei (white) and β-actin RNA staining as a protocol control (orange). Scale bars, 50 µm.*

D) *The number of spots per positive cells (spots > 0) detected across probe hybridisation concentrations for BU6, BU10, BU23, and BU25 using an ImageJ-based spot-counting pipeline. V = VBS probe at the indicated concentration in nM; U = UBS probe at the indicated concentration in µM. Error bars indicate standard deviation across ROIs (n = 3).*

Three clonal TRAMP-C1-derived 15K populations (F, H, I) and four clonal RM-1-derived 2M lines (J, W, Y, Z), previously used for CyTOF analysis, were adhered to microscope slides using a modified STAMP protocol [167]. The X-CODE protocol for plated cells was performed, systematically altering the concentrations of the padlock and universal probes. The effect of increasing BU complexity on universal probe demand was evaluated by adjusting hybridisation conditions previously optimised for the simultaneous detection of four sequences. This involved each padlock probe at a final concentration of 100 nM and the universal probe at 1 µM. Detection of the 15K or 2M libraries requires binding 18- or 27-padlock probes and a universal probe to the barcode transcript at each BU position; thus, the concentration of the universal probe was increased to 1.8 µM or 2.7 µM, respectively. Furthermore, a 500 nM padlock probe concentration was tested to assess its impact on X-CODE detection performance.

Fluorescent detection probes targeting BU6 and BU10 were hybridised to 15K barcode lines, while probes targeting BU23 and BU25 were applied to 2M barcode lines. All fluorescent detection probes were used at a final concentration of 500 nM (**Figure 6-1**B and C). β-actin RNA padlock-based detection was performed as a protocol control. The effect of altering universal and padlock probe concentrations was assessed using an ImageJ-based spot-counting pipeline to determine the number of BU spots per cell across conditions (**Figure 6-1**D). Across the range of universal probe concentrations tested, the mean number of BU spots detected per cell remained largely unchanged across X-CODE 15K and 2M clonal lines, suggesting that the availability of the universal probe at 1 µM is not a limiting factor for high-plex X-CODE barcode detection for the tested sequences. Similarly, increasing the padlock probe concentration did not significantly increase the mean number of barcode unit spots detected per cell for the tested sequences.

Together, these results support the use of the initially optimised, lowest concentrations of padlock and universal probe tested for further X-CODE detection, without compromising detection efficiency.

6.3 Expansion of fluorescence X-CODE detection capacity using PhenoCycler

To extend high-dimensional, sequencing-free X-CODE barcode detection to spatially resolved imaging, the Akoya PhenoCycler-Fusion was repurposed for direct RNA barcode readout. By default, the PhenoCycler is configured for high-parameter visualisation of fluorescence-based antibody staining, built on the CO-Detection by indEXing (CODEX) framework developed by *Garry Nolan* and colleagues [201,202]. Antibodies are conjugated to short, unique oligonucleotides to enable hybridisation with complementary fluorescent probes (reporters) purchased from Akoya. In practice, the tissue section is stained with the full antibody panel, followed by several fixation steps to ensure staining stability on the instrument. Then, the PhenoCycler performs automated, iterative cycles of fluorescent reporter hybridisation, imaging, and removal before the next cycle. The PhenoCycler can distinguish three unique fluorophores per imaging round; therefore, each cycle images antibodies in triplicate.

To enable accurate decoding of antibody identity across cycles, a pre-designed reporter plate is loaded into the instrument. The positional layout of fluorescent reporters in this 96-well plate is specified in the experimental run file, enabling the system to assign each signal within an imaging cycle to its corresponding antibody during image reconstruction. For X-CODE adaptation, fluorescent X-CODE detection probes targeting BUs 1–27 (**Supplementary Table 10**) were added directly to the reporter plate. Thereby being processed using the unmodified Akoya PhenoCycler-Fusion fluidics, cycling, and imaging workflow, without altering instrument hardware or acquisition parameters (**Figure 6-2**A). Furthermore, X-CODE detection probes and Akoya Reporters can be used within the same experimental run and even within imaging cycles, provided that three distinct fluorophores are applied per cycle, consistent with the instrument's capacity.

The four RM-1 X-CODE 2M populations (J, W, Y, Z) were used to assess the compatibility of RNA barcode detection on the Akoya PhenoCycler-Fusion. Each line

carries a unique barcode encoded by a combination of 12-of-27 BUs, visualised in this defined mixture with two distinguishing barcode units per clone (**Figure 6-2**A, top left). Equal mixtures or single populations of J, W, Y, and Z were seeded on a multi-chamber microscope slide, then the X-CODE protocol for plated cells was performed with appropriate RCA-negative (RCA–) and wild-type (WT) controls. Visual assessment of X-CODE detection showed that BU signals were robust and confined to the expected line in single-clone samples, and analysis of the mixed population enabled accurate spatial resolution of all four clones (**Figure 6-2**B). Barcode signal was minimal in WT RM-1 cells and in RCA– controls, and X-CODE detection was compatible with simultaneous antibody staining, as exemplified by co-detection of sodium–potassium ATPase (Na/K-ATPase).

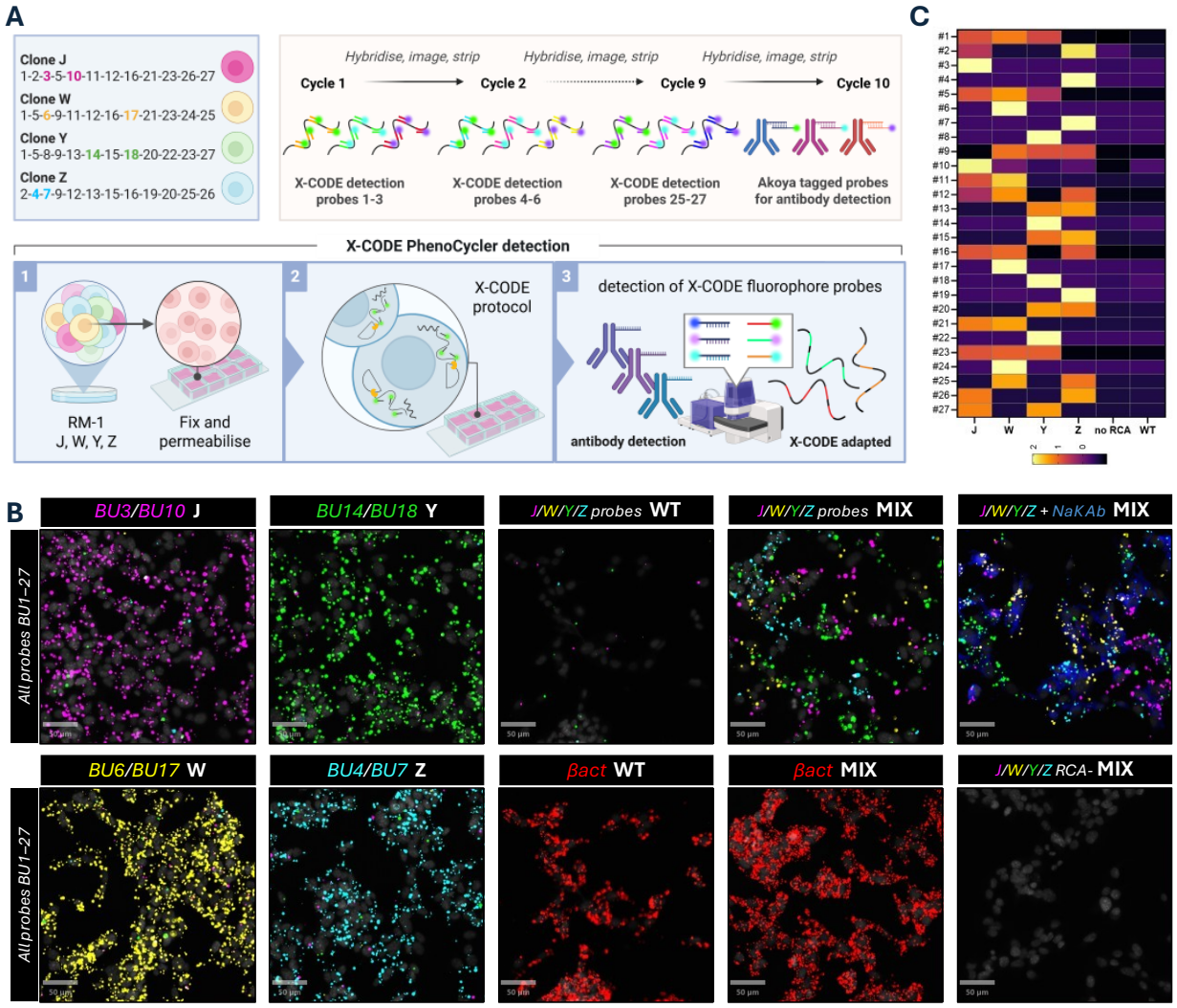


Figure 6-2. Validation of high-plex fluorescence detection of X-CODE barcode units using Akoya PhenoCycler.

A) *The Akoya PhenoCycler-Fusion, designed for high-plex protein staining via automated cyclic hybridisation of fluorescent antibody-matched oligos, was repurposed for X-CODE barcode unit (BU) detection. X-CODE fluorescent detection probes were directly processed by the unmodified PhenoCycler hardware which performs automated*

hybridisation of probes for imaging, followed by stripping and rehybridization of the next batch of probes in cycles. RM-1 2M clonal J, W, Y, and Z lines were mixed equally, or plated as single populations, stained with for X-CODE BUs 1–27 and readout by Akoya PhenoCycler, along with RM-1 WT control cells. Created with BioRender.com.

***B)** Representative RM-1 2M clonal populations are shown, visualised by distinguishing barcode unit (BU) sequences: J (pink, BU3&10); W (yellow, BU6&17); Y (green, BU14&18); and Z (cyan, BU4&7). DAPI staining for nuclei (white) and β-actin RNA staining as a protocol control (red), and no rolling circle amplification (RCA) control (RCA–). Co-detection of sodium–potassium (Na/K) ATPase antibody (NaK Ab, blue). Scale bars, 50 μM.*

***C)** The number of BU spots per segmented cell and their signal intensity were quantified using a QuPath-based spot-counting pipeline. Heatmap showing Z-score (normalised spot intensity per cell) across conditions: RM-1 barcoded J, W, Y, and Z cells plated separately, also a rolling circle negative (RCA–) and unbarcoded, wild-type cells.*

The efficiency of X-CODE detection on the PhenoCycler-Fusion platform was assessed using a QuPath-based spot-counting pipeline to quantify the number of BU spots visualised by fluorescence probes and the mean fluorescent intensity per cell. Per-probe spot counts across conditions further demonstrated specificity and low background (**Figure 6-3**). As observed for suspension-based X-CODE detection using mass cytometry**,** barcode sequence detection showed variability across BUs, as evidenced by differences in the mean number of detected spots and per-cell fluorescence intensity. For example, probe #23, present in clones J, W, and Y, demonstrated ≥100 detected spots per cell, whereas probe #6, also present in clones J, W, and Y, exhibited approximately 30–40 spots per cell (**Figure 6-3**). Despite this variability, probe-specific signals remained readily detectable across all BUs using consistent spot-detection thresholds, enabling reliable binary assignment of BU presence at the single-cell level. Quantification of Z-score (normalised spot intensity per cell) confirmed clear BU-specific signal patterns across samples (**Figure 6-2**C).

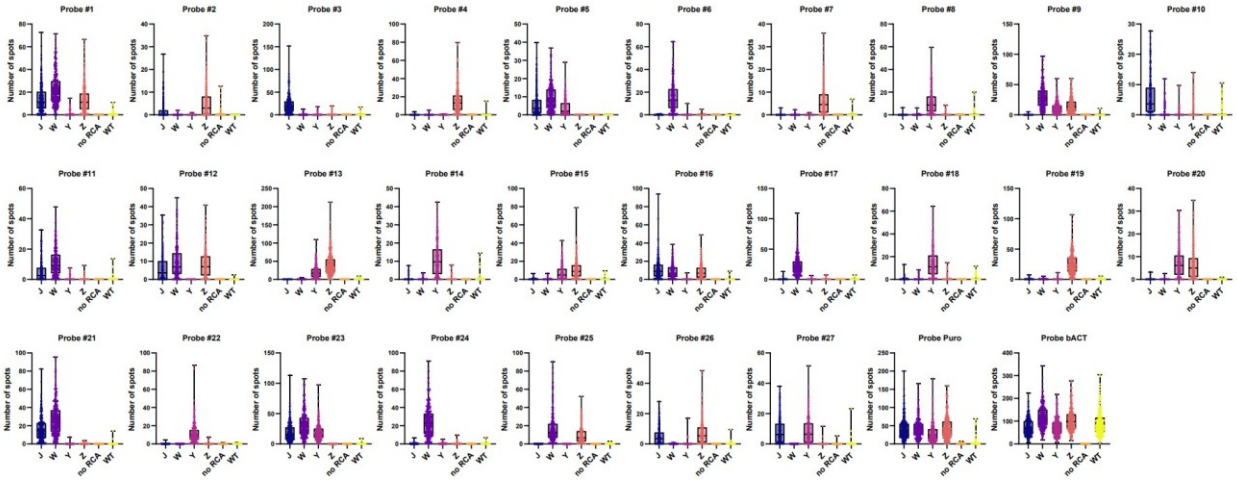


Figure 6-3. Highly multiplexed fluorescence detection of barcode units 1-27 by Akoya PhenoCycler-Fusion.

The number of BU spots per segmented cell was quantified using a QuPath-based spot-counting pipeline. Box plots showing BU spot counts per BU probe across conditions: RM-1 barcoded J, W, Y, and Z cells plated separately, also a rolling circle negative (RCA–) and unbarcoded, wild-type cells.

Collectively, these results demonstrate that high-plex, spatially resolved X-CODE barcode detection is compatible with the Akoya PhenoCycler-Fusion workflow while retaining specificity and single-cell resolution.

6.4 Adaptation of X-CODE detection for imaging mass cytometry

Building on the establishment of X-CODE barcode detection in suspension using mass cytometry and spatial fluorescence imaging, these principles were combined to adapt X-CODE detection for imaging mass cytometry (IMC). Isotope-conjugated X-CODE detection probes, previously validated for barcode unit (BU) detection in suspension-based mass cytometry, were directly utilised in the spatial setting (**Supplementary Table 8**). TRAMP-C1 cells barcoded with the X-CODE 15K library were seeded onto microscope slides, and the X-CODE detection protocol for plated cells was performed with minor adaptations to allow efficient visualisation of cells and BUs by imaging mass cytometry (**Figure 6-4**A). These adaptations preserved the core steps of the X-CODE workflow while enabling compatibility with laser ablation-based mass cytometry image acquisition. The notable changes involve the use of isotope-conjugated probes, as described, and nuclei staining with a DNA intercalator containing iridium isotopes (intercalator-Ir).

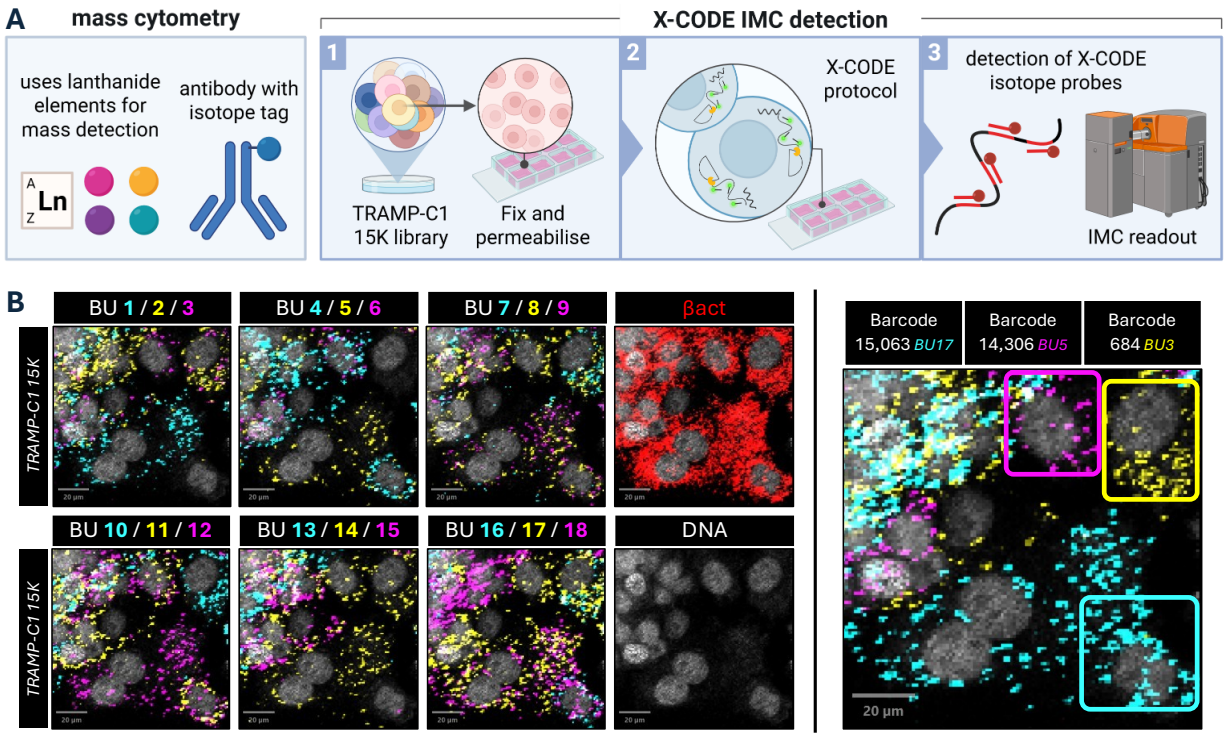


Figure 6-4. Protocol adaptation for X-CODE detection using high-parameter imaging mass cytometry in cells.

A) *Schematic representation of the X-CODE protocol for detection in plated cells by imaging mass cytometry. Mass cytometry traditionally uses antibodies with lanthanide isotopic tags; X-CODE utilises these tags for barcode detection. Created with BioRender.com.*

B) *Left: Representative TRAMP-C1 15K barcoded cells with barcode units (BUs) 1–18 visualised across six panels in triplicate (cyan, pink, and yellow). DNA intercalator staining visualising nuclei (white) and β-actin RNA staining as a protocol control (red). Right: Three representative cells with decoded X-CODE 15K barcodes, each visualised by a BU unique to one of the three identified barcodes (BC). BU17 (cyan) unique to BC15,063; BU5 (pink) to BC14,306; BU3 (yellow) to BC684. Scale bars, 20 µm.*

Cells were stained for BUs 1–18 using the 15K X-CODE isotope detection probe set to visualise the entire X-CODE 15K library, with β-actin serving as a protocol control, and nuclei were stained with intercalator-Ir (**Figure 6-4**B). Efficient and specific X-CODE signal detection for sequences 1–18 was observed across all targeted BUs (**Figure 6-4**B, left). Each unique barcode in the X-CODE 15K library is encoded by a combination of 8-of-18 BUs. In line with this design, individual cells showed clear positivity for a specific subset of eight barcode units, with minimal off-target signals detected in the remaining probes. Decoding of three representative cells is shown (**Figure 6-4**B, right), with each cell clearly positive for only eight sequences. Cross-referencing each BU combination against the whitelist of X-CODE 15K barcodes correctly identified barcodes 684, 14,306, and 15,063, with each cell visualised by a BU unique to one of the three identified barcodes (**Figure 6-4**B, right).

These results demonstrate that isotope-based X-CODE detection supports accurate barcode recovery and spatially resolved clonal identification at single-cell resolution using imaging mass cytometry.

6.5 Comparison of X-CODE detection across tissue embedding methods

Following the expansion of highly multiplexed X-CODE barcode unit (BU) detection in cells using fluorescence imaging and high-parameter imaging mass cytometry, X-CODE performance was evaluated in tissue sections. The four RM-1 X-CODE 2M populations J, W, Y, and Z were used to generate mixed (JWYZ) subcutaneous tumours in immunocompetent C57BL/6 mice. JWYZ tumours were processed using standard tissue preparation and embedding methods, including fresh freezing followed by optimal cutting temperature (OCT) embedding and formalin fixation with paraffin embedding (FFPE). Additionally, fresh-frozen JWYZ tumours were embedded in an HPMC/PVP hydrogel, selected for its reported compatibility with downstream molecular techniques

such as mass spectrometry imaging [164,165]. Since the success of X-CODE detection in tissue sections depends heavily on RNA stability, careful sample preparation of barcoded tumours was crucial. Particular attention was given to tissue handling, processing timelines, and storage conditions across all preparation methods, as is standard operating procedure for RNA *in situ* hybridisation protocols [177].

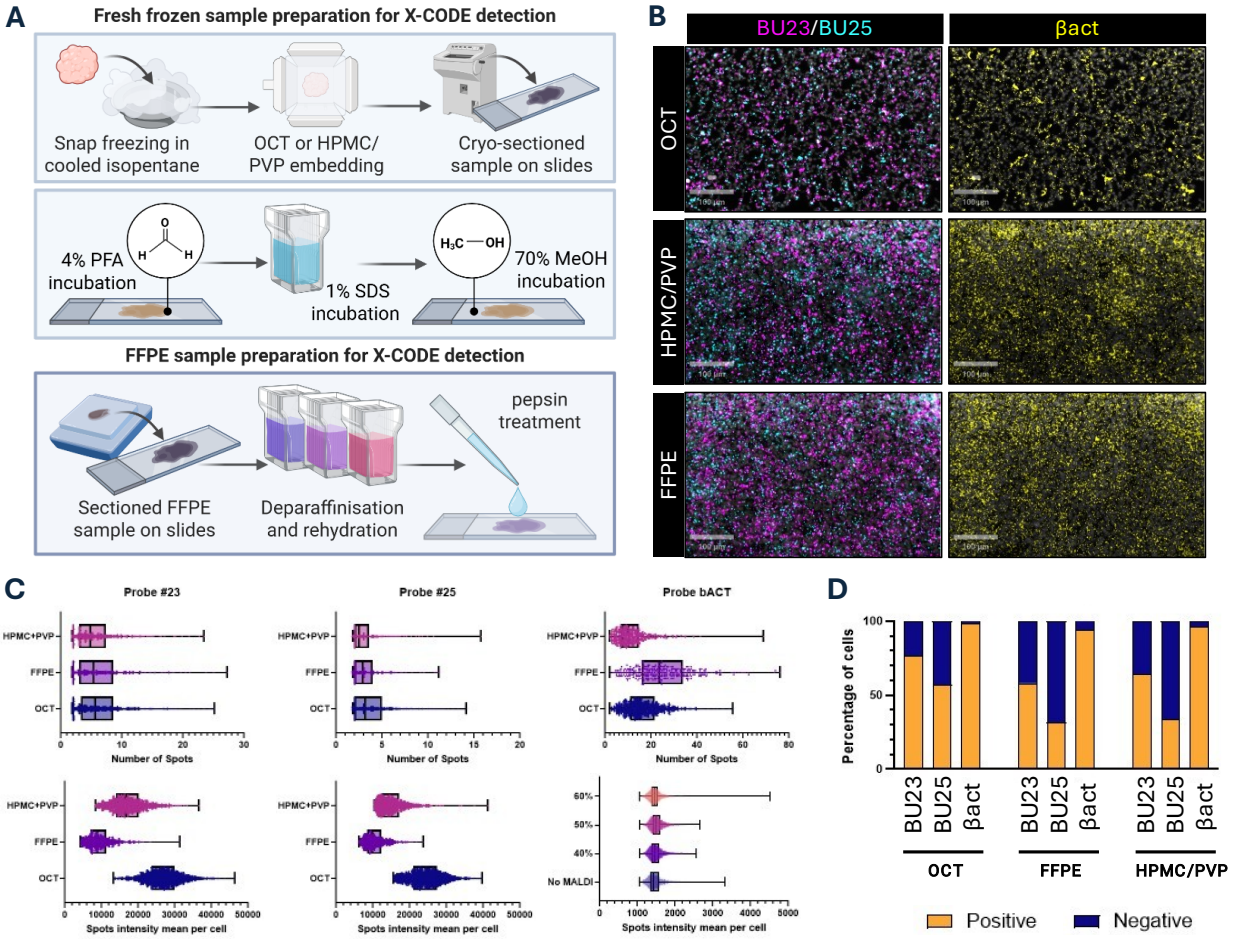


Figure 6-5. Assessment of X-CODE detection across fresh-frozen OCT or HPMC/PVP hydrogel embedded tissue, and FFPE tissue.

A) *Overview of the X-CODE sample preparation workflows for fresh-frozen optimal cutting temperature (OCT) or HPMC/PVP hydrogel-embedded tissue, and formalin-fixed paraffin-embedded (FFPE) tissues. HPMC = hydroxypropyl methylcellulose; PVP = polyvinylpyrrolidone; SDS = sodium dodecyl sulphate. Created with BioRender.com.*

B) *Representative RM-1 2M mixed JWYZ subcutaneous tumours underwent OCT, FFPE, or HPMC/PVP sample preparation, and were stained with probes detecting barcode unit (BU) 23 (yellow) and BU25 (cyan). DAPI staining visualising nuclei (white) and β-actin RNA staining as a protocol control (magenta). Scale bars, 100 µm.*

C) *Quantification of BU spots (top) and their mean intensity (bottom) per cell was performed using a QuPath-based spot-counting pipeline for representative probes (#23, #25, β-actin) in mixed JWYZ tumours processed using the three tissue embedding methods: OCT, FFPE, and HPMC/PVP.*

D) *Stacked bar graph showing the percentage of cells positive (spots > 0) and negative for representative probes (#23, #25, β-actin) across the three tissue embedding methods.*

The sample preparation workflow for fresh-frozen OCT-embedded tumours followed established practices to maintain RNA integrity in intact tissues, adapted from

demonstrated protocols developed for high-sensitivity spatial transcriptomic applications[1] (**Figure 6-5**A, top). After collecting JWYZ tumours, samples were immediately frozen and stored at –80 °C to preserve RNA quality. OCT embedding was performed promptly using dry ice–cooled isopentane to prevent thawing. Using similar principles, freshly frozen JWYZ tumours intended for HPMC/PVP hydrogel embedding were quickly processed following established protocols [164,165] (**Figure 6-5**A, middle). JWYZ tumours intended for formalin fixation and paraffin embedding were handled according to standard histological procedures [203] (**Figure 6-5**A, bottom). Briefly, freshly resected JWYZ tumours were submerged in 4% paraformaldehyde overnight at 4 °C before paraffin embedding. Additionally, FFPE blocks were stored at 4 °C rather than at room temperature, as is standard, to reduce RNA degradation during storage.

Fluorescence imaging, via confocal microscopy, was used to optimise X-CODE barcode detection in tissue sections due to its accessibility and suitability for protocol development. This approach enabled practical optimisation of robust tissue-compatible X-CODE detection workflows across embedding conditions prior to validation using high-parameter spatial platforms, including the Akoya PhenoCycler and imaging mass cytometry. OCT- and HPMC/PVP-embedded tissues were cryosectioned at 10 µm, and the X-CODE detection protocol optimised for fresh-frozen tissue was applied to visualise barcode sequences. In parallel, FFPE blocks were sectioned at 6 µm, and barcode sequences were visualised using the X-CODE detection protocol optimised for formalin-fixed tissues.

Across tissue preparation methods, sections were stained with fluorescent detection probes targeting BU23 and BU25, together with β-actin RNA as a protocol control (**Figure 6-5**B). Across all embedding conditions, barcode signals for representative BU probes and the β-actin control were consistently detected by fluorescence imaging. The efficiency of X-CODE detection across tissue preparation methods was quantitatively assessed using a QuPath-based spot-counting pipeline to determine the number of BU spots per cell. While spot counts, signal intensities, and the proportion of BU-positive cells varied between embedding conditions, signals remained quantifiable and interpretable across all methods, as reflected by spot count distributions, mean spot intensity per cell, and proportions of cells positive for BU23, BU25, and β-actin (**Figure**

[1] 10x Genomics Xenium In Situ Demonstrated Protocol: Fresh-Frozen Tissue Preparation Guide (CG000579, Rev D).

6-5C and D). Notably, the barcode of clones J, W, and Y contain BU23, and barcodes of clones W and Z contain BU25. Furthermore, near-universal cell positivity for β-actin RNA across tissue preparation techniques provided confidence in the optimised RNA detection protocols (**Figure 6-5**D).

Overall, these results demonstrate that X-CODE RNA barcode detection is compatible with multiple tissue-preservation and embedding strategies, enabling quantitative spatial barcode readout across diverse sample-preparation methods.

6.6 Spatial detection of X-CODE barcode libraries in fresh-frozen tissues

6.6.1 High-parameter fluorescence imaging for X-CODE detection in fresh-frozen tissues

Optimised X-CODE protocols for tumour sections were utilised for high-parameter spatial detection of X-CODE barcode libraries in fresh-frozen OCT- and HPMC/PVP hydrogel-embedded tissues using the Akoya PhenoCycler-Fusion. RM-1 JWYZ OCT and HPMC/PVP subcutaneous tumours were cryosectioned at 10 µm for processing and staining using the X-CODE–PhenoCycler workflow. Following completion of the X-CODE detection protocol for fresh-frozen tissues, sections underwent additional fixation and permeabilisation steps recommended by Akoya to support repeated imaging cycles on the PhenoCycler platform. Although these stages were optimised for highly multiplexed antibody staining, they were also incorporated into the X-CODE–PhenoCycler workflow, with or without antibody co-detection, to enable repeated cycles of fluorescent probe hybridisation and stripping from X-CODE rolling circle amplicons during PhenoCycler imaging. Tissue sections were stained for barcode units (BUs) 1–27 using X-CODE fluorescent detection probes (**Supplementary Table 10**), and β-actin RNA as a protocol control (**Figure 6-6**A).

Each RM-1 J, W, Y, and Z clonal line carries a unique barcode encoded by a combination of 12-of-27 BUs, with two distinguishing BUs per clone used for visualisation within the JWYZ tumour (**Figure 6-6**A). Visual assessment of X-CODE detection showed that BU signals were robust and confined to patterns expected of the

clonal lines in both OCT and HPMC/PVP hydrogel samples. Additionally, simultaneous visualisation of all four clones by their unique BU combinations demonstrated clear spatial separation of clonal populations within tumour sections across embedding media. X-CODE detection on the PhenoCycler platform in fresh-frozen tissue sections was quantitatively assessed using the QuPath-based spot-counting pipeline to detect BU signal per segmented cells in JWYZ tumours. Percentages of positive cells per BU demonstrated reliable detection across all 27 X-CODE sequences in OCT and HPMC/PVP-embedded samples. β-actin positivity showed near-universal cellular detection (**Figure 6-6**B).

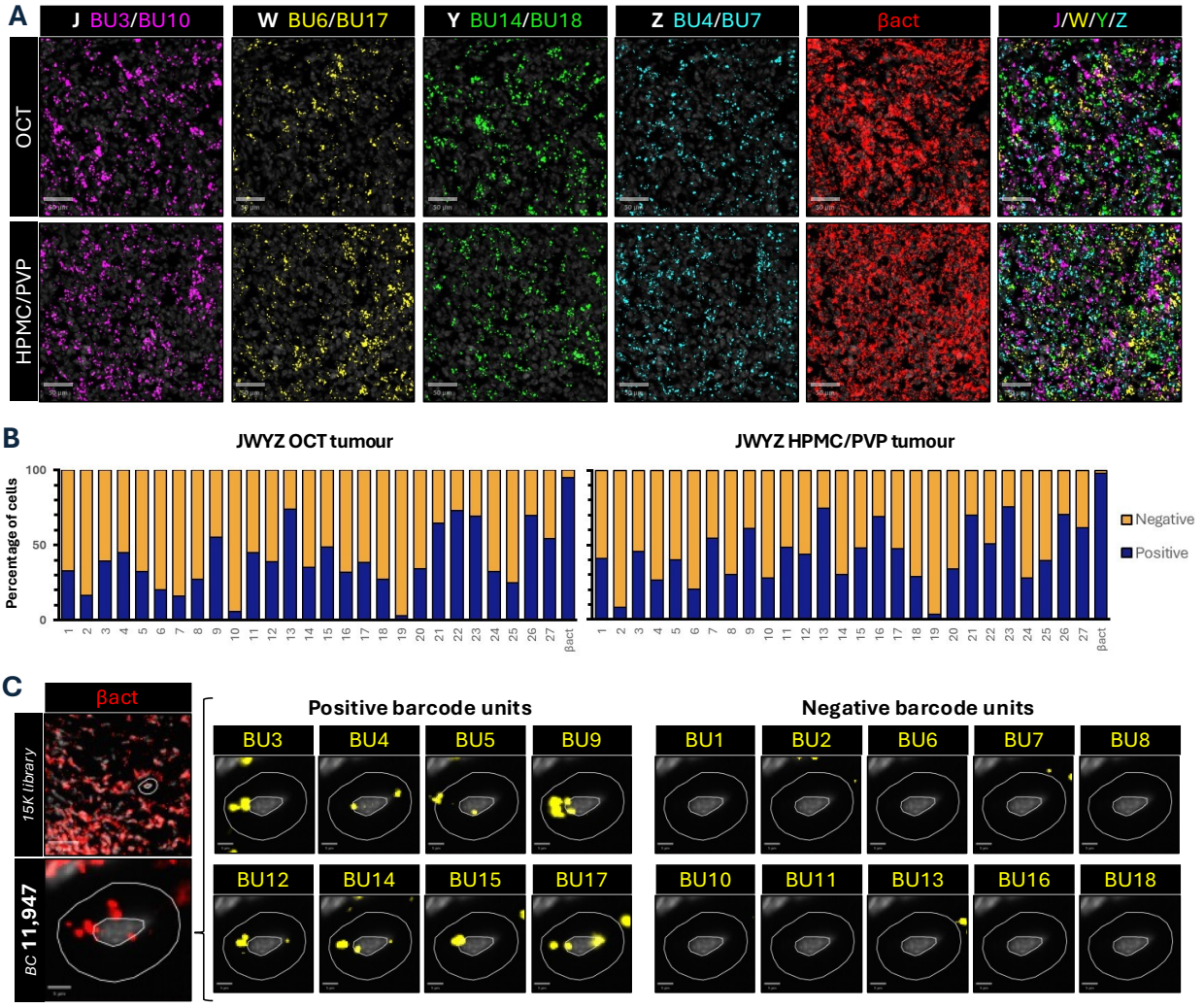


Figure 6-6. X-CODE detection of 2M clonal lines and 15K library in OCT and HPMC/PVP embedded tissues by Akoya PhenoCycler.

A) *Fresh-frozen OCT- and HPMC/PVP hydrogel-embedded RM-1 JWYZ tumours were stained for X-CODE barcode units (BUs) 1–27 and β-actin RNA as a protocol control and visualised by Akoya PhenoCycler-Fusion. Representative RM-1 2M clonal populations are shown, visualised by distinguishing BU sequences: J (pink, BU3&10); W (yellow, BU6&17); Y (green, BU14&18); and Z (cyan, BU4&7). DAPI staining for nuclei (white). Scale bars, 50 µm.*

B) *Quantification of BU signal per segmented cell was performed using a QuPath-based spot-counting pipeline. Stacked bar graphs show percentages of positive and negative cells per BU across embedding media.*

C) *HPMC/PVP RM-1 15K tumour sections were stained for BUs 1–18 (yellow) and β-actin (red) RNA, visualised by Akoya PhenoCycler-Fusion. A representative cell showing its positive and negative BU units, decoded against the X-CODE 15K library whitelist as barcode (BC) 11,947. Tissue-focused scale bar, 50 µm. Cell-focused scale bars, 5 µm.*

Following *in situ* detection of all X-CODE sequences using JWYZ tumours, tissue samples from the X-CODE 15K library were also analysed. An RM-1 population harbouring the X-CODE 15K library were used to generate subcutaneous tumours in immunocompetent C57BL/6 mice. Tumours were fresh-frozen, embedded in an HPMC/PVP hydrogel, and cryosectioned at 10 µm for processing and staining using the X-CODE–PhenoCycler workflow. Tissue sections were stained for BUs 1–18 using X-CODE fluorescent detection probes (**Supplementary Table 10**), and β-actin RNA as a protocol control (**Figure 6-6**C, left). Representative X-CODE decoding shows a 15K-cell that is positive for only 8-of-18 sequences (**Figure 6-6**C, right). Cross-referencing the BU combination against the X-CODE 15K barcode whitelist correctly identified the barcode as 11,947.

Collectively, these findings establish that X-CODE 15K and 2M barcode libraries can be robustly and quantitatively resolved at single-cell resolution in fresh-frozen OCT- and HPMC/PVP-embedded tumour sections using the Akoya PhenoCycler-Fusion platform.

6.6.2 Detection of X-CODE barcodes in fresh-frozen tissues by imaging mass cytometry

To evaluate the compatibility of spatial *in situ* X-CODE detection with an orthogonal high-parameter modality, detection of the X-CODE 15K library in fresh-frozen tumour sections was next assessed using imaging mass cytometry. TRAMP-C1 cells barcoded with the X-CODE 15K library, and their wild-type (WT) counterparts not expressing X-CODE barcodes, were injected into immunocompetent C57BL/6 mice to generate subcutaneous tumours. 15K and WT tumours were processed according to the X-CODE fresh-frozen OCT workflow and the previously described HPMC/PVP hydrogel embedding protocols [164,165]. 15K and WT tumours were cryosectioned at 10 µm and subjected to the X-CODE protocol for fresh-frozen samples. Barcoded and WT tissue sections were stained for barcode units (BUs) 1–18 using the 15K X-CODE isotope detection probe set (**Supplementary Table 8**), including a rolling-circle amplification negative (RCA–) control and β-actin RNA staining as a protocol control. Nuclei were visualised using a DNA intercalator containing iridium isotopes (**Figure 6-7**A).

Consistent with X-CODE visualisation using the PhenoCycler platform, imaging mass cytometry (IMC) enabled robust and specific detection of X-CODE BUs 1–18 in both OCT- and HPMC/PVP hydrogel-embedded tissue sections. Representative IMC images display signal from X-CODE isotope-conjugated detection probes targeting BUs 1–18, visualised in triplicate with three repeated colour assignments (**Figure 6-7**A, left). BU-specific signals were clearly distinguishable, with no off-target or non-specific signals detected in WT tissue. Additionally, the lack of barcode signal in the no–rolling circle amplification (RCA–) control confirms that φ29-mediated amplification is necessary for successful X-CODE detection. Padlock probes targeting β-actin RNA consistently produced isotope-mediated signals in both X-CODE 15K and WT tumours, across different embedding conditions, serving as a control for RNA detection via padlock probe-mediated RCA in tissue sections (**Figure 6-7**A, right).

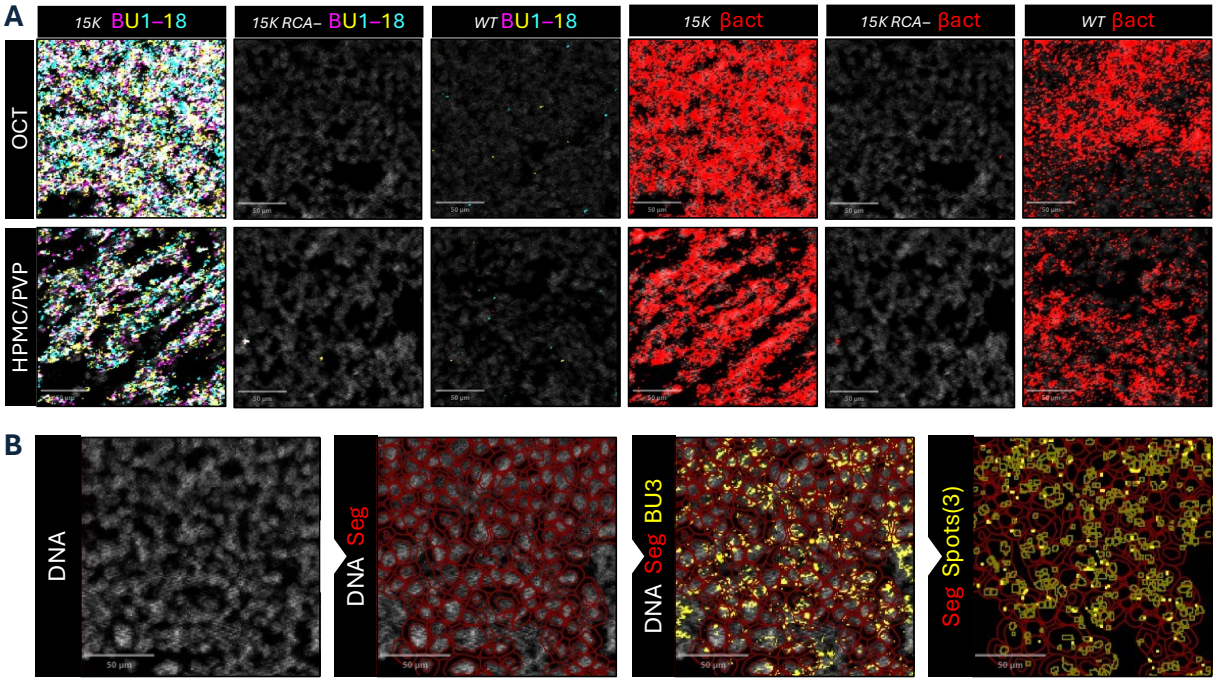


Figure 6-7. X-CODE 15K library detection in fresh-frozen OCT and HPMC/PVP embedded tissues by imaging mass cytometry.

A) Imaging mass cytometry–based detection of the X-CODE 15K library in fresh-frozen OCT and HPMC/PVP hydrogel embedded tumour sections, generated from TRAMP-C1 15K cells. Isotope-conjugated X-CODE detection probes targeting BUs 1–18 were visualised in representative IMC images, with wildtype (WT), no–rolling circle amplification (RCA–), and β-actin RNA detection controls. Scale bars, 50 µm.

***B)** Representative IMC analysis workflow illustrating computational QuPath-based analysis steps. From left to right: nuclei visualisation; nuclear segmentation based on DNA intercalator signal; overlay of segmentation masks with representative BU3 signal; and BU3 spot masks generated by spot detection from intensity-based thresholding. Scale bars, 50 µm.*

In contrast to fluorescence-based imaging, IMC data exhibit a granular signal profile and reduced visual contrast between neighbouring cells, reflecting fundamental differences

in acquisition modality and signal generation. Representative computational analysis of X-CODE IMC data from the fresh-frozen OCT-embedded X-CODE 15K tumour is shown in **Figure 6-7**B. Although mass cytometric imaging achieves single-cell resolution, nuclear segmentation based on the DNA intercalator signal was less well-defined than DAPI-based segmentation used for fluorescence imaging. Nonetheless, the existing QuPath-based spot counting pipeline was successfully applied to IMC data. Sequential processing steps are illustrated, including nuclear segmentation, overlay of the segmentation masks with the representative BU3 signal, and generation of final BU3 spot masks via intensity thresholding, which can be used for quantitative X-CODE analysis. Despite the altered visual characteristics of IMC images, BU-specific spots were accurately detected and assigned within segmented regions, demonstrating that X-CODE BU detection and quantification are compatible with imaging mass cytometry using a unified analysis workflow.

These results demonstrate that high-parameter spatial detection and quantification of X-CODE barcode libraries in fresh-frozen tissues is compatible with imaging mass cytometry workflows across OCT- and HPMC/PVP hydrogel-embedded samples.

6.7 Antibody co-detection considerations in fresh-frozen tissue

During the X-CODE–PhenoCycler protocol, tissue sections undergo additional fixation and permeabilisation steps recommended by Akoya to support repeated imaging cycles on the PhenoCycler platform, as previously described in 6.6.1. Although these steps were optimised for highly multiplexed antibody staining, they were also incorporated into the X-CODE–PhenoCycler workflow, with or without antibody co-detection, to enable repeated cycles of fluorescent probe hybridisation and stripping from X-CODE rolling circle amplicons during PhenoCycler imaging. Co-detection of a single protein (sodium–potassium ATPase) with X-CODE barcodes was demonstrated in 6.6.1, without comparison to Akoya's recommended conditions for antibody staining. Therefore, to further explore the compatibility of X-CODE–protein co-detection, multiple antibodies were co-detected with X-CODE BUs. The same antibody panel was applied to serial sections without X-CODE BU staining, following Akoya's recommended workflows for PhenoCycler-based antibody staining.

To assess the compatibility of X-CODE and protein co-detection in fresh-frozen tissue sections, RM-1 JWYZ tumours embedded in OCT or HPMC/PVP hydrogel were processed with X-CODE and antibody staining workflows on the Akoya PhenoCycler-Fusion. Consecutive tissue sections from JWYZ tumours were stained using either Akoya's recommended PhenoCycler protocol (Ab only) or the optimised X-CODE–PhenoCycler protocol (X+Ab). Across both embedding conditions, protein staining with a greater number of antibodies did not interfere with X-CODE BU detection, and a robust X-CODE signal from fluorescent probes was retained (**Figure 6-8**A). Whereas protein staining performance varied across markers: Ki67, vimentin, and sodium–potassium ATPase (Na/K-ATPase) showed broadly comparable signal intensity and spatial distribution between the Ab-only and X+Ab protocols. However, CD4 and CD8 staining were reduced following X-CODE processing (**Figure 6-8**B).

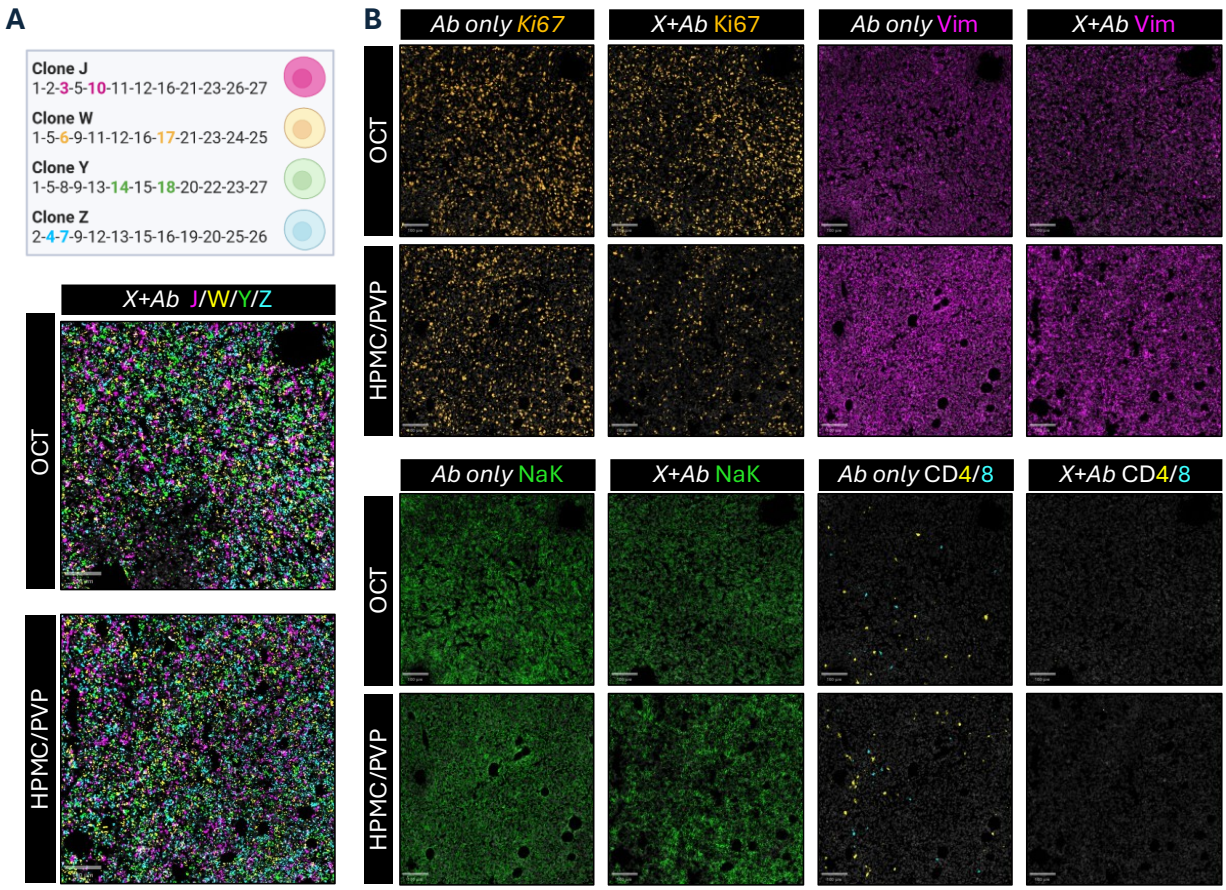


Figure 6-8. Considerations for co-detection of antibodies with X-CODE staining in fresh-frozen OCT- and HPMC/PVP-embedded tissues.

A) *Fresh-frozen OCT and HPMC/PVP RM-1 JWYZ tumours were stained for X-CODE BUs 1–27, and an antibody staining panel, visualised by PhenoCycler readout. Representative RM-1 2M clonal populations are shown, visualised by distinguishing BU sequences: J (pink, BU3&10); W (yellow, BU6&17); Y (green, BU14&18); and Z (cyan, BU4&7). DAPI staining for nuclei (white). Scale bars, 100 µm. Key created with BioRender.com.*

B) *Comparison of antibody staining performance in fresh-frozen OCT- and HPMC/PVP hydrogel-embedded RM-1 JWYZ tumour sections processed using the standard Akoya PhenoCycler protein staining protocol (Ab only) or the*

combined X-CODE–PhenoCycler workflow (X+Ab). A representative, coherent tumour region from consecutive sections is shown. Representative antibody markers include Ki67, vimentin (Vim), Na/K-ATPase (NaK), and CD4/CD8. Scale bars, 100 µm.

Overall, these results demonstrate that antibody co-detection is compatible with X-CODE barcode detection in fresh-frozen OCT- and HPMC/PVP-embedded tissues, while highlighting marker-specific sensitivity to the integrated workflow.

6.8 Discussion

This chapter presents X-CODE as a flexible platform for high-parameter spatial barcode detection using multiplexed imaging technologies. Initial optimisation of padlock and universal probe concentrations demonstrated that efficient, specific barcode detection could be achieved at the lowest concentrations tested, providing a robust baseline for downstream spatial applications. Building on this, X-CODE barcode detection was successfully implemented on the Akoya PhenoCycler-Fusion, enabling high-plex, spatially resolved fluorescence-based detection at single-cell resolution. Hybridisation of padlock probes for barcode units 1–27, spanning the 15K and 2M-plex X-CODE libraries, yielded a robust signal via rolling-circle amplification. Importantly, barcode detection remained specific and quantitative despite inherent sequence-dependent variability in signal intensity, consistent with observations from suspension-based X-CODE detection.

Extension of X-CODE detection in cells to imaging mass cytometry further demonstrated that isotope-labelled barcode probes support accurate barcode recovery and spatially resolved clonal identification at single-cell resolution. Although imaging mass cytometry exhibited distinct visual characteristics compared with fluorescence imaging, including increased signal granularity, barcode units could be reliably detected for quantification using a unified computational pipeline. Additionally, decoding of 15K library cells was achieved by detecting only 8 of 18 sequences per cell, despite visual differences relative to fluorescence imaging data. Together, these findings demonstrate that X-CODE detection is compatible with complementary high-parameter imaging modalities, enabling spatial barcode readout across platforms with minimal protocol divergence.

Robust X-CODE barcode detection was further validated across multiple tissue preservation and embedding strategies. Fresh-frozen OCT- and HPMC/PVP hydrogel-

embedded tissues, as well as FFPE samples, supported quantitative barcode detection, establishing compatibility with standard and alternative embedding approaches. High-parameter detection of X-CODE barcode libraries in fresh-frozen tissues was achieved using both fluorescence-based imaging and imaging mass cytometry, confirming that barcode readout is retained across platforms and tissue preparation methods. Antibody co-detection experiments further demonstrated that combined barcode and protein detection is feasible in fresh-frozen tissues, while highlighting marker- and protocol-dependent considerations that may require optimisation for reliable multiplexed phenotyping.

Previous implementations of spatial barcoding have primarily relied on manual, iterative rounds of fluorescence imaging, which are labour-intensive [58,72], or more recently, on adaptations to spatial transcriptomic platforms [159,160] . Although spatial transcriptomic approaches enable rich molecular profiling, they remain costly per sample and often require specialised proprietary sample preparation workflows that limit compatibility with complementary molecular phenotyping. Spatial detection of protein-based barcodes has also been demonstrated [91] ; however, library complexity is inherently constrained by the number of available epitopes, thereby restricting its applicability to resolving large or highly diverse clonal populations. To our knowledge, this work represents the first demonstration of a high-complexity RNA-based lineage-tracing barcode detected using the Akoya PhenoCycler-Fusion platform, and the first demonstration of compatibility with HPMC/PVP hydrogel embedding for combined PhenoCycler-Fusion antibody staining and spatial barcode detection.

Similarly, while imaging mass cytometry has been used for spatial detection of protein barcodes [91] and for RNA detection [150], this work extends its application to the detection of RNA-based lineage-tracing barcodes. Demonstrating that X-CODE detection can be seamlessly transferred between fluorescence-based imaging and mass cytometry modalities. Across platforms, fluorescence imaging provided sharper visual contrast and more clearly defined nuclear segmentation compared to imaging mass cytometry. Nevertheless, barcode spots were accurately detected and quantified across both modalities using the same QuPath-based analysis pipeline, without the need for specialised computational tools or platform-specific analysis software. This unified analysis approach lowers the technical barrier to adoption and enables direct comparison of spatial barcode data across imaging platforms. Significantly, X-CODE

detection relies on a shared core protocol and reagent set, with only minor adaptations required for different readout technologies, allowing barcode identity to be retained seamlessly across platforms.

6.8.1 Limitations

The X-CODE system has been demonstrated as a flexible, multi-modal barcode for cross-platform detection, enabled by intensive optimisation of sample preparation and processing protocols. However, optimisation of padlock and universal probe concentrations was performed using a limited subset of barcode units (BUs) for each of the 15K and 2M-plex clonal lines. As observed in this chapter and in the previous chapter focusing on suspension-based X-CODE detection, individual BUs exhibit sequence-dependent variability in detection efficiency. While this variability can be accommodated computationally during analysis by intensity- and size-based thresholding during spot quantification, additional probe-specific optimisation of hybridisation conditions may improve the uniformity of detection across all X-CODE sequences.

X-CODE barcode detection was validated across multiple tissue-embedding strategies; however, full-scale compatibility of X-CODE barcode libraries in formalin-fixed paraffin-embedded (FFPE) tissue was not demonstrated in this chapter. While no foreseen limitations are expected to prevent the detection of X-CODE 15K- or 2M-plex libraries in FFPE samples, successful integration with the PhenoCycler-Fusion may require protocol adaptations and reagent substitutions aligned with established Akoya protocols for FFPE tissue sections. Namely, Akoya recommends different combinations of fluorophore conjugations for imaging of FFPE tissue sections [2], and has optimised FFPE-specific antibody panels. In addition, X-CODE detection in FFPE tissues was not demonstrated by imaging mass cytometry (IMC). Nonetheless, *in situ* hybridisation of endogenous RNA transcripts using RNAscope in FFPE tissue sections, with antibody co-detection, has been demonstrated using IMC [150]. This work by *Schulz et al.* indicates that robust IMC protocols for probe-based *in situ* RNA detection exist; however, X-CODE-specific detection may require optimisation of the relevant protocols.

[2] Akoya Biosciences PhenoCycler Fusion (formerly CODEX) User Guide [accessed 10/01/2026]

While combined antibody co-detection with X-CODE staining was evaluated using fluorescence-based PhenoCycler workflows, it was not demonstrated using imaging mass cytometry. Additionally, X-CODE–protein co-detection in FFPE tissue was not explored using either the PhenoCycler or imaging mass cytometry workflows. However, reports and workflows demonstrate successful probe-based RNA detection, combined with immunofluorescence staining, by fluorescence imaging in FFPE tissues [204,205] [3]. Therefore, X-CODE–protein fluorescence co-detection in FFPE sections is likely feasible, ensuring that X-CODE-specific protocol optimisations are carried out. Similarly, *Schulz et al.* demonstrate RNA–protein co-detection on FFPE tissue sections using imaging mass cytometry. Co-detection using both modalities would likely require per-target optimisation, as described for RNA–protein staining in fresh-frozen sections. Therefore, further work will be necessary to establish robust combined barcode and protein detection for tissues sections across fluorescence and imaging mass cytometry platforms.

Finally, spatial detection of the 2M-plex clonal X-CODE lines was not demonstrated using imaging mass cytometry. However, all isotope-conjugated probes targeting BUs 1–27 were validated in suspension mass cytometry in the preceding chapter. Additionally, no foreseen imaging-modality-specific constraints have been identified that would preclude extending IMC detection to higher-complexity barcode libraries. A potential consideration is the increased granularity of IMC images relative to fluorescence data. This may complicate analysis, as the assignment of a larger number of BUs per cell is required for the 2M-plex library (12 of 27 combinations) than for the 15K-plex library (8 of 18 combinations). Optimisation of computational pipelines for improved cellular segmentation in IMC data is expected to facilitate accurate spot detection and assignment to individual cells.

6.8.2 Conclusions

In conclusion, this chapter establishes X-CODE as a versatile and platform-agnostic system for high-parameter spatial barcode detection using multiplexed imaging technologies. X-CODE barcodes were robustly detected, quantified, and decoded using fluorescence-based imaging and imaging mass cytometry, across multiple tissue

[3] RNAscope® Multiplex Fluorescent v2 Assay combined with Immunofluorescence | Bio-Techne [accessed 10/01/2026]

embedding strategies, and in combination with antibody-based phenotyping. By enabling seamless barcode readout across imaging platforms using shared protocols and analysis pipelines, X-CODE provides a scalable framework for spatial clonal tracing integrated with high-dimensional phenotypic analysis in intact tissues.

Chapter 7 Integration of X-CODE spatial workflows with unbiased mass spectrometry imaging

7.1 Research in context

Recently, substantial experimental and computational efforts have focused on integrating multiple single-modality datasets into unified multiomic analyses of biological samples [133,200] . These approaches aim to reveal fundamental molecular mechanisms that govern cellular behaviour, development, physiology, and disease progression that could not be discovered from a single modality alone [104]. Within this context, integrating molecular landscapes derived from mass spectrometry imaging (MSI) with spatially resolved lineage information offers a promising avenue for investigating clonal molecular heterogeneity and its significance in tissues. Building on the robust high-plex fluorescence RNA barcode detection demonstrated in the preceding chapter, the compatibility of integrating the X-CODE framework with MSI can be explored.

Nonetheless, multimodal integration presents challenges. Practically, integrating techniques on the same tissue section is limited by the compatibility of samples and analytes across methodologies [104,106]. MSI–RNA integration is particularly difficult because molecular analytes are highly sensitive. Therefore, MSI must be performed first [152,154–156,206–208], as reagents for RNA detection can degrade metabolites. However, this creates uncertainty about RNA survival post-MSI, as the technique can induce chemical and physical sample degradation. Furthermore, computational integration of MSI and fluorescence imaging data is challenging. MSI typically operates at a lower resolution than fluorescence imaging, creating pixel-level disparities that complicate co-registration [200]. Accurate alignment of these datasets depends on dedicated computational pipelines, and large-scale efforts to develop integration tools are underway as the volume of multimodal data increases [104,106].

As discussed previously in Section 1.3.5, MetaFISH [152] and OpenFISH [156] pipelines have established combined matrix-assisted laser desorption/ionisation (MALDI) and fluorescence *in situ* hybridisation (FISH) workflows. Although these works present sophisticated MSI–RNA workflows, cellular barcode detection for lineage tracing following MSI has not yet been demonstrated. Here, maintaining RNA integrity is

essential for accurate recovery of cellular barcodes, as degradation can lead to missing barcodes or misidentified cells, thereby compromising clonal resolution. Additionally, workflows using high-plex imaging platforms for post-MSI RNA detection have not yet been optimised. Together, this motivates the exploration of alternative imaging-based strategies for probe-based RNA detection of cellular barcodes after MSI, providing high-plex spatial readouts of clonal architecture overlaid with molecular landscapes.

The MetaFISH workflows emphasise that MSI signal quality and RNA preservation are inherently conflicting, necessitating experimental optimisation to achieve a balanced dataset rather than maximising results from either modality [152,206]. MALDI acquisition parameters that enhance ion yield and molecular coverage, particularly increased laser energy or repeated ablation, inevitably compromise tissue integrity and reduce the recoverability of RNA for downstream *in situ* detection. Moreover, the applied MALDI matrix directly influences metabolite ionisation efficiency and the range of analyte classes detected, as some ionise more effectively at higher primary laser intensities, whereas others produce stronger signals with increased laser shots per pixel [206]. Consequently, integrated workflows typically converge on balanced operating parameters that preserve sufficient molecular information from both readouts.

Furthermore, sample-preparation considerations influence the compatibility between MSI and RNA-based imaging. Tissue embedding media must be selected to avoid interference with ionisation, as the commonly used optimal cutting temperature (OCT) compound for fresh-frozen tissue embedding can suppress analyte signals during MSI by competing for available ions [209]. While MSI compatibility with OCT-embedded tissue was recently demonstrated, the workflow was optimised using nano-DESI [210]. Alternative embedding strategies, like hydrogel with hydroxypropyl methylcellulose and polyvinyl-pyrrolidone (HPMC/PVP), demonstrate improved compatibility across multiple molecular techniques and have been adopted as a reliable embedding method for fresh-frozen tissues intended for MSI [164,165].

Overall, these developments establish a clear rationale for systematically evaluating X-CODE detection following MALDI-MSI. Drawing on established MALDI–FISH principles, this chapter explores the integration of unbiased MALDI-MSI with high-plex, spatially resolved clonal identification. Thereby, aiming to provide a practical workflow to achieve

comprehensive *in situ* phenotyping with spatially resolved cellular barcoding information within a single tissue section.

7.1.1 Chapter aims

- To evaluate the feasibility of integrating unbiased mass spectrometry imaging with X-CODE *in situ* RNA detection on the same tissue section.
- To establish post-MALDI X-CODE detection workflows compatible with fresh-frozen tissue and commonly used MALDI matrices.
- To define the trade-off between MALDI-MSI acquisition parameters and post-MALDI X-CODE RNA detectability.

7.2 Establishing post-MALDI *in situ* X-CODE RNA detection on the same tissue section

7.2.1 X-CODE detection following MALDI-MSI using a DHB matrix

Having established the compatibility of X-CODE detection with spatial imaging across multiple embedding conditions in Chapter 6, the integration of X-CODE detection with matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) as an upstream modality was assessed. The previously described MetaFISH protocol [152,206], which integrates MSI–FISH workflows, indicates careful optimisation is required to balance molecular signal acquisition with RNA integrity preservation. Given that robust and well-established methodologies exist for the detection of various analyte classes via MALDI-MSI on fresh-frozen tissue [133,211], the initial focus of this chapter was to evaluate the impact of MALDI acquisition on downstream X-CODE RNA detection.

Based on prior reports that MALDI-FISH workflows require matrix- and acquisition-parameter optimisation [152,206], X-CODE detection was assessed following MALDI-MSI in positive-ionisation mode, using a 2,5-dihydroxybenzoic acid (DHB) matrix and increasing laser power. DHB was selected as an initial test matrix due to its widespread use and universal applicability for compound detection across analyte classes in MALDI-MSI [212,213]. Previously described RM-1 subcutaneous tumours harbouring the X-CODE 15K library were used to optimise MALDI–X-CODE integration. Samples were fresh-frozen and embedded in HPMC/PVP hydrogel to ensure MSI compatibility [164,165].

10 µm tissue sections were subjected to MALDI-MSI acquisition at three laser powers (30%, 70%, 90%) to systematically explore RNA survival. Subsequently, the tissue sections were processed for X-CODE detection (**Figure 7-1**A).

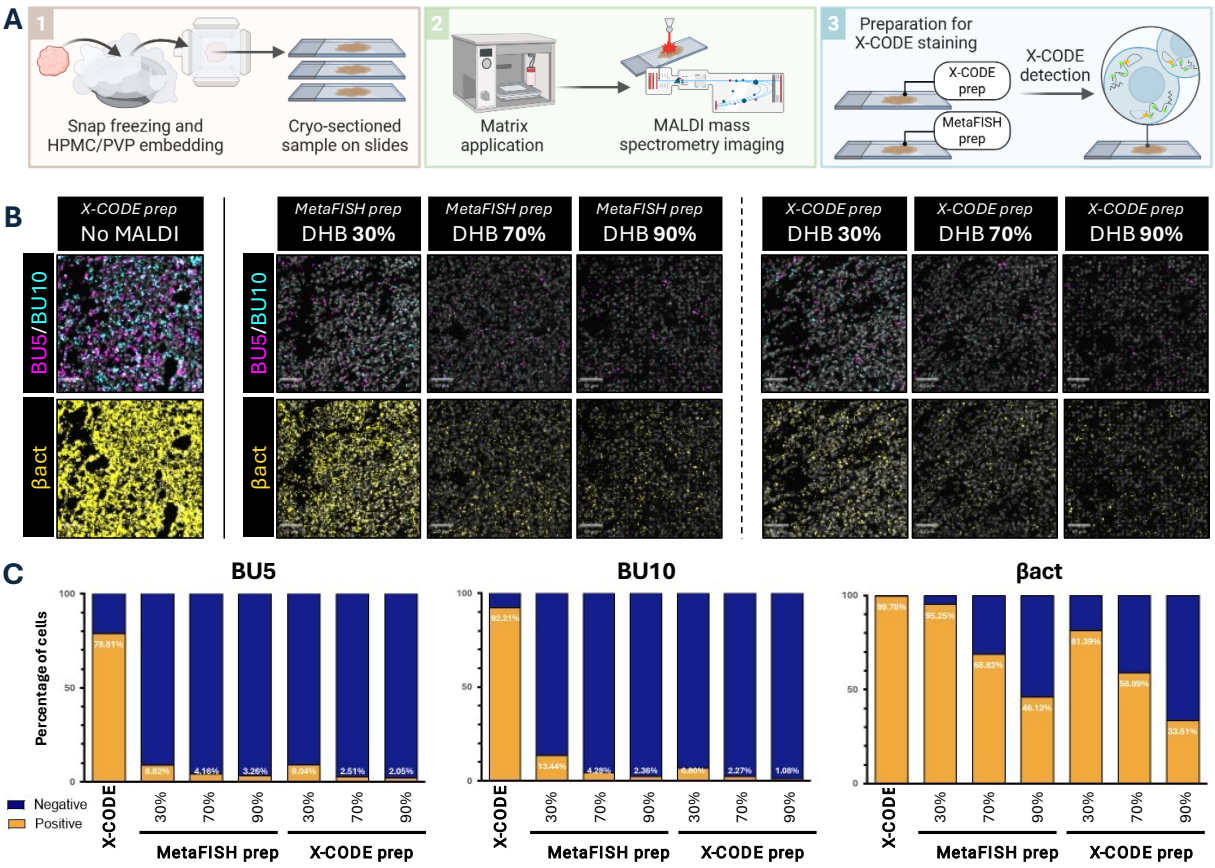


Figure 7-1. Evaluation of rolling circle-mediated RNA detection of X-CODE barcode units and β-actin following MALDI-MSI using DHB matrix.

A) *Schematic overview of the workflow for HPMC/PVP hydrogel-embedded, fresh frozen tissue sections undergoing matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI), followed by X-CODE detection under two post-MSI tissue section preparations: X-CODE-style preparation (prep) and MetaFISH-style prep. Created with BioRender.com.*

B) *RM-1 X-CODE 15K subcutaneous tumour sections underwent MALDI-MSI at laser powers of 30%, 70%, or 90% using a 2,5-dihydroxybenzoic acid (DHB) matrix. Consecutive tissue sections were processed post-MALDI using either the MetaFISH or X-CODE sample preparation protocol. X-CODE barcode units (BUs) 1–18 were detected using the Akoya PhenoCycler-Fusion platform. Representative images show BU5 (magenta) and BU10 (cyan) across laser powers and preparation conditions, alongside an X-CODE-only control that did not undergo MALDI-MSI. DAPI staining visualising nuclei (white) and β-actin RNA staining as a protocol control (yellow). Scale bars, 50 µm.*

C) *Quantification of BU5, BU10, and β-actin RNA amplicon spots per cell across conditions using a QuPath-based spot-counting pipeline. Stacked bar graphs show the percentages of positive (spots > 0) and negative cells for BU5, BU10, and β-actin across laser powers and sample-preparation conditions. X-CODE-only control without MALDI-MSI compared to MetaFISH-style prep and X-CODE-style prep following MALDI at 30, 70, and 90% laser power. Percentage values (white) indicate the proportion of positive cells.*

The initial stage of tissue preparation for RNA *in situ* hybridisation involves fixation with paraformaldehyde (PFA). Additionally, alcohol-based fixatives, methanol and ethanol, which enhance permeability, can be used in conjunction with PFA or alone. Tissue permeabilisation stages using low concentrations of Tween-20, Triton X-100, or sodium

dodecyl sulphate (SDS) can also be performed [177]. The optimised X-CODE protocol for fresh-frozen tissue (X-CODE prep) involves a 30-minute fixation with 4% PFA, followed by a 2-minute permeabilisation with 1% SDS, and an additional 1-hour incubation in 70% methanol. In contrast, the MetaFISH workflow (MetaFISH prep) for fresh-frozen tissue preparation following MALDI indicates a fixation for 1 hour in 2% PFA, followed by a standard FISH protocol.

Following samples preparation with the MetaFISH or X-CODE prep, X-CODE detection was performed using probes targeting barcode units (BUs) 1–18, with β-actin as a protocol control. RNA signals were visualised using the Akoya PhenoCycler platform following the previously optimised X-CODE imaging protocol. This comparison aimed to determine whether the more extensive fixation and permeabilisation procedures usually required for X-CODE detection, as opposed to those outlined by MetaFISH, remain suitable after MALDI-induced tissue disruption.

Representative images illustrate the effect of increasing MALDI laser power on detected RNA signal for BU5, BU10, and β-actin, compared to X-CODE-only control sections that were not subjected to MALDI-MSI (**Figure 7-1**B). In line with prior MALDI-FISH studies [152,206] , increasing laser power resulted in a visually marked reduction in detectable RNA signal, with pronounced tissue damage evident at the highest power setting. Quantitative analysis using the QuPath-based spot-counting pipeline confirmed a substantial signal reduction of BU5, BU10, and β-actin cell positivity following MALDI-MSI relative to X-CODE-only controls (**Figure 7-1**C). β-actin provided a sensitive readout of global RNA preservation because of its expected universal expression across all cells. In X-CODE-only control sections, 99.78% of cells were positive for β-actin. Following MALDI-MSI and MetaFISH prep, β-actin positivity decreased progressively with increasing laser power (lp), from 95.25% at 30% lp to 53.88% at 90% lp.

This reduction was more pronounced when tissues were processed using the standard X-CODE preparation protocol following MALDI-MSI. Under these conditions, β-actin positivity decreased from 81.39% of cells at 30% lp to only 33.51% of cells at 90% lp. These findings suggest that the additional permeabilisation steps included in the X-CODE protocol may exacerbate RNA loss when applied to tissues that have already undergone MALDI-induced disruption. Furthermore, the reduction in signal detection of barcode units was more pronounced following MALDI than that of β-actin. No-MSI X-

CODE controls showed 78.81% and 92.21% of cells positive for BU5 and BU10, respectively. In contrast, BU5 and BU10 cellular positivity was lower than 15% across all MALDI laser power and post-MSI tissue section prep conditions (**Figure 7-1**C).

Collectively, these results indicate that while rolling circle amplification enables robust signal generation under standard conditions, it cannot compensate for extensive RNA degradation caused by high MALDI laser energies. The data further suggest that post-MALDI tissue disruption may reduce the need for multi-step fixation and permeabilisation, and that milder fixation-only protocols may better preserve residual RNA for X-CODE detection. These observations highlight the need to explore alternative MALDI matrices and acquisition conditions, as reported in prior MetaFISH studies, as well as sample preparation protocols to improve RNA detection in integrated workflows.

7.2.2 X-CODE detection following MALDI-MSI using a DHAP matrix

Given the substantial RNA loss observed following MALDI-MSI with a DHB matrix, alternative MALDI matrix conditions were explored to assess whether post-MSI X-CODE RNA detection could be improved. Specifically, the 2,5-dihydroxyacetophenone (DHAP) matrix was evaluated, as it has also been demonstrated in post-MALDI RNA *in situ* hybridisation workflows [206]. Previously described RM-1 subcutaneous JWYZ tumours generated from a mixture of X-CODE 2M clonal lines were selected to assess RNA preservation and barcode detection under conditions more favourable to downstream *in situ* hybridisation. Tumour samples were fresh-frozen and embedded in HPMC/PVP hydrogel to maintain compatibility with MALDI-MSI [164,165]. 10 µm tissue sections were analysed by MALDI-MSI at laser powers of 40% and 50%. Powers were selected to investigate RNA detection under intermediate-energy acquisition conditions while aiming to maintain an effective MSI signal intensity, consistent with prior MSI analysis using a DHAP matrix [206].

Consecutive tissue sections were processed following MALDI-MSI using an adapted MetaFISH-style preparation protocol that incorporated a 10-minute 30% methanol incubation step to remove residual DHAP matrix prior to RNA detection. This additional step was required because DHAP exhibited lower solubility in aqueous buffers, such as phosphate-buffered saline, than DHB under the conditions used in this work. This is

consistent with previous reports that DHAP is more hydrophobic than most other commercially available matrices [214], necessitating organic solvent-based removal of the matrix to prevent interference with downstream hybridisation reactions. Following fixation and matrix removal, tissue sections underwent padlock probe–mediated rolling circle amplification for detection of X-CODE RNA barcodes. Fluorescent probes targeting BUs 1–27 were applied alongside β-actin as a protocol control, and signals were visualised on the Akoya PhenoCycler platform using the previously optimised X-CODE imaging protocol.

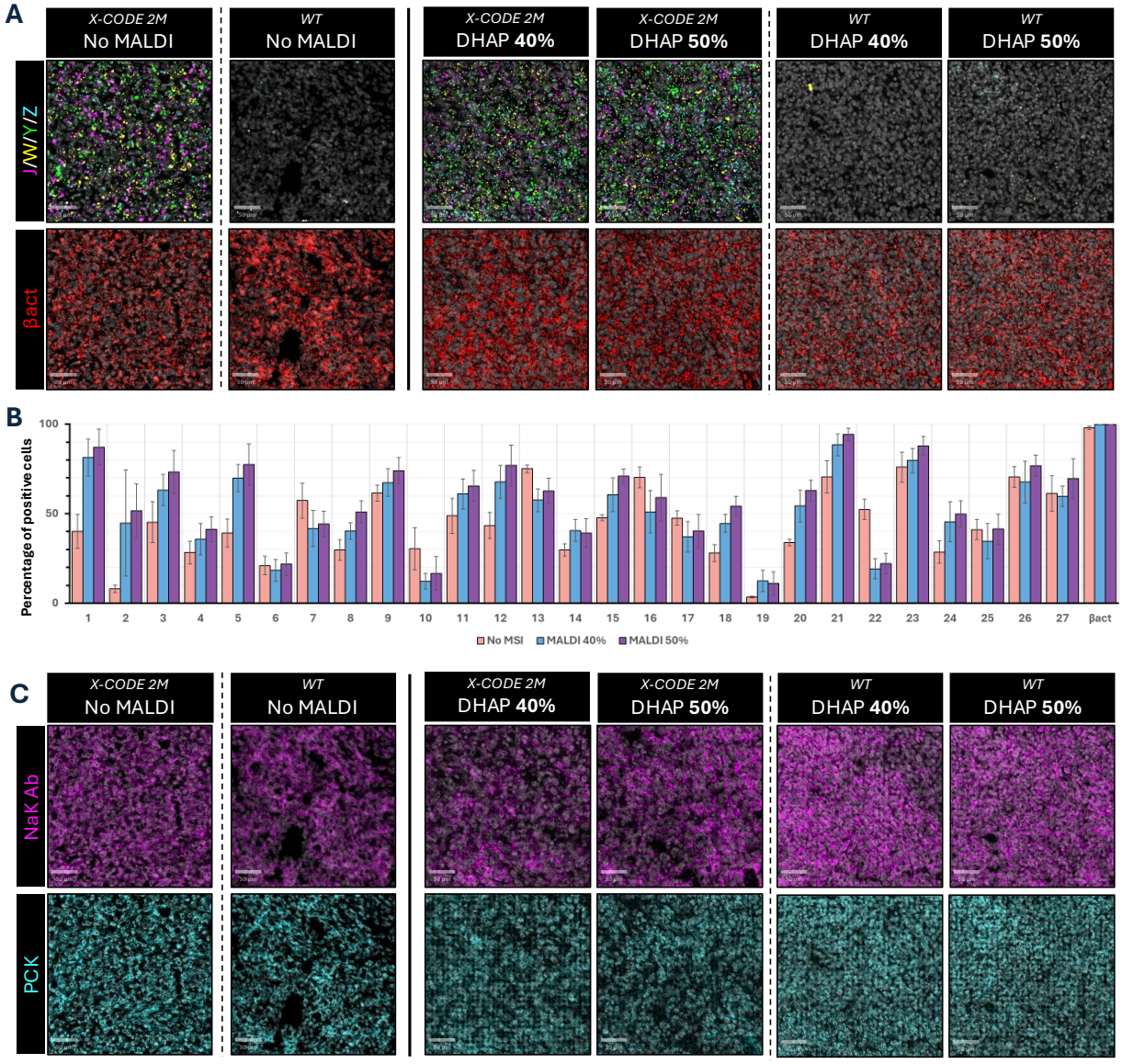


Figure 7-2. Assessing detection of X-CODE barcode units and antibody staining following MALDI-MSI using DHAP matrix.

A) *RM-1 X-CODE JWYZ subcutaneous tumour sections, embedded in HPMC/PVP hydrogel, and their no-barcode RM-1 wildtype (WT) controls underwent MALDI-MSI at laser powers of 40% and 50% using a 2,5-dihydroxyacetophenone (DHAP) matrix. Consecutive tissue sections were processed post-MALDI using the post-*

MALDI X-CODE protocol. Consecutive no-MALDI controls were processed using the standard X-CODE protocol for HPMC/PVP hydrogel-embedded fresh-frozen tissues. X-CODE barcode units (BUs) 1–27 were detected using the Akoya PhenoCycler-Fusion platform. RM-1 2M clonal populations are shown across conditions, visualised by distinguishing BU sequences: J (pink, BU3&10); W (yellow, BU6&17); Y (green, BU14&18); and Z (blue, BU4&7). DAPI staining for nuclei (white) and β-actin RNA staining (red). Scale bars, 50 µm.

***B)** Quantification of BUs 1–27 and β-actin RNA amplicon spots per cell across conditions using the QuPath-based spot-counting pipeline. Bar graph showing the percentage of positive cells (spots > 0) detected across samples without having undergone MALDI (no-MSI), or after MALDI-MSI at 40% and 50% laser power. Error bars indicate standard deviation across ROIs (n = 5).*

***C)** Representative images of sodium–potassium ATPase (Na/K-ATPase, i.e. NaK) and pan cytokeratin (PCK) antibody-based protein detection across conditions.*

Representative images demonstrated robust detection of X-CODE barcode units and β-actin following MALDI-MSI at both laser power settings, with signal distributions comparable to the no-MALDI control (**Figure 7-2**A). To assess whether MALDI-MSI or DHAP matrix removal introduced non-specific probe binding, RM-1 wild-type tumour sections lacking X-CODE barcodes were processed under identical conditions and stained using the same probe set. Minimal non-specific signal was observed across both laser power settings, indicating that integrating MALDI-MSI does not contribute to detectable background hybridisation (**Figure 7-2**A).

Quantitative analysis using the QuPath-based spot-counting pipeline was used to assess the impact of increasing laser power from 40% to 50% on X-CODE detection across BUs 1–27 and β-actin, relative to no-MALDI control sections (**Figure 7-2**B). Assessment of β-actin RNA signal detection following MALDI-MSI, compared with the no-MALDI setting, revealed near-universal cellular detection across conditions, regardless of laser power (**Figure 7-2**B, β-act). This indicated substantially improved RNA preservation relative to DHB matrix conditions. Across several BUs, the percentage of positive cells increased slightly with increasing laser power relative to the no-MSI control. Also, BU-specific effects were apparent: barcode sequences with lower baseline signal under standard X-CODE conditions, such as BU10, tended to exhibit greater sensitivity to MSI-induced signal reduction following MALDI.

In addition to RNA detection, the compatibility of post-MALDI tissue sections with antibody-based protein detection was preliminarily assessed by staining for extracellular sodium–potassium ATPase (Na/K-ATPase) and pan-cytokeratin (PCK). Representative images showed consistent detection of both proteins under both no-MALDI and post-MALDI conditions (**Figure 7-2**C). However, variable MALDI-MSI–induced tissue disruption was visually apparent at both laser power settings, resulting in a reduced delineation of cellular boundaries compared to no-MALDI controls.

Taken together, these results demonstrate that use of a DHAP matrix and lower MALDI laser powers substantially improves post-MSI RNA preservation. These conditions enable X-CODE barcode detection following MALDI-MSI, although BU-specific sensitivity to laser exposure remains evident. Nevertheless, rolling-circle–mediated RNA amplification enables consistent and clear detection of all BUs used in the 15K- and 2M-plex X-CODE libraries in post-MALDI settings. These findings highlight the importance of selecting the appropriate matrix and optimising acquisition parameters to enable integrated MSI and X-CODE workflows.

7.3 Balancing X-CODE RNA detection with MALDI-MSI acquisition parameters

Once MALDI-MSI conditions were optimised for consistent downstream X-CODE RNA detection, their effectiveness in generating reliable MSI data was assessed. Fresh-frozen RM-1 15K HPMC/PVP hydrogel-embedded tumours were analysed to evaluate the compatibility between the two modalities. MALDI-MSI was performed across a broader range of laser power settings to determine the minimum acquisition parameters required to obtain meaningful MSI signal while retaining X-CODE detectability. Four consecutive 10 µm tissue sections were analysed by MALDI-MSI using a DHAP matrix at laser powers of 30%, 40%, 50%, and 60%. Then, James McKenzie from the Takats Lab at Imperial College London analysed the MALDI-MSI data to assess signal quality across acquisition conditions.

After preprocessing MSI data, approximately 1,600 spectral features were detected across the four samples. Visualisations of principal component analysis (PCA) showed that datasets acquired at laser powers of 50% or higher exhibited coherent spatial structure, whereas those at lower laser powers produced less informative spatial patterns, particularly at 30% (**Figure 7-3**A). Inspection of feature sparsity across acquisition conditions indicated that reduced interpretability at lower laser powers was primarily driven by insufficient ionisation. At 30% laser power, the most intense 1% of detected variables exhibited a median sparsity below 50%, reflecting limited and uneven signal across tissue pixels (**Figure 7-3**B, left). As laser power increased, feature sparsity decreased, consistent with improved ionisation efficiency and more comprehensive molecular coverage across the tissue section (**Figure 7-3**B, right). These findings

indicate that lower laser power settings (30%) were insufficient to generate robust MALDI-MSI datasets under the conditions tested.

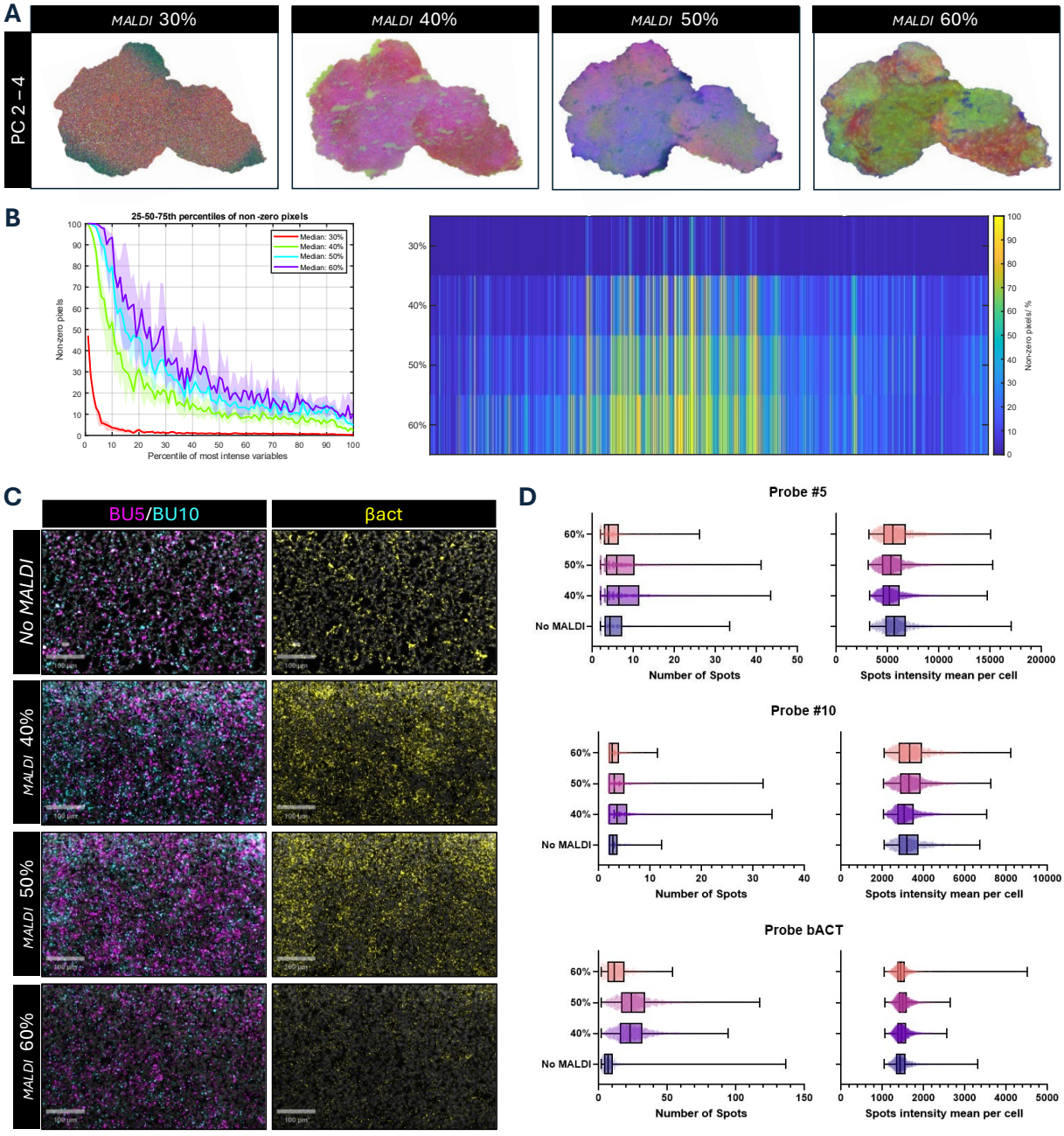


Figure 7-3. Balancing MALDI-MSI data quality with X-CODE RNA detection across laser power settings.

A) *HPMC/PVP hydrogel-embedded RM-1 X-CODE 15K subcutaneous tumour sections underwent MALDI-MSI at laser powers of 30%, 40%, 50%, 60% using a 2,5-dihydroxyacetophenone (DHAP) matrix. MALDI PCA composite maps (PC2–4) highlight spectral variance patterns across acquisition parameters.*

B) *Left: Sparsity distribution, where variables are ranked by intensity and grouped by centile. The X-axis shows the top-ranked variables, and the Y-axis shows the percentage of non-zero pixels, reflecting signal occupancy. Right: Heatmap showing feature sparsity across MALDI laser power settings (30%, 40%, 50%, 60%). Each column represents a variable, and each row corresponds to a MALDI condition. Colour coding indicates the percentage of non-zero pixels per condition, reflecting signal occupancy.*

C) *Post-MALDI or no-MALDI consecutive tissue sections were processed using the optimised MALDI–X-CODE or X-CODE only sample preparation protocols, respectively. Representative X-CODE BU5 and BU10 were detected using fluorescence imaging. Images show detected signal for BU5 (magenta) and BU10 (cyan) across laser powers,*

compared to an X-CODE-only control. DAPI staining visualising nuclei (white) and β-actin RNA staining as a protocol control (yellow). Scale bars, 50 µm.

***D)** Box plot showing the number of detected spots per BU, and mean spot intensity per cell, across conditions.*

To confirm that X-CODE RNA detection remained compatible with acquisition settings that produced informative MSI data, X-CODE detection was repeated using the optimised MALDI–X-CODE workflow on tissue sections acquired at 40, 50, and 60% power. Two representative BUs and β-actin RNA detection were visualised across conditions using fluorescence imaging. Representative images demonstrated robust detection of X-CODE BU5 and BU10, and β-actin, following MALDI-MSI at laser powers of 40–60%, with signal comparable to that of no-MALDI controls (**Figure 7-3**C). QuPath-based analysis of spot counts and mean probe intensity demonstrated that X-CODE probe detection was preserved at laser powers of 40–60% (**Figure 7-3**D), consistent with the RNA preservation observed under DHAP matrix conditions in the preceding section.

Overall, these results indicate that laser intensity settings of at least 50% are a practical operating power for integrating MALDI-MSI, with a DHAP matrix, and downstream X-CODE RNA detection. Under these conditions, MALDI-MSI acquisition yields coherent spatial molecular information while maintaining sufficient RNA integrity to support reliable rolling circle–mediated barcode detection.

7.4 Discussion

This chapter aimed to assess the feasibility of combining X-CODE spatial RNA detection with unbiased mass spectrometry imaging on the same tissue section, and to identify the experimental conditions needed to obtain reliable data from both methods. Collectively, the results demonstrate that while rolling circle–mediated amplification enables robust RNA signal generation under standard conditions, it cannot compensate for extensive RNA degradation caused by high MALDI laser energies. When performing MALDI-MSI with a DHB matrix, increasing laser power resulted in pronounced signal reductions, indicating that RNA integrity is severely compromised under aggressive acquisition conditions. These results align with earlier MALDI-FISH studies [152,206] and highlight the main challenge of balancing MSI acquisition parameters with preserving RNA integrity for subsequent *in situ* detection.

Comparison of post-MALDI sample preparation strategies further indicated that MALDI-induced tissue disruption may reduce the requirement for multi-step fixation and permeabilisation typically used for X-CODE detection in fresh-frozen tissues under standard conditions. In particular, fixation-only approaches appeared to better preserve residual RNA following MALDI-MSI than the standard X-CODE preparation, which includes additional methanol and detergent-based permeabilisation. This indicates that, following MALDI, tissue sections may require tailored preparation conditions, as excessive permeabilisation could increase RNA loss rather than improve probe accessibility. These findings highlight the importance of developing RNA detection protocols specifically for post-MALDI samples, rather than directly transferring workflows optimised for standard conditions.

Using a DHAP matrix with lower MALDI laser power significantly enhanced RNA preservation compared with DHB-based acquisition. Under these optimised conditions, X-CODE detection of barcode units (BUs) 1–27 was retained following MALDI-MSI in tumours, with near-universal β-actin detection and minimal non-specific probe binding in wild-type controls. Although BU-specific sensitivity to laser exposure was evident, padlock-probe-based rolling-circle amplification enabled clear detection across all library BUs. These results highlight matrix selection, in conjunction with optimised sample preparation workflows, as a critical determinant of post-MSI RNA detectability. Also, in this context, they show that DHAP-based MALDI-MSI provides a more permissive setting for integrated workflows than the DHB-based conditions that were tested.

Systematic evaluation of DHAP-based MALDI-MSI data quality across a range of laser powers showed that acquisition parameters compatible with X-CODE detection may not produce informative MSI results. Analysis of MALDI-MSI datasets revealed that laser powers below 50% produced sparse, spatially uninformative molecular signals, whereas acquisitions at 50% or higher yielded coherent spatial patterns with reduced feature sparsity. Importantly, X-CODE RNA detection remained robust at these higher laser powers when DHAP matrix conditions were used. Together, these findings support using laser powers of at least 50% as a practical operating regime for combining MALDI-MSI with downstream X-CODE detection, balancing molecular coverage with retained RNA integrity.

RNA detection has previously been integrated with MSI as reported by the MetaFISH workflow, which combines MALDI-MSI with traditional fluorescence *in situ* hybridisation (FISH) to spatially associate metabolites and microbial identities in host–microbe systems [152,206] . More recently, the OpenFISH workflow has demonstrated post-MALDI RNA detection in mammalian tissues, using padlock-probe–based rolling-circle amplification to detect hundreds of endogenous transcripts by fluorescence imaging [156] . In parallel, several studies have combined MALDI- or DESI-MSI with commercial spatial transcriptomics platforms to achieve transcriptome-wide RNA detection following MSI acquisition [153–155] . While these methods are advanced, MSI–X-CODE workflows offer an accessible alternative strategy focused on scalable clonal identification, compatible with multiplexed imaging, for integrating lineage information and unbiased molecular imaging.

Moreover, to our knowledge, this work represents the first demonstration of cellular barcode detection for lineage tracing after MSI acquisition using a high-plex imaging platform. By integrating X-CODE detection with MALDI-MSI on the Akoya PhenoCycler-Fusion, this study extends existing MSI–RNA workflows to include scalable clonal identity readouts. Additionally, preliminary results on the compatibility of post-MALDI tissue sections with antibody-based protein detection suggest that multimodal phenotyping beyond RNA measurements alone may be feasible within the same tissue section. Previous work has demonstrated fluorescence and mass cytometry imaging of antibody staining following DESI- and MALDI-MSI [215–217] . While these reports provide robust evidence of MSI–immunophenotyping, further work will be required to unite MSI molecular analysis with robust protein and RNA detection on the same tissue section [200] .

The observed variability in post-MALDI X-CODE BU signal could pose a challenge to accurate barcode retrieval if RNA detection were substantially reduced. The disparity in signal likely reflects intrinsic differences in sequence composition and baseline detection efficiency, as previously seen under standard X-CODE conditions in spatial and suspension settings. BUs with lower baseline signal appeared more sensitive to laser-induced degradation, resulting in greater signal reductions following MSI acquisition. The OpenFISH workflow alleviates RNA loss by capturing transcripts with poly(A) probes post-MSI and embedding them in a polymerised hydrogel matrix prior to imaging [156] . While this strategy provides strong RNA preservation, it involves extensive

enzymatic digestion and polymerisation steps that limit compatibility with additional phenotyping modalities, including antibody-based protein detection and downstream histological staining.

By contrast, the MSI–X-CODE protocol prioritises workflow flexibility and compatibility with antibody-based or morphological phenotyping. The initial protein-detection results here suggest that post-MALDI tissues retain sufficient antigenicity for antibody phenotyping, although higher laser power may reduce the spatial delineation of the staining. Nevertheless, if RNA detection alone post-MSI were the primary objective, adoption of RNA preservation strategies, such as those applied in OpenFISH, could further enhance X-CODE signal following MSI. Combining such approaches with high-parameter imaging platforms may represent a promising direction for optimisation. Further improvements to MSI–X-CODE workflows could explore additional matrices and alternative ionisation modes [218], or RNA-protective acquisition strategies, including post-ionisation approaches such as MALDI-2 [136]. Systematic evaluation of these parameters may further improve molecular coverage while preserving RNA integrity, extending the applicability of combined MSI and clonal imaging workflows.

Finally, inherent differences in MALDI matrix chemistry, which dictate the detection of analyte classes, should be acknowledged. DHAP matrices are commonly associated with efficient ionisation of lipids and small metabolites [219], whereas DHB is more widely used as a general-purpose matrix that supports a broader range of analytes, including peptides, lipids, and other low-molecular-weight compounds [220]. The OpenFISH pipeline employs DHB-based MALDI at 90% laser intensity [156], likely to maximise MSI signal, whereas the results presented here demonstrate that DHAP-based acquisition provides a more favourable compromise for RNA integrity without requiring additional preservation stages. These observations reinforce that matrix selection and acquisition parameters for specific samples is crucial to balance MSI performance with downstream molecular detection.

Collectively, the results presented here define a practical framework rather than a universally optimal solution for integrating MALDI-MSI with X-CODE detection. While specific acquisition parameters and matrix choices were systematically explored for the tested tissue types, the principles outlined in this chapter provide a generalisable

strategy for balancing MSI data quality with downstream RNA detection across diverse experimental settings.

7.4.1 Limitations

This chapter focused on establishing the feasibility of integrating X-CODE RNA detection with MALDI-MSI on fresh-frozen tissue sections and defining acquisition parameters compatible with both modalities. While robust fluorescence-based X-CODE detection was demonstrated following MALDI-MSI, compatibility with downstream imaging mass cytometry for barcode detection was not assessed. Similarly, RNA detection following MALDI-MSI was not evaluated in formalin-fixed paraffin-embedded (FFPE) tissues, where extended preparation steps may further impact RNA integrity [221] .

BU–specific sensitivity to MALDI-MSI was observed, where X-CODE sequences that exhibit lower baseline signals under standard conditions show greater vulnerability to signal reduction following MSI acquisition. This sequence-dependent variability may affect debarcoding robustness in post-MALDI settings, so careful calibration of spot-detection thresholds is required when analysing post-MSI X-CODE data. In addition, an increase in background autofluorescence was qualitatively observed in the PhenoCycler channel detecting Alexa Fluor 488-conjugated probes following MALDI-MSI, necessitating more conservative thresholding during spot quantification; however, this effect was not systematically quantified in this chapter.

The integration of workflows to perform MALDI-MSI, RNA detection, and antibody-based protein phenotyping on a single tissue section was preliminarily explored. While post-MALDI tissues retained sufficient antigenicity for staining with two cytoplasmic markers, a systematic evaluation of multiplexed antibody panels in combination with MALDI-MSI and X-CODE detection was not performed. Achieving fully integrated multimodal workflows incorporating MSI, RNA, and protein detection will require careful optimisation of fixation, permeabilisation, and staining protocols [200].

The results presented here also highlight inherent constraints associated with MALDI matrices and acquisition parameters. Different matrices favour distinct classes of analytes [222], and the choice of DHAP in this study prioritised RNA preservation over broader molecular coverage achieved with DHB. Consequently, the molecular species

detected by MSI are influenced by matrix selection and laser power, and the operating settings evaluated here may not be optimal for all experimental objectives. In addition, MALDI acquisition at laser powers above 60% was not evaluated for DHAP matrix conditions, and it remains possible that higher laser energies could further enhance MSI signal while still retaining sufficient RNA integrity for X-CODE detection.

Computational alignment of MSI and fluorescence imaging data was not performed. The pixel-level differences between MALDI-MSI and fluorescence imaging data complicate their alignment, which is important to consider when analysing and interpreting integrated datasets [200]. Accurate co-registration of modalities with differing spatial resolutions remains challenging and requires specialised computational approaches outside the scope of this chapter [106]. Several dedicated frameworks have been developed to address multimodal spatial alignment [104,143]. The future application of such methods will be essential for fully integrated analysis of MSI and X-CODE datasets.

Finally, this chapter did not address detailed annotation or validation of specific molecular features detected by MALDI-MSI. While MSI enables spatially resolved detection of molecular signals, compound identification typically requires additional analytical steps, such as tandem mass spectrometry or orthogonal chromatography-based validation [152] , which is outside the scope of this work. Accordingly, the MALDI-MSI data generated here were used to assess signal quality and spatial coherence under different acquisition conditions, rather than to interpret specific metabolite identities.

7.4.2 Conclusions

This chapter demonstrates the feasibility of integrating X-CODE–based spatial RNA detection with unbiased mass spectrometry imaging on the same tissue section. By systematically evaluating MALDI matrix selection, acquisition parameters, and post-MSI RNA detection protocols, this work defines experimental conditions that preserve both molecular imaging and clonal barcode detection. More broadly, this chapter extends X-CODE detection beyond fluorescence and imaging mass cytometry–based phenotyping towards integration with unbiased molecular imaging. By linking clonal identity to spatial metabolic and biochemical landscapes, the integrated workflow presented here

provides a foundation for future multimodal studies seeking to resolve lineage, phenotype, and molecular state within intact tissue sections.

Chapter 8 Application of X-CODE to an *in vivo* model of prostate cancer evolution

8.1 Research in context

Prostate cancer is initially an androgen-driven disease, with tumour growth dependent on androgen receptor (AR) signalling. Following recurrence after definitive local therapies such as radical prostatectomy or radiotherapy, androgen deprivation therapy (ADT) is the standard systemic treatment for advanced disease and typically induces tumour regression [21] . However, despite initial responsiveness, a substantial proportion of patients ultimately develop resistance. At this stage, the disease progresses to a castration-resistant state (CRPC), characterised by continued tumour growth despite low circulating androgen levels. CRPC exhibits considerable phenotypic and molecular plasticity, characterised by altered AR signalling and heterogeneous lineage adaptation. Once metastatic and castration-resistant, prostate cancer remains incurable, underscoring the clinical need to understand the mechanisms underpinning therapy response and relapse [223] .

Research shows that CRPC encompasses a spectrum of phenotypic and molecular adaptations rather than a single uniform state. During progression under sustained androgen deprivation, tumours can transition from an AR-driven luminal adenocarcinoma phenotype toward mixed or hybrid states, characterised by reduced AR dependence and increased lineage plasticity. These intermediate states may include basal-like or stem-like programmes. Continued selective pressure can promote further divergence toward androgen-indifferent phenotypes, including neuroendocrine (NE) prostate cancer, which is typically marked by loss of AR expression and acquisition of neuroendocrine features. Consequently, CRPC represents a heterogeneous evolutionary landscape in which luminal, basal-like, proliferative, and neuroendocrine-like states may coexist within the same tumour, reflecting diverse adaptive responses to therapy [30,224] .

The MyC-CaP cell line is a murine prostate cancer model originally derived from a prostate tumour in a *c-myc* transgenic mouse [225] . These tumours arise in the Hi-Myc transgenic model, in which human *MYC* is expressed under the control of the androgen-

responsive ARR2PB probasin promoter, thereby linking oncogenic initiation to androgen signalling within the prostate epithelium [226] . The line is characterised by epithelial morphology and an amplified androgen receptor (AR) expression, despite no prior exposure to androgen withdrawal. These cells retain AR expression and demonstrate androgen-dependent proliferation *in vitro* and following implantation into mice. They perform as a clinically representative model of androgen-sensitive prostate cancer prior to castration-induced regression, followed by subsequent relapse under androgen deprivation therapy (ADT) conditions [225] .

The well-defined transition of MyC-CaP tumours from androgen-sensitive to hormone-refractory growth under continued androgen deprivation provides a reproducible temporal framework for interrogating prostate cancer evolution *in vivo* [225,227,228] . MyC-CaP cells have been widely used in preclinical research to investigate mechanisms of the androgen-deprivation response and resistance in immunocompetent hosts. The MyC-CaP model has been used to demonstrate clinically relevant immune-mediated resistance to ADT [227] and immune-modulation therapies [229]. The MyC-CaP line has also been utilised in an orthotopic mouse model, recapitulating primary prostate cancer, to investigate resistance to ADT [230] . Additionally, metastatic disease models have used intracardiac injection to study prostate cancer bone metastasis [231], reflecting the diverse utility of this model for examining prostate cancer biology.

The MyC-CaP prostate cancer mouse model offers a balance of androgen dependence, immune competence, and phenotypic relevance, making it particularly suitable for testing novel multimodal analytical frameworks. The subcutaneous MyC-CaP model provides a rapid, reproducible, and well-controlled experimental system in which tumour burden, clonal composition, and spatial organisation can be monitored longitudinally and attributed directly to androgen deprivation. These features make it particularly well-suited as a proof-of-principle setting for evaluating the X-CODE framework for *in vivo* clonal investigation of cancer evolution during therapy, rather than for the detailed dissection of specific biological mechanisms.

This chapter applies the X-CODE framework to a subcutaneous MyC-CaP model of prostate cancer evolution under androgen deprivation. Clonal dynamics are first assessed longitudinally during therapy response and relapse using next-generation sequencing, and clonal organisation is subsequently mapped *in situ* within tumours.

Multimodal spatial phenotyping is then integrated to analyse clone-associated protein expression and metabolic features. Collectively, this chapter demonstrates the feasibility of applying X-CODE in a complex *in vivo* therapeutic setting and establishes a framework for integrated clonal, spatial, and phenotypic analysis rather than aiming to comprehensively define the underlying tumour biology.

8.1.1 Chapter aims

- To apply the X-CODE framework in an *in vivo* therapeutic setting to investigate clonal dynamics during cancer evolution.
- To develop a computational spatial debarcoding pipeline for assigning X-CODE clonal identities *in situ*.
- To integrate spatial multimodal phenotyping to characterise clonally associated cellular features.

8.2 Assessment of clonal dynamics in a MyC-CaP tumour model during androgen deprivation

To utilise X-CODE in an *in vivo* therapeutic setting, clonal dynamics were examined during androgen deprivation in the MyC-CaP mouse model of prostate cancer. MyC-CaP cells were labelled with the X-CODE 15K library and implanted into immunocompetent mice. Tumours were sampled longitudinally before and after Degarelix-induced chemical castration. The tumours were processed to retain maximum versatility for downstream experiments, with portions allocated for DNA extraction, cell culture, and line establishment. Additionally, the central tumour region was halved, fresh-frozen, and stored for spatial analyses. This enabled tumour slices to be embedded in OCT or in an HPMC/PVP hydrogel, depending on the downstream spatial analysis techniques (**Figure 8-1**A). This design enabled comprehensive longitudinal analysis of clonal behaviour across multiple assays using matched samples, providing a defined temporal framework for studying tumour response to androgen deprivation and the development of castration-resistant disease.

Roberto Rota, in the Tumour Functional Heterogeneity Lab at the Institute of Cancer Research, monitored the *in vivo* experiment and analysed tumour growth patterns. Consistent with the established androgen dependence of the MyC-CaP model [225],

tumour growth dynamics were significantly altered after castration. Tumours underwent an initial phase of regression following androgen deprivation at days 4 and 10 post-castration, indicating an effective therapeutic response, before relapsing at later time points. In contrast, tumours in untreated control mice continued to expand progressively over the same period. These growth patterns created a distinct temporal separation between treatment response and relapse, providing the basis for subsequent analysis of clonal behaviour across pre-castration, early post-castration, and castration-resistant tumour states (**Figure 8-1**B).

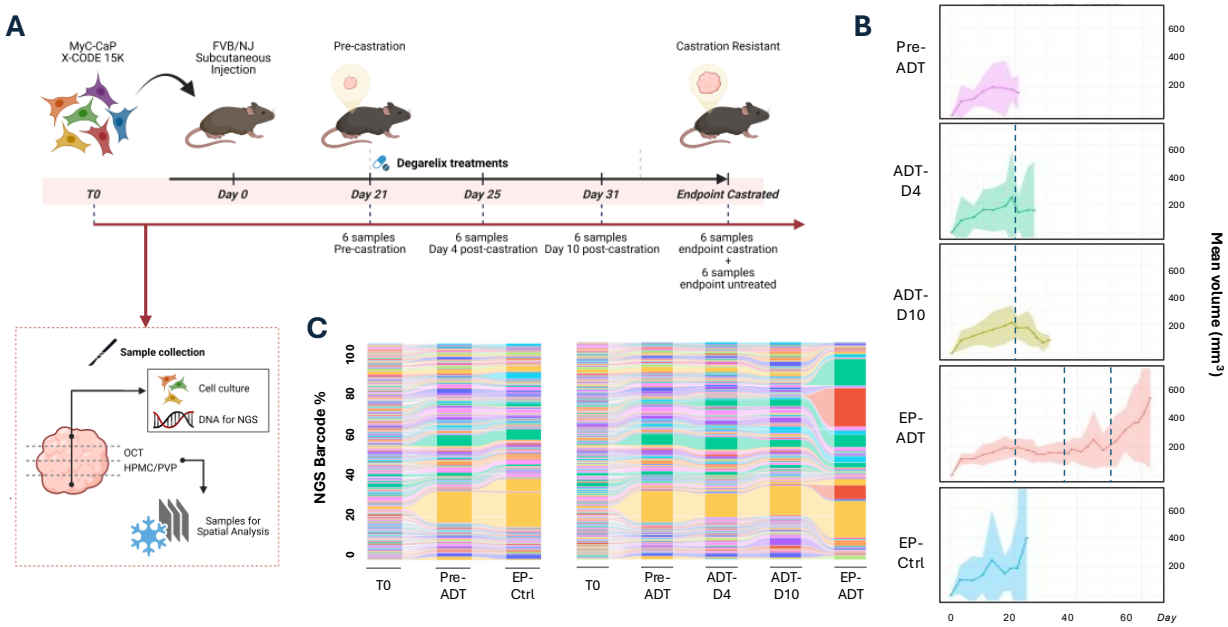


Figure 8-1. Investigating clonal dynamics using X-CODE in a MyC-CaP mouse model of prostate cancer during androgen deprivation.

A) *Schematic of experimental design for longitudinal clonal tracking in the MyC-CaP model. MyC-CaP cells transduced with the 15K X-CODE library (T0) were injected subcutaneously into FVB/NJ mice. Tumours were collected at 4 days and 10 days post-castration with Degarelix, and at the experimental endpoint under castrated or untreated conditions. Tumours were divided to retain material for DNA barcode sequencing, cell culturing, and in situ spatial analysis. Created with BioRender.com.*

B) *Tumour growth curves across experimental groups showing mean tumour volume (n = 5 – 6) over time for each experimental group. Shaded areas indicate 95% confidence intervals.*

C) *Alluvial plots showing barcode composition dynamics derived from debarcoding of DNA sequencing (NGS) data. Each bar represents the normalised barcode frequencies calculated from pooled data across all tumour samples within the corresponding experimental group.*

Clonal dynamics were initially characterised using DNA-level barcode next-generation sequencing (NGS) across all tumours and time points by Roberto Rota. Analysis of barcode composition revealed highly similar clonal representations across pre-castration, early post-castration (day 4 and day 10), and untreated endpoint samples (**Figure 8-1**C, top). In contrast, castration-resistant endpoint tumours exhibited

pronounced clonal remodelling, characterised by the selective expansion of a restricted subset of clones (**Figure 8-1**C, bottom). Notably, these expanding clones were not substantially enriched at early post-castration time points, indicating that resistance was not associated with immediate selective outgrowth following treatment. Instead, these findings are consistent with differential clonal fitness emerging under sustained androgen deprivation, reflecting heterogeneity in the adaptive potential of different clonal lineages.

8.2.1 Spatial mapping of X-CODE barcodes reveals clonal organisation during tumour evolution

Based on clonal dynamics identified by sequencing, a minimum of two tumour samples per time point were selected for spatial profiling to detect X-CODE sequences *in situ*. The chosen tumour samples were embedded in HPMC/PVP hydrogel multi-blocks according to previously described protocols [165] . Multi-blocks were cryosectioned at 10 µm, and tumour sections were processed using the X-CODE spatial detection protocol (**Figure 8-2**A). Samples were stained for barcode units (BUs) 1–18 using X-CODE fluorophore detection probes (**Supplementary Table 10**) and read out using the Akoya PhenoCycler-Fusion. The tissue sections were then stained with haematoxylin and eosin (H&E) following PhenoCycler analysis (**Figure 8-2**B, top). Representative images of samples from each experimental time point, with randomly coloured BU signal visualised across all 18 channels, demonstrate well-defined cellular patches in various colours (**Figure 8-2**B, bottom).

To determine whether the observed variation in these coloured regions from all BU signals reflected underlying differences in clonal identity, a spatial decoding strategy for X-CODE barcodes was developed. Subcellular spots corresponding to all 18 BUs were quantified following cell segmentation using the QuPath-based detection pipeline. Each cell was assigned a set of eight BUs based on the most highly detected signals within its segmentation boundaries. Then, barcode identities were inferred by matching these BU combinations to the X-CODE 15K library whitelist. Cells lacking an initial barcode assignment were reevaluated under a stringent criterion for barcode reassignment. Unassigned cells with minor BU mismatches to a spatially proximal confidently assigned cell were subsequently attributed to the corresponding barcode and grouped accordingly. This debarcoding workflow enabled the generation of spatially resolved

clonal maps, revealing that the distinct BU signal patches observed across tissue sections corresponded to assigned X-CODE clonal identities (**Figure 8-2**C, bottom).

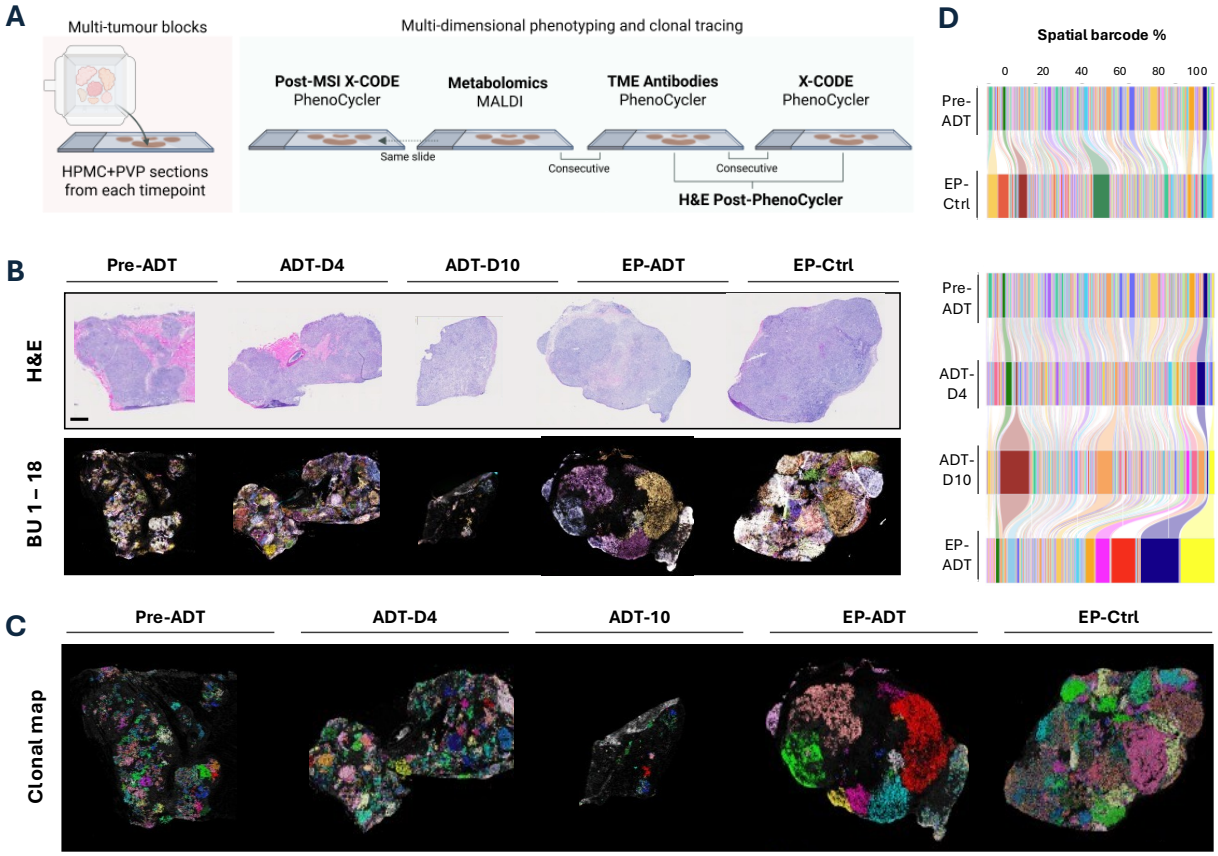


Figure 8-2. X-CODE resolves clonal dynamics spatially during androgen deprivation in vivo.

A) *Schematic representation of the multi-dimensional phenotyping and analysis of clonal dynamics in X-CODE barcoded MyC-CaP subcutaneous tumours. Samples across multiple experimental timepoints following treatment with androgen deprivation therapy (ADT), via chemical castration using Degralix, were embedded into HPMC/PVP hydrogel blocks. Created with BioRender.com.*

B) *Representative tumour sections from samples across the experimental groups. Top: H&E staining showing overall tumour morphology. Bottom: Combined fluorescent signal of barcode units 1 – 18 using X-CODE fluorescent probes detecting the X-CODE 15K library, acquired using the Akoya PhenoCycler-Fusion. Pre-castration = Pre-ADT; untreated endpoint = EP-Ctrl; 4 days post-castration = ADT-D4; 10 days post-castration = ADT-D10; and castration endpoint = EP-ADT. Scale bar, 400 µm.*

C) *Spatially resolved clonal maps following QuPath-based spot detection and computational debarcoding, illustrating distinct spatial clonal distribution within tumours across experimental timepoints.*

D) *Alluvial plots showing barcode composition dynamics derived from debarcoding of spatial fluorescence data. Each bar represents the normalised barcode frequencies calculated from pooled data across all tumour samples within the corresponding experimental group.*

Pre-castration and untreated tumours exhibited a highly heterogeneous clonal architecture, characterised by numerous small clones that were closely packed and often intermingled with nearby lineages. In contrast, castration-resistant tumours displayed a markedly different spatial arrangement, with large contiguous areas dominated by a few clones, resulting in clear spatial separation (**Figure 8-2**C). This

change in spatial structure was reflected in barcode composition variations over time across all samples. Barcode composition remained relatively stable in the untreated tumours, consistent with minimal clonal selection during androgen-driven growth (**Figure 8-2**D, top). Although barcode diversity remained largely preserved at early post-treatment time points following castration, the castration-resistant endpoint tumours were mainly dominated by a limited subset of barcodes (**Figure 8-2**D, bottom). This indicates the expansion of adapted clonal lineages over time rather than immediate selective outgrowth following castration, mirroring the clonal dynamics identified by the DNA-level barcode NGS (**Figure 8-1**C).

Together, these results show that X-CODE is a versatile, comprehensive platform for investigating tumour evolution *in vivo*. The framework supports high-throughput clonal tracking across extensive cohorts and multiple time points, enables informed sample selection for spatial analysis, and provides integrated identification of clonal identities during therapy response.

8.3 Clone-associated phenotypic and metabolic characterisation during tumour evolution under androgen deprivation

8.3.1 Immunophenotypic characterisation of X-CODE–defined clones

Following characterisation of clonal dynamics *in situ*, spatial clonal maps were integrated with tumour phenotyping to identify phenotypic states associated with dominant clones during androgen deprivation. An antibody panel was used to characterise tumour cell identity and the surrounding microenvironment. Epithelial lineage was assessed with cytokeratin (CK) markers: CK8 for luminal epithelial tumour cells, consistent with the known MyC-CaP phenotype [225] , and CK5 for basal-like cells. Androgen receptor (AR) staining was evaluated to assess androgen signalling status, with MyC-CaP tumour cells expected to show predominantly nuclear AR expression before treatment and reduced nuclear localisation after androgen deprivation. Microenvironment components were identified using vimentin (VIM) for stromal fibroblasts, CD11b for myeloid cells (macrophages and neutrophils), and CD45 for lymphoid cells. T cell subsets were distinguished by CD4 and CD8, which mark helper and cytotoxic T lymphocytes, respectively.

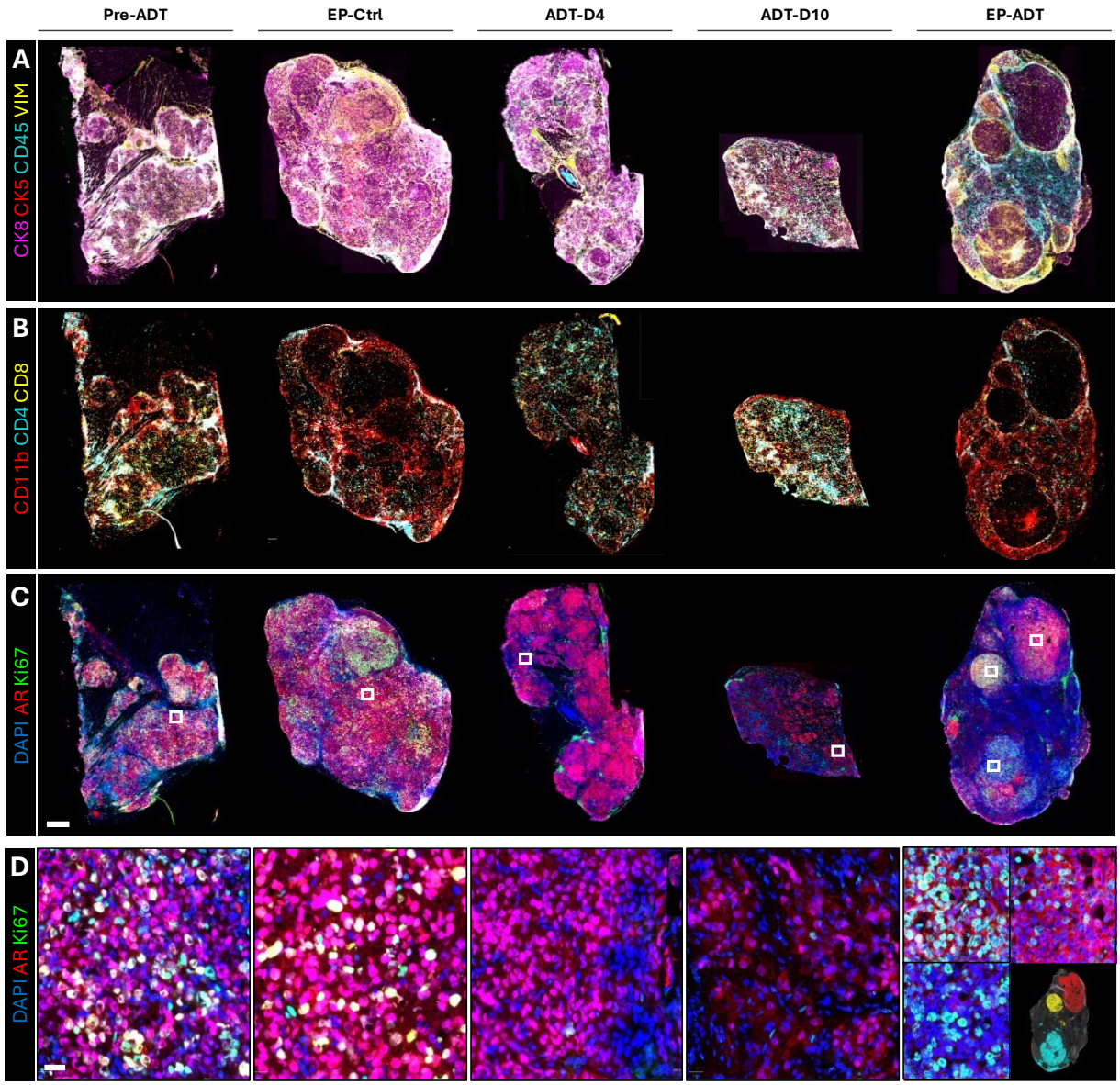


Figure 8-3. Immunophenotyping of MyC-CaP tumour states during tumour evolution under androgen deprivation.

Representative images of multiplexed immunophenotyping of tumour sections, across experimental time points, read out using the Akoya PhenoCycler-Fusion. Pre-castration = Pre-ADT; untreated endpoint = EP-Ctrl; 4 days post-castration = ADT-D4; 10 days post-castration = ADT-D10; and castration endpoint = EP-ADT. Horizontal panels display 3 or 4 marks simultaneously. Scale bar, 500 µm.

***A)** Displaying epithelial state of tumour cells determined by CK8 (luminal, magenta) and CK5 (basal, red); immune cells by CD45 (cyan) and stromal cells by vimentin (VIM, yellow).*

***B)** Displaying myeloid cells by CD11b (red) and T cell subsets were distinguished using CD4 (blue) for helper cells and CD8 (yellow) for cytotoxic cells.*

***C)** Displaying markers for proliferation by Ki67 (green) and androgen receptor (AR, red) with DAPI (blue).*

***D)** Higher-magnification views of selected regions highlighting AR and Ki67 staining. Scale bar, 20 µm. The bottom-right panel shows the spatial clonal map of the three dominant clones identified in the EP-Cast tumour.*

Consecutive sections from the same HPMC/PVP-embedded tumour blocks used to create the clonal maps were stained and analysed to enable spatial alignment of clonal identity with multiplex immune-phenotyping. The Akoya PhenoCycler-Fusion was used

for multiplexed imaging of the stained tumour sections, revealing significant differences in tumour architecture and microenvironmental organisation across treatment conditions (**Figure 8-3**A). Pre-castration and untreated tumours were phenotypically homogeneous, displaying a uniform $CK8^+$ luminal epithelial phenotype with variable but broadly distributed stromal and immune components. In contrast, castration-resistant endpoint tumours displayed significant spatial heterogeneity, with distinct tumour regions showing decreased luminal marker expression. Furthermore, immune and stromal cells were largely excluded from regions dominated by single clonal expansions and instead accumulated at the interfaces between clonally distinct tumour areas (**Figure 8-3**B). In contrast, untreated and early post-castration samples exhibited a widespread distribution of immune cells throughout the tumour, encompassing both myeloid and lymphoid compartments.

Visual analysis of AR and Ki67 showed the expected temporal localisation pattern following androgen deprivation. There was reduced proliferation and cytoplasmic re-localisation of AR at early post-castration time points, followed by re-emergence of proliferative activity in resistant tumours (**Figure 8-3**C). Notably, within a castration-resistant tumour dominated by three clones, distinct spatially resolved AR and proliferation states could be observed, including AR-negative proliferative regions, AR-cytoplasmic proliferative regions, and AR-cytoplasmic low-proliferative regions, indicating clone-associated modes of adaptation (**Figure 8-3**D). Moreover, increased basal marker expression in the castration-resistant sample is consistent with phenotypic plasticity under sustained androgen deprivation (**Figure 8-3**A).

Based on the analysed clonal dynamics, individual clones identified at different time points and enriched at either the untreated or castration endpoints were examined. Clones expressing barcodes 9948, 8385, 8958, 8466, 8147, and 15129 were detected across two or more timepoints (**Figure 8-4**A). Clone 9948 was minimally detected in pre-castration and in early post-castration samples, yet was enriched in two castrated endpoint samples (samples #292 and #285). Similarly, clones 8385, 8958, and 8147 were detected at low levels in pre-castration or early-castration samples but showed enrichment in castrated endpoint samples. In contrast, clone 8466 was detected only in a castrated endpoint at a relatively high proportion relative to other barcodes, whereas clone 15129 was identified solely in an untreated endpoint tumour (#306).

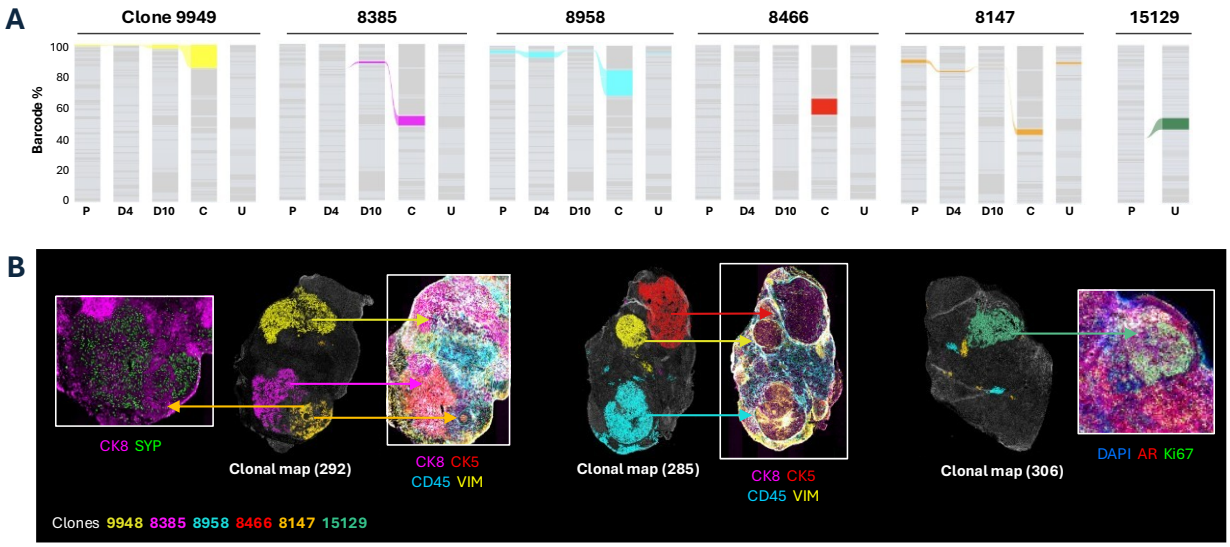


Figure 8-4. Clone-associated phenotypic states during tumour evolution under androgen deprivation.

A) *Clonal dynamics across time points are depicted using alluvial plots showing the relative abundance of selected clones based on pooled data for each experimental group. Clones of interest expressing barcodes 9948, 8385, 8958, 8466, 8147, and 15129.*

B) *Representative images of multiplexed immunophenotyping highlighting the spatial distribution and associated phenotypic features of clones of interest in endpoint-treated tumours (292, 285) and an endpoint-un-treated tumour (306).*

Clones enriched in castration-resistant tumours showed distinct spatial phenotypic associations (**Figure 8-4**B). Clone 9948 showed varied states across tumours, from luminal areas with associated lymphoid infiltration (#292) to immune-excluded zones with reduced CK8 expression (#285). Furthermore, clone 8385 showed a distinct basal phenotype, with higher synaptophysin expression and lymphoid infiltration, which were not observed at earlier time points. Clone 8958, while detected in various conditions, exhibited a specific increase in vimentin-positive cells at the castration-resistant endpoint (#285). Clone 8466, detected exclusively at the castration-resistant endpoint, exhibited a marked loss of epithelial markers (#285).

Taken together, these observations suggest that MyC-CaP dominant clones, which expand during androgen deprivation, exhibit distinct spatial phenotypic arrangements rather than evolving into a single resistant state.

8.3.2 Integration of X-CODE with MALDI–MSI reveals clone-associated metabolic features

To extend clone-associated phenotypic characterisation, matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) was performed on

HPMC/PVP hydrogel-embedded MyC-CaP tumour sections. MALDI-MSI was conducted under conditions that yielded robust metabolic data while maintaining compatibility with downstream X-CODE detection, as previously established in Chapter 7. Tissue sections were coated with a 2,5-dihydroxyacetophenone (DHAP) matrix and analysed at 50% laser power. Following the MSI acquisition, the same tissue section was subjected to X-CODE barcode detection using the optimised post-MSI protocol, enabling evaluation of X-CODE detection performance after MALDI. Importantly, tissue sections processed using the combined MSI–X-CODE workflow were near-consecutive to sections that underwent X-CODE detection alone, as previously analysed in Section 8.2.1.

The workflow established in Chapter 7 provides a flexible framework for post-MSI X-CODE detection that requires further sample- and experiment-specific optimisation, rather than a fixed protocol. Here, the optimised operating regime was applied to assess its performance in a proof-of-principle application using a complete X-CODE library. Subsequently, direct comparison of barcode detection and clonal decoding between MSI-exposed and non-exposed tissue within matched tumour samples was performed. Near-consecutive sections spanning pre-castration, early post-castration, untreated endpoint, and castration-resistant endpoint samples were stained with fluorescent X-CODE detection probes complementary to barcode units (BUs) 1–18 (**Supplementary Table 10**), and barcode signal was acquired using the Akoya PhenoCycler-Fusion platform.

Representative images of the endpoint-castrated sample #285 showed comparable spatial patterns of dominant X-CODE signal areas in the no- and post-MALDI images when BUs were coloured consistently across samples (**Figure 8-5**A, left). Following MSI, BU signal remained spatially specific, although post-MALDI images exhibited a diffuse signal appearance. Additionally, the ability to visually distinguish smaller, distinct regions of the BU-associated signal was reduced after MALDI. β-actin RNA signal also appeared diminished, with a diffuse pattern post-MALDI, indicating an overall decrease in RNA integrity following MSI (**Figure 8-5**A, middle). To assess whether this affected barcode assignment, the previously described computational debarcoding pipeline was applied to generate maps of X-CODE clones. The top 10 most abundant barcodes identified in the X-CODE-only section are highlighted across both conditions (**Figure**

8-5A, right). While barcode identities could still be assigned after MSI, the total number of successfully decoded and barcode-assigned cells decreased.

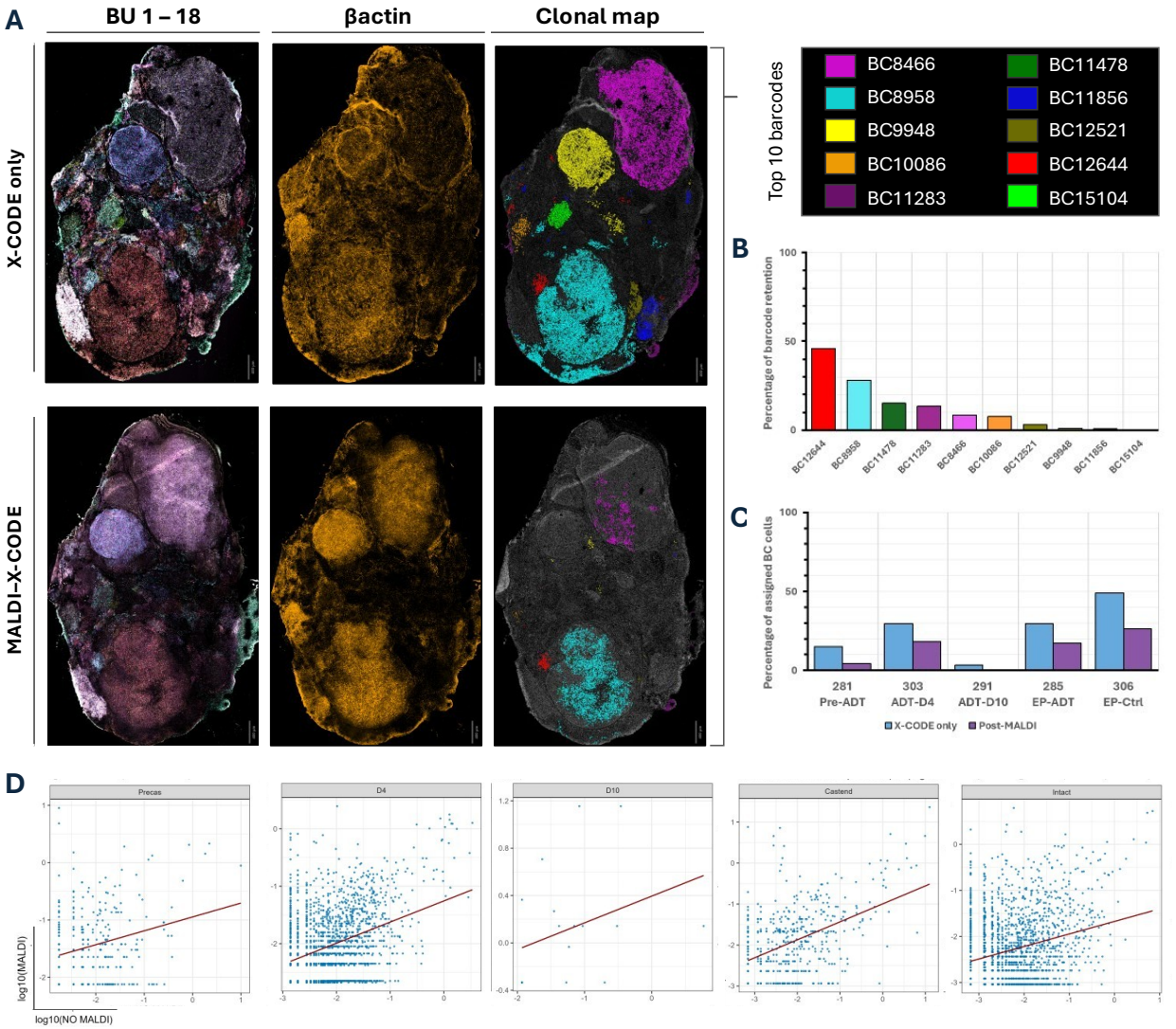


Figure 8-5. X-CODE spot detection and spatial 15K library decoding following MALDI-MSI.

A) *Near-consecutive sections of MyC-CaP underwent X-CODE detection alone, or X-CODE detection following MALDI-MSI using a DHAP matrix at 50% laser power. Both conditions were stained for barcode units (BUs) 1 – 18 using appropriate X-CODE protocols, and the signal was visualised by the Akoya PhenoCycler-Fusion. Representative endpoint castrated sample #285 shows BU 1 – 18 uniquely coloured across both conditions and β-actin RNA staining. The QuPath-based spot detection and decoding pipeline was used to assign segmented cells their corresponding X-CODE 15K barcode; the top 10 quantified barcodes are shown in the colours indicated across both conditions.*

B) *Bar graph showing the percentage of cells in the MALDI–X-CODE #285 sample retaining the barcode assigned in the X-CODE only condition. BC15104 was not detectable in the MALDI–X-CODE sample.*

C) *Bar graph showing the percentage of cells in samples #281, 303, 291, 285, and 306 in the X-CODE only and MALDI–X-CODE conditions that were successfully assigned an X-CODE 15K barcode.*

D) *Comparison of barcode assignments between X-CODE–only and post-MALDI condition across experimental timepoints. Scatter plots show pooled, cell-level comparisons for pre-castration (Precas), 4 days post-castration (D4), 10 days post-castration (D10), castration endpoint (Castrated), and untreated intact endpoint (Intact) samples. Each point represents a cell successfully decoded in both conditions, plotted by its assigned barcode count in the X-CODE–only section (x-axis) versus the post-MALDI section (y-axis). Axes are shown on a log10 scale to visualise the distribution of barcode identities. Red lines indicate linear regression fits summarising the relationship between assignments across conditions.*

9 of the top 10 barcodes quantified in the X-CODE only section were identified in the MALDI-X-CODE sample. The highest barcode retention was observed for clone 12644 at approximately 50%; whereas BC15104 was not computationally decoded in the post-MSI section (**Figure 8-5**B). This reduction aligns with the MALDI laser degrading RNA, causing a loss of BU signals and hindering cellular identity decoding. The decrease in barcode assignment in the post-MALDI sections, relative to the X-CODE-only sample, was also seen in samples #281, 303, 291, and 306 across experimental timepoints. (**Figure 8-5**C). Pooled comparison of barcode assignments from all samples per timepoint across conditions demonstrated that, among successfully decoded cells, barcode identities were largely preserved following MALDI-MSI. Yet, a reduced decoding efficiency is observed, indicating attenuation rather than random reassignment of clonal identities (**Figure 8-5**D). These data indicate that while MALDI–MSI reduces the efficiency of barcode recovery, it does not substantially alter barcode identity among cells that remain decodable.

To circumvent this reduction in post-MALDI barcode recovery for the analyses of clone-associated metabolic features, MSI data were integrated with X-CODE identities derived from non-MSI consecutive sections. Co-registration of spatial clonal maps with metabolic profiles was performed by James McKenzie from the Takats Lab at Imperial College London. Principal component analysis revealed tumour- and clone-specific metabolic patterning, with spatially localised m/z features aligning with dominant clonal regions (**Figure 8-6**A). PC2 – 4 reveal distinct differences between early post-castration and intact samples in comparison to later stage castration and endpoint castration samples (**Figure 8-6**A, middle). Additionally, PC4 – 6 demonstrate specific patterns corresponding to major clonal regions in castrated endpoint samples #285 (BC8466, 8958, 9948), and #292 (BC8147, 8385, 9948) as well as the dominant clone identified in the intact endpoint sample #306 (**Figure 8-6**A, bottom).

Next, specific clone-associated m/z features that were determined to be selectively enriched relative to other clones were examined. Representative ion images of m/z 690.44, 884.657, 833.654, 848.55, 825.562, and 872.333 are shown, illustrating spatially localised metabolic variables associated with the clones of interest (**Figure 8-6**B). For instance, in the castrated endpoint samples, clone 8385 is associated with 833.654 m/z signal intensity in sample #292. Clone 9948 exhibits variable dominant outgrowth in two castrated endpoint samples, showing increased 572.33 m/z signal

intensity in sample #292 but not in sample #285. Finally, tentative variable annotations were also performed using the bulk structure search function of the LIPID MAPS Structural Database (LMSD) [171,172]. These clone-associated features corresponded to lipid species with putative annotations, including phosphatidylserines, phosphatidylglycerols, phosphatidyl-inositols, ceramide-linked phosphorethanolamines, and lysophospholipids. This data indicated that distinct dominant clones are associated with non-overlapping metabolic configurations.

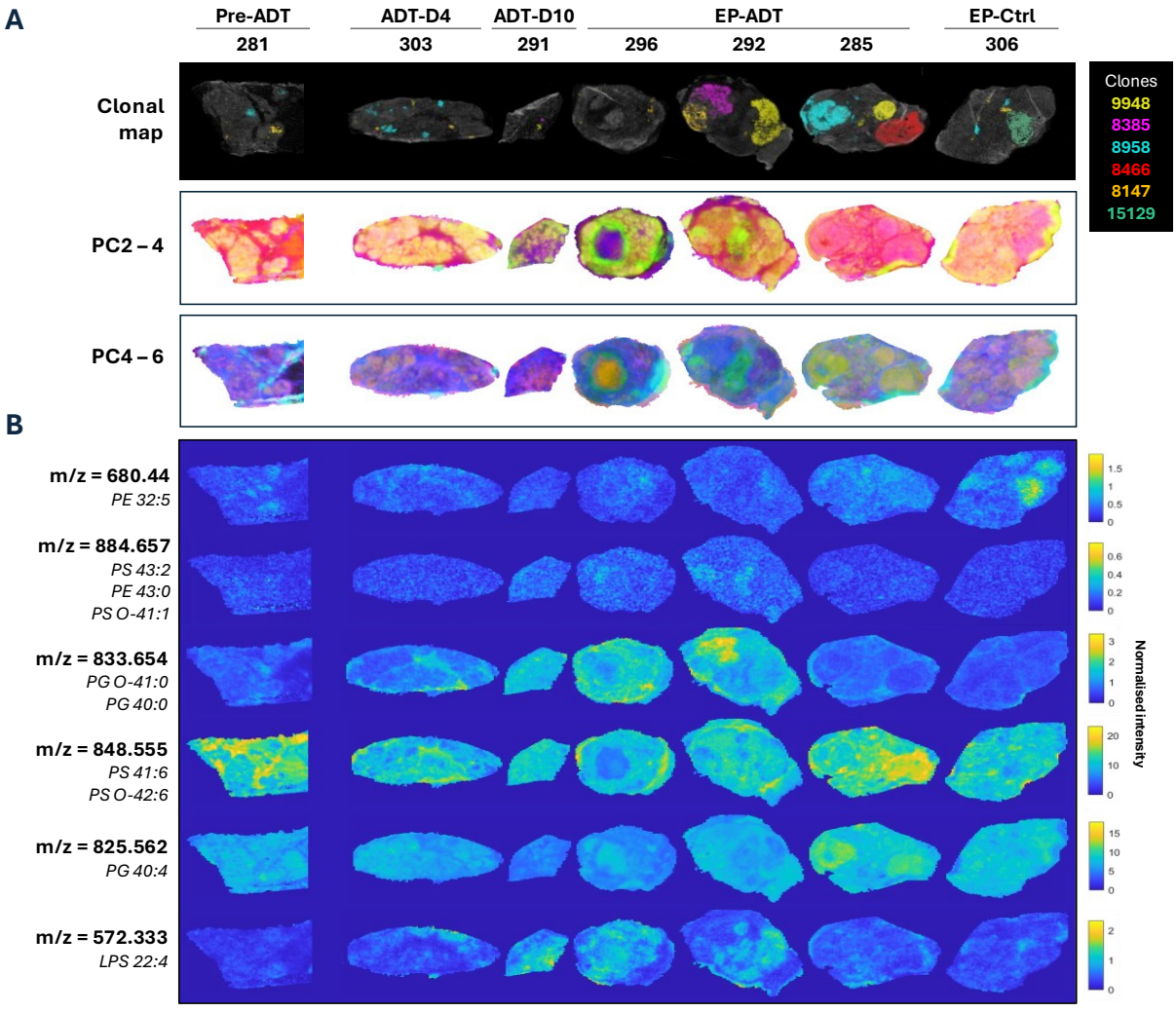


Figure 8-6. Clone-associated metabolic features during tumour evolution under androgen deprivation.

A) *Top: Clonal maps of decoded X-CODE signals for samples across experimental timepoints, colours of clones of interest as indicated. Pre-castration = Pre-ADT; untreated endpoint = EP-Ctrl; 4 days post-castration = ADT-D4; 10 days post-castration = ADT-D10; and castration endpoint = EP-ADT. Middle and bottom: MALDI-derived principal component (PC) maps aligned to tumour sections across time points.*

B) *Representative ion images corresponding to clone-associated m/z features with putative annotations, illustrating spatially localised metabolic variables associated with the clones of interest.*

Collectively, these results establish the feasibility of integrating X-CODE clonal identity with MALDI–MSI data in a therapeutic tumour model. This framework enables the

spatial association of dominant clones with distinct metabolic features across stages of tumour evolution under androgen deprivation.

8.4 Discussion

In this chapter, the X-CODE framework was applied to an *in vivo* MyC-CaP model of prostate cancer evolution under androgen deprivation, enabling longitudinal and spatial interrogation of clonal dynamics during therapy response and relapse. By combining high-complexity clonal barcoding with the X-CODE 15K library, spatial imaging, and multimodal phenotyping, this work demonstrates the feasibility of resolving clonal behaviour within intact tumour tissue across multiple experimental readouts. Together, the results establish X-CODE as a flexible and scalable platform for studying tumour evolution *in vivo*, combining bulk analysis of barcode dynamics with single-cell, spatially resolved clonal, phenotypic, and metabolic information.

Analysis of clonal dynamics during androgen deprivation revealed that castration-resistant tumours arise through pronounced clonal remodelling rather than uniform outgrowth of pre-existing dominant lineages. Early post-castration time points retained high clonal diversity, whereas resistant endpoint tumours were characterised by selective expansion of a limited subset of clones. Importantly, these expanding clones were not strongly enriched immediately after treatment, indicating that resistance is not driven by rapid early selection alone but rather emerges over time under sustained androgen deprivation. These findings are consistent with a model in which differential adaptive potential among clonal lineages shapes long-term tumour evolution during therapy, consistent with previous studies [232] .

A computational strategy for decoding clonal identities *in situ* was developed that used barcode unit detection to assign each cell its combinatorial X-CODE barcode from the 15K library barcode whitelist. Spatial mapping of X-CODE barcodes further demonstrated that clonal expansion in resistant tumours is accompanied by marked reorganisation of tumour architecture. Whereas pre-castration and untreated tumours exhibited highly intermingled clonal patterns, castration-resistant tumours were organised into large, spatially segregated clonal regions. This spatial segregation suggests that clonal dominance is associated with local niche occupation rather than homogeneous mixing, highlighting the importance of spatial context in shaping tumour

evolution. By enabling direct visualisation of clonal organisation *in situ*, X-CODE provides a means to link clonal selection with spatial tumour structure during disease progression.

Integration of immunophenotypic profiling revealed that dominant clones expanding under androgen deprivation exhibited distinct protein expression patterns rather than converging on a single resistant phenotype. Differences in epithelial markers, proliferation-associated proteins, and microenvironmental composition were observed between clones occupying distinct spatial regions. These observations indicate that castration resistance in the MyC-CaP model is phenotypically heterogeneous, with multiple coexisting resistant clones adopting divergent cellular states. These observations challenge simplified models of resistance based on uniform lineage switching or phenotypic convergence and instead support a view of resistance as a multi-lineage evolutionary outcome.

The integration of X-CODE with MALDI–MSI further extended this framework by enabling spatial association of clonal identity with metabolic features. Although MALDI–MSI reduced barcode recovery efficiency, clonal identity was largely preserved among successfully decoded cells. The experimental design enabled integration of clonal maps with metabolic imaging via near-consecutive sections. This approach revealed that dominant clones were associated with distinct, spatially localised metabolic features, predominantly corresponding to lipid species with putative annotations. Importantly, different resistant clones exhibited non-overlapping metabolic profiles, indicating that clonal dominance under androgen deprivation is compatible with multiple metabolic configurations rather than a single convergent programme.

Taken together, these findings support a model in which prostate cancer evolution under androgen deprivation does not converge on a uniform phenotypic or metabolic state, consistent with observations in patients [30]. Instead, resistant tumours comprise multiple dominant clones, each occupying distinct spatial niches and associated with specific phenotypic and metabolic features. This heterogeneity likely reflects differences in clonal adaptive potential and local microenvironmental context, rather than deterministic progression toward a single resistant endpoint. By integrating clonal identity with spatial phenotype and metabolism *in situ*, X-CODE enables direct resolution of this complexity within intact tumours.

More broadly, this chapter demonstrates the utility of static, high-complexity spatial barcoding as a tool for studying tumour evolution *in vivo*. To our knowledge, this represents the first application of a static clonal barcode system enabling direct spatial debarcoding of thousands of unique clonal identities within tissue. The computational debarcoding pipeline and multimodal integration workflow presented here establish a generalisable framework that can be adapted to other tumour or disease models, therapeutic contexts, and spatial readouts. As such, X-CODE provides a foundation for future studies to dissect how clonal identity, spatial organisation, and cellular state interact to shape tumour evolution during therapy.

8.4.1 Limitations

While clone-associated multi-modal phenotyping is presented in this chapter, antibody staining was performed on consecutive tissue sections rather than on the same section used for X-CODE barcode detection. This approach gave robust visualisation of X-CODE barcode units and protein markers in their respective conditions, enabling comparison across matched tumour regions. However, this experimental design precludes direct, cell-level co-measurement of clonal identity and protein expression. As demonstrated in Chapter 6, careful optimisation of each protein marker is likely required to ensure compatibility with RNA-based detection, and further investigation would be necessary to enable fully integrated, same-section multimodal measurements in this context. In addition, although quantitative analysis of antibody staining was performed, the results are not presented here, as they fall outside the scope of this chapter. Nonetheless, numerous well-established computational pipelines for protein-based analysis can be integrated into the X-CODE framework [106,233,234] .

Similarly, integration of MALDI-MSI data with X-CODE clonal identity was achieved using near-consecutive sections rather than the same tissue section. Chapter 7 established a framework for post-MSI X-CODE RNA detection on the same section. However, it also highlighted the need for further optimisation of multimodal generation from the same tissue section, which requires tailoring to individual samples and experimental contexts, consistent with established reports [152,206]. In this chapter, consecutive sections were used to assess detection and debarcoding of the full-scale 15K X-CODE library when applied to a proof-of-principle cancer evolution model. While clonal patterns were partially retained following MSI, loss of RNA barcode unit signal

reduced the number of computationally decodable barcodes, indicating that additional optimisation of MALDI–X-CODE workflows is required to achieve robust same-section debarcoding after MALDI-MSI.

A further, related limitation concerns the sparsity of MyC-CaP MALDI–MSI data, which necessitated pixel-level smoothing during analysis. In this experiment, 12 tumour sections of varying sizes across experimental time points were distributed across two slides and subjected to MALDI–MSI. This experimental setup resulted in MALDI acquisition times exceeding 24 hours. Prolonged acquisition may increase the evaporation of volatile metabolites and matrices and extend the duration for which RNA is exposed to suboptimal conditions, likely contributing to reduced post-MSI barcode recovery [206]. Future studies could mitigate these issues by analysing fewer tissue sections per MALDI run; however, this may increase per-sample reagent costs. Alternatively, the spatial resolution of the MALDI run could be reduced to shorten acquisition times, but this may introduce challenges for accurate co-registration with high-resolution fluorescence imaging data [200].

While robust multi-modal alignment and integration pipelines were demonstrated in this chapter, achieving same-section integration of MALDI, RNA, and antibody measurements will require additional optimisation. Experimental compromises, such as reduced sample throughput, would likely be necessary, as would extensive protein-target antibody optimisation. While costly and potentially labour-intensive, these considerations are not unique to the X-CODE framework and are a bottleneck in highly multiplexed protein staining and multimodal workflows [106,234]. Finally, although putative m/z identities were assigned using the LIPID MAPS database, orthogonal validation of compounds was not performed. The aim of this chapter, with respect to MSI, was to establish the feasibility of integrating high-complexity clonal barcoding with multimodal spatial phenotyping, such as MALDI, in a therapeutic tumour model. Future studies can build on this framework to optimise multi-omic workflows and to conduct targeted validation of compounds tailored to specific biological questions.

8.4.2 Conclusions

In conclusion, this chapter demonstrates the successful application of the X-CODE framework in an *in vivo* therapeutic model of prostate cancer evolution under androgen

deprivation. X-CODE enabled longitudinal assessment of clonal dynamics during treatment response and relapse, alongside spatial mapping of clonal organisation within intact tumour tissue. The development and application of a computational spatial debarcoding pipeline enabled robust *in situ* assignment of clonal identities, supporting integration with downstream spatial phenotyping analyses. By combining X-CODE with immunophenotyping and MALDI–MSI, this work further establishes the feasibility of multimodal spatial characterisation of clonally associated cellular features. Together, these findings position X-CODE as a flexible platform for resolving clonal heterogeneity and its phenotypic and metabolic context during tumour evolution *in vivo*.

Chapter 9 Thesis discussion, impact, and outlook

9.1 X-CODE as a scalable, probe-based cellular barcoding framework

Lineage tracing has become a central tool for studying development, tissue homeostasis, disease, and cancer evolution [48]. Identifying clonal identity alongside cellular state is increasingly recognised as crucial for understanding population dynamics under selection pressure [89]. However, existing barcoding strategies typically impose practical trade-offs between achievable clonal resolution, phenotypic depth, spatial information content, and experimental scalability. High-complexity barcode libraries are most commonly read out by sequencing [56,70,76], which limits experimental throughput and complicates spatial analysis, while probe-based and imaging-compatible barcodes are often restricted in library scale [72,92,111,160]. As a result, integrating clonal identity with extensive phenotypic and spatial information within a coherent experimental framework remains challenging. This design gap has constrained the study of clonal behaviour across analytical platforms and biological contexts, thereby motivating the development of barcoding strategies that are scalable, provide spatial resolution, and are compatible with diverse readout technologies.

X-CODE addresses this challenge by introducing a combinatorial RNA barcode architecture that is probe-detectable, enabling multi-platform readout whilst decoupling barcode complexity from detection burden. Clonal identity is encoded through the modular assembly of short barcode units (BUs), each of which can be independently targeted by probe-based detection. Barcode diversity is generated combinatorially, enabling libraries comprising 15,876 (15K) or 2,000,376 (2M) unique barcodes without requiring a proportional increase in the number of detection probes. In practice, detecting these libraries requires only 18 padlock and detection probes for the 15K library and 27 for the 2M library, along with a universal probe. Thus, the overall detection burden remains manageable, even with substantial increases in library complexity. This design contrasts with fully random barcode libraries, in which each barcode is a randomly generated, unique short sequence. Thus, the design is inherently incompatible with *in situ* hybridisation-based targeting using a fixed, reusable probe set.

A defining feature of the X-CODE framework is its dual-barcode architecture, in which each clonal identity is represented by both a long, probe-detectable RNA barcode and a matched short barcode optimised for sequencing. The long barcode is designed for padlock-probe-based transcript capture, followed by rolling-circle amplification of the target sequence, enabling sensitive and specific detection by cytometric and imaging platforms. The short barcode supports validation and investigation of DNA or single-cell RNA sequencing, as well as barcode-specific clonal retrieval for downstream functional analyses. Importantly, the correspondence between long and short barcodes is fixed during library construction, ensuring that the same clonal identity can be explicitly tracked across probe-based and sequencing-based workflows. This intrinsic pairing eliminates the need for post hoc barcode matching, which could introduce ambiguity and incomplete recovery in systems that rely on random barcode architectures.

To effectively utilise barcoding tools, accurate and consistent identification of clonal identities across experiments and platforms is essential. Employing a predefined barcode whitelist further strengthens the analytical reliability and recovery of X-CODE sequences. Because X-CODE is built from a combinatorial barcode design, the analysis is restricted to a known set of valid barcode combinations. This reduces misassignment, improves signal interpretation, and increases tolerance to variability in probe performance or detection efficiency. This is particularly important at high library complexity, where small classification errors could create substantial ambiguity in clonal identification. Furthermore, in the context of cross-platform barcode retention, the whitelist strategy ensures that the same clonal identities can be reliably recovered from cytometric, imaging, and sequencing data. Thereby enabling direct comparison and integration of results across analytical modalities.

X-CODE builds on established probe-based RNA detection concepts, including rolling-circle amplification strategies developed for endogenous transcript detection. In particular, the proximity ligation assay for RNA (PLAYR) [137,141] and the STARmap strategy [118]. However, the use of sequence elements as genomically encoded, modular barcode units represents a conceptual departure from prior applications. In existing systems, such sequences function as passive detection elements within fixed assays, whereas in X-CODE, they are repurposed as the variable components of a combinatorial clonal barcode. Compared with probe-based barcode implementations developed for perturbation screens, such as the BARC-FISH strategy [184], X-CODE

operates at substantially greater library scale and is explicitly designed for clonal tracing.

In addition, this RNA-based design also overcomes fundamental constraints associated with protein-based barcoding approaches, in which barcode diversity is limited by epitope availability and antibody compatibility. Thereby, the X-CODE framework enables tracing of greater clonal complexity while retaining compatibility with probe-based and spatial detection across multiple platforms. Nonetheless, this design shifts the primary constraint from barcode library scalability to detection capacity. Resolving combinatorial barcode identities requires sufficient measurement channels to detect multiple barcode units per cell, where higher-complexity libraries necessitate more parameters. Detection of the X-CODE 15K and 2M libraries requires simultaneous discrimination between 18- and 27-barcode units, respectively. Consequently, effective deployment of X-CODE requires high-parameter analytical platforms, such as mass cytometry and highly multiplexed imaging techniques.

9.2 X-CODE enables cross-platform barcode retention and analytical flexibility

A central strength of the X-CODE framework is that clonal identity is retained across fundamentally different analytical modalities. The same barcode can be recovered across assays governed by distinct physical and chemical principles, such as fluorescence imaging, mass cytometry, and sequencing-based readouts. Importantly, this platform-agnostic design provides substantial experimental flexibility. Provided that samples are appropriately preserved and stored, initial analysis using one modality can guide subsequent analytical choices, enabling targeted use of resource-intensive and costly techniques. In practice, this allows adaptation to experimental workflows based on preliminary results, rather than committing to a single readout strategy. For example, this project employed bulk sequencing of all MyC-CaP samples across experimental time points to select specific samples for detailed spatial analyses. Thereby demonstrating that X-CODE facilitates iterative, decision-driven experimental design rather than inflexible, single-method approaches.

Suspension mass cytometry provides a particularly powerful readout of X-CODE barcodes by enabling sequencing-free, high-throughput clonal analysis at single-cell

resolution. Mass cytometry by time-of-flight (CyTOF) enables the analysis of millions of cells in a single experiment, facilitating the detection of rare clonal populations that may be difficult to resolve using sequencing-based approaches alone. Building on well-established RNA detection protocols using CyTOF [137,141], robust workflows for X-CODE detection were optimised. Importantly, X-CODE barcode readout by CyTOF is compatible with simultaneous antibody-based immunophenotyping, enabling direct linkage of clonal identity to cellular phenotype in heterogeneous populations. This analytical context is especially relevant for liquid malignancies, where flow-based techniques have historically been central to disease characterisation. Moreover, as spectral flow cytometry platforms continue to expand in marker capacity, the principles underlying CyTOF-based X-CODE detection are likely to be extensible to these systems, further broadening the range of accessible platforms for high-throughput, single-cell analysis of clonal dynamics.

The X-CODE framework also supports spatial clonal readout across high-parameter imaging platforms, including multiplexed fluorescence imaging and imaging mass cytometry. Using shared reagents and closely related protocols, barcode units were robustly detected at single-cell resolution across both modalities, despite inherent differences in image characteristics between fluorescence and mass cytometry data. Furthermore, the development of a unified spatial decoding strategy, implemented within a QuPath-based analysis pipeline, enabled consistent assignment of X-CODE barcodes. While visual features such as signal granularity and segmentation quality varied across platforms, computational identification of barcode units remained robust. This unified approach lowers the technical barrier to spatial clonal analysis and enables direct analysis of results across imaging modalities without reliance on platform-specific analysis software. These shared sample-preparation workflows, detection protocols, and analysis pipelines highlight the X-CODE framework's versatile, readout platform-independent design.

Beyond the platforms directly evaluated in this thesis, the probe-based detection design of X-CODE positions it for integration with additional high-parameter spatial detection technologies. The mass spectrometry-based imaging approach termed multiplexed ion beam imaging (MIBI) was designed for targeted visualisation of isotope-labelled antibodies *in situ* [235]. While conceptually similar to imaging mass cytometry, the physical processes and demands of each technique differ. Nonetheless, MIBI's high multiplexing

capacity could represent a natural extension of the X-CODE framework to an orthogonal detection strategy. Furthermore, there are increasing efforts to adapt lineage-tracing systems for readout by spatial transcriptomic assays, enabling simultaneous cellular RNA profiling. Recent approaches have demonstrated compatibility between evolvable barcodes and spatial transcriptomic platforms [159,160]. Nonetheless, evolvable systems often require more optimisation for tracing deployment and impose greater analytical demands than static systems, owing to their complexity. Additionally, barcodes can be context-specific, thereby limiting their widespread application.

These developments highlight the need for a flexible, accessible static barcoding method compatible with modern spatial transcriptomic platforms. Within the context of the X-CODE framework, integration with probe-based spatial transcriptomics platforms is particularly significant. The 10x Genomics Xenium platform employs probe-based transcript targeting and rolling-circle amplification, which aligns with the chemistry underlying X-CODE barcode readout. Based on this shared methodology, X-CODE is anticipated to integrate with Xenium's spatial transcriptomic workflows. Optimisation of this integration is currently being pursued as an extension of this work, but lies beyond the scope of the results presented here. Nonetheless, this potential of readout adaptation highlights the X-CODE's forward-looking design and presents an opportunity to integrate an additional layer of *in situ* phenotyping. Moreover, by carefully redesigning probe architectures, X-CODE detection may be extendable to other probe-based spatial transcriptomic platforms, such as the NanoString CosMx, without requiring construct-level redesign of the X-CODE libraries.

Despite its broad flexibility, effective deployment of the full-scale X-CODE 15K and 2M libraries across technologies may require platform-specific optimisation and careful parameter allocation. Detection of high-complexity libraries requires dedicating a subset of available measurement channels to barcode units, and probe-level variability must be accounted for during analysis. These considerations reflect trade-offs inherent in probe-based detection of RNA sequences compared to sequencing-based readout. For instance, decoding cells from the 15K and 2M X-CODE libraries requires successful positive detection of only 8 or 12 BUs per cell, respectively. Resolving multiple barcode parameters per cell within densely packed tissue sections can therefore pose challenges for accurate *in situ* decoding. In addition, barcode units that exhibit consistently lower detection efficiency across platforms may disproportionately reduce

overall barcode assignment rates. However, these effects can be partially mitigated by computational decoding strategies, including optimised cellular segmentation, attentive BU signal thresholding, and spatially informed reassignment approaches that leverage confidently decoded neighbouring cells.

9.3 Multimodal integration: linking clonal identity and phenotype

It is increasingly recognised that resolving clonal behaviour requires integration of lineage information with phenotypic, molecular, and spatial contexts. Across developmental biology and cancer research, recent work has emphasised that lineage-only approaches lack the resolution to account for functional heterogeneity, whereas phenotype- or state-based measurements alone cannot distinguish inherited from adaptive behaviours [48]. This has driven a broader shift toward multimodal and multi-omic workflows, particularly in cancer, to better understand therapeutic response and resistance that arise from interactions between clonal identity, cellular state, and the microenvironment [32]. Within this landscape, lineage tracing strategies that can be directly integrated with protein, transcriptional, and metabolic measurements are especially valuable. X-CODE was explicitly designed to operate within this emerging multimodal paradigm, enabling mapping of clonal identity as a foundational layer that can be linked to multiple biological readouts.

Integration of RNA-based barcode detection with multiplexed protein phenotyping enables direct association of clonal identity with functional cellular states. This thesis demonstrated the spatial association between specific MyC-CaP clones and their immunophenotyped states under androgen deprivation by analysing consecutive tissue sections. This reflects the practical challenges of same-section RNA–protein co-detection, which are not unique to the X-CODE workflow. Multimodal RNA–protein assays generally require careful optimisation of fixation, permeabilisation, and staining conditions to balance RNA preservation with antigen accessibility, and established workflows demonstrate that such integration is feasible but non-trivial [110]. The results presented here show that spatial X-CODE barcode mapping is compatible with protein phenotyping, while highlighting that same-section integration of clonal identities with full antibody panels will require target-specific optimisation. These constraints represent tractable technical challenges rather than conceptual barriers, and established RNA–

protein co-detection methodologies provide a foundation for optimising X-CODE protocols toward fully integrated workflows in future studies [147,183].

Beyond integration with protein phenotyping, linking clonal identity to unbiased molecular features provides an additional layer of biological context. Combining X-CODE with mass spectrometry imaging (MSI) enables association of clonal identity with spatially resolved metabolic landscapes of tumour heterogeneity. However, MSI-based workflows impose distinct constraints on RNA integrity, reflecting an inherent trade-off between efficient molecular ionisation and nucleic acid preservation. This challenge is not unique to X-CODE and has been reported in prior MSI–RNA *in situ* hybridisation workflows that explored the integration of matrix-assisted laser desorption/ionisation (MALDI) [206]. Laser energy, matrix chemistry, and acquisition parameters directly influence transcript preservation and MSI data quality. Within this thesis, optimisation of MALDI acquisition parameters and matrix selection defined practical operating regimes that retained sufficient barcode unit detection while yielding informative MSI data.

Application of this optimised regime to samples from the *in vivo* barcoded prostate cancer evolution model resulted in reduced recovery of computationally decodable barcode identities after MSI acquisition, consistent with partial RNA loss under MSI conditions. Nonetheless, integration of clonal identity using consecutive tissue sections enabled robust association of dominant clones with spatially localised metabolic features. Overall, these findings suggest that with present workflows, X-CODE–MSI integration is most effective for specifically linking dominant clones to molecular features, rather than providing comprehensive clonal-molecular integration across all cells. Nonetheless, recent studies that integrate spatial transcriptomics after MSI acquisition demonstrate that, under carefully optimised conditions, RNA integrity can be preserved sufficiently to enable downstream transcriptomic profiling [154,155] .

These advances highlight experimental strategies that mitigate RNA degradation and suggest that further optimisation of RNA-compatible MSI workflows is feasible. In this context, alternative ionisation approaches, such as desorption electrospray ionisation (DESI), may advance integrated workflows compared with MALDI. DESI operates under milder acquisition conditions than MALDI because ionisation is achieved through solvent-mediated desorption at ambient pressure rather than laser-induced ablation [135], thereby reducing thermal and physical disruption of nucleic acids. Implementation of

DESI–X-CODE workflows may improve compatibility with post-MSI RNA detection and offer promising avenues for future optimisation of X-CODE–MSI integration. However, the choice of ionisation strategy must also account for the specific classes of metabolites and molecular species of interest, as DESI and MALDI differ in their ionisation efficiencies, spatial resolution, and molecular coverage.

Furthermore, spatial transcriptomics platforms present the opportunity for *in situ* detection of cellular RNA barcodes with simultaneous transcript-based profiling, as implied in Section 9.2. While recent lineage-tracing studies have demonstrated evolvable barcode readout within spatial transcriptomic assays, these approaches use dynamic barcodes that prioritise lineage ancestry over stable clonal identity. In contrast, X-CODE provides a static clonal identifier that can serve as a lineage anchor within spatial transcriptomic workflows, enabling interpretation of transcriptional states in the context of clonal identity. As discussed, compatibility with probe-based spatial transcriptomics platforms provides a practical means to integrate these data types. Therefore, demonstrating X-CODE detection on the 10x Xenium platform would show that clonal identity can be recovered alongside spatially resolved transcriptional information while preserving tissue architecture.

The challenges in same-section integration of RNA, protein, and molecular imaging modalities reflect broader bottlenecks in multimodal spatial biology [106]. Achieving fully integrated measurements often requires experimental compromises, including trade-offs between sample throughput, phenotyping depth, cost, and protocol complexity. Specific instances encountered in this work include limited spatial analyses of potentially biologically relevant samples and the inability to conduct extensive per-target antibody optimisation due to cost constraints. Also, compromising dataset quality when combining conflicting analytical techniques. Overcoming the challenges encountered in multiomic biology will be critical to maximising the performance of a multimodal X-CODE framework. This thesis prioritised optimising the X-CODE framework to enhance tool flexibility, library scalability, and cross-platform barcode readout compatibility. Therefore, multi-modal integration strategies can be tailored and optimised to specific biological questions in future studies. This is particularly relevant for *in vivo* applications, where spatial resolution of clonal dynamics and their intersection with phenotype and molecular state is essential for understanding biological behaviour.

9.4 X-CODE reveals spatial clonal dynamics during tumour evolution

To demonstrate the biological utility of the X-CODE framework, it was applied to an *in vivo* MyC-CaP model of prostate cancer progression under androgen deprivation. This system provides a well-defined selective pressure that drives tumour adaptation and resistance [225]. By combining high-complexity clonal barcoding with longitudinal sampling and spatially resolved analysis, this model enabled interrogation of how clonal dynamics unfold over time and space during therapy response. Importantly, the experimental design enabled multimodal assessment of clonal behaviour across bulk and single-cell readouts with spatial resolution within the same biological samples. This provided a setting to evaluate X-CODE's ability to investigate clonal identity during tumour evolution *in vivo*.

Longitudinal analysis revealed that castration resistance in the MyC-CaP model arose from delayed clonal remodelling rather than from rapid selection of pre-existing dominant clones. Early post-castration tumours retained high clonal diversity, indicating that the initial therapeutic response did not immediately constrain the clonal landscape. In contrast, resistant endpoint tumours were characterised by pronounced expansion of a limited subset of clones. Notably, these dominant clones were not strongly enriched at early time points, suggesting that resistance reflects differential adaptive potential that develops over time under sustained selective pressure [232]. This temporal pattern argues against simple models of sudden clonal sweeps and instead supports a dynamic process of clonal restructuring during therapy-driven tumour evolution.

Spatial mapping of clonal identities revealed that temporal clonal remodelling was accompanied by marked reorganisation of tumour architecture. Whereas untreated and early post-castration tumours exhibited highly intermingled clonal patterns, resistant tumours were dominated by large, spatially segregated clonal regions. This separation indicates that clonal expansion during resistance is spatially structured and likely influenced by local microenvironmental conditions. By enabling direct visualisation of clonal organisation *in situ*, X-CODE reveals that clonal dominance is not only a population-level phenomenon but is also tightly linked to spatial niche occupation, underscoring the importance of spatial context in shaping tumour evolution under therapy.

Integration of clonal identity with immunophenotypic profiling showed that dominant resistant clones did not converge on a single phenotypic state. Instead, distinct clones exhibited divergent expression patterns across epithelial markers, proliferation-associated proteins, and microenvironment-related features. These differences were spatially organised, with phenotypic variation aligning with clonal territories within the tumour. This heterogeneity challenges simplified models of resistance based on uniform lineage switching or phenotypic convergence and supports a view in which multiple resistant clones coexist, each adopting distinct cellular states. Such diversity suggests that resistance to androgen deprivation is a multi-lineage outcome shaped by both intrinsic clonal properties and local environmental context.

Further integration with mass spectrometry imaging (MSI) enabled the association of clonal identity with spatially resolved metabolic features. Despite reduced barcode recovery following MSI acquisition, dominant clones from a consecutive section could be linked to distinct, non-overlapping metabolic signatures. Subsequent putative structure assignment revealed that predominantly lipid-associated compounds were present. Importantly, distinct resistant clones exhibited different metabolic features, reinforcing the observation that resistance does not converge on a single molecular programme. These findings indicate that clonal dominance under androgen deprivation is compatible with multiple metabolic configurations, highlighting metabolic plasticity as an additional feature of clonal heterogeneity during tumour evolution.

Together, these results support a model in which prostate cancer evolution under androgen deprivation is driven by multi-lineage, spatially structured adaptation rather than deterministic progression toward a uniform resistant state, consistent with observations in patients [30]. Resistance emerges through delayed clonal remodelling, spatial segregation of dominant clones, and divergence across phenotypic and metabolic dimensions. By integrating clonal identity with spatial phenotype and molecular state *in situ*, X-CODE enables direct resolution of this complexity within intact tumours. This integrative view reveals how clonal dynamics, spatial organisation, and cellular state interact to shape evolutionary trajectories during therapy.

While this study demonstrates the biological power of spatial clonal tracing, several limitations should be acknowledged. Multimodal integration was primarily achieved by using consecutive tissue sections rather than by fully integrating measurements from

the same section, reflecting current technical constraints in multiomic spatial biology as discussed in Section 9.3. In addition, the time and resource demands of high-dimensional spatial and multimodal acquisitions required balancing the number of samples analysed against the depth of phenotypic and molecular information obtained from each sample. These limitations do not detract from the development of the novel barcoding technology, its detection, or the core biological conclusions. Instead, they highlight areas for future workflow refinement. Importantly, the framework established here provides a foundation for increasingly comprehensive evolutionary analyses as multimodal methodologies continue to advance.

Chapter 10 Thesis conclusions

This thesis aimed to create a flexible RNA barcoding system for scalable clonal tracing, ensuring barcode retention across multiple platforms to provide spatial resolution and integrated multimodal phenotyping. To achieve this, the X-CODE system was engineered as a combinatorial, probe-detectable RNA barcode platform compatible with sequencing, flow cytometry, and spatial imaging technologies. Through systematic technical validation and application in an *in vivo* cancer model, this work establishes X-CODE as a versatile framework for resolving clonal architecture and associated cellular states across diverse experimental contexts.

The principal conclusions of this thesis are as follows:

- **X-CODE establishes a scalable, combinatorial RNA barcoding strategy for clonal tracing.**

  By assembling barcodes from predefined barcode units, X-CODE decouples barcode diversity from detection burden, enabling libraries comprising tens of thousands to millions of unique clonal identities to be resolved using a fixed, reusable probe repertoire.

- **Cross-platform barcode retention enables flexible experimental design.**

  X-CODE barcodes were reliably detected across sequencing, suspension mass cytometry, and high-parameter spatial imaging platforms. This platform-agnostic design allows adaptation of analytical strategies, using preliminary data to guide the selection of downstream readouts without compromising clonal identity.

- **Sequencing-free, high-throughput clonal analysis is enabled by mass cytometry.**

  Integration of X-CODE with CyTOF supports large-scale, single-cell clonal analysis alongside multiplexed protein phenotyping, providing sensitivity to rare clones and extending lineage-aware analysis to contexts traditionally dominated by flow-based approaches.

- **X-CODE supports spatial clonal mapping and multimodal integration.**

  Probe-based detection enabled spatially resolved barcode readout using multiplexed fluorescence imaging and imaging mass cytometry, supported by a unified spatial decoding pipeline. Integration of protein phenotyping and mass spectrometry imaging further demonstrated that clonal identity can be linked to spatial phenotypic and molecular features, with clearly defined trade-offs that are typical of multiomic workflows.

- **Application to an *in vivo* prostate cancer model reveals non-convergent clonal evolution under therapy.**

  In a MyC-CaP *in vivo* model of androgen deprivation, X-CODE revealed delayed clonal remodelling, spatial segregation of dominant clones, and divergence across phenotypic and metabolic dimensions. These findings support a model of therapy resistance driven by multi-lineage, spatially structured adaptation rather than deterministic progression toward a single resistant state.

- **The X-CODE framework is positioned for integration with emerging spatial and multimodal technologies.**

  Compatibility with probe-based spatial transcriptomics platforms reflects X-CODE's forward-looking design and supports its potential integration into emerging spatial biology workflows as a stable clonal identifier.

Collectively, these findings demonstrate that X-CODE provides a robust and extensible framework for clonal tracing that bridges sequencing, cytometric, and spatial imaging modalities. By enabling direct integration of lineage identity with phenotypic and molecular information, this work lays the foundation for future studies to dissect how clonal architecture and cellular state interact to shape biological systems, particularly in the context of tumour evolution and therapeutic response.

Chapter 11 Supplementary information

11.1 Simple barcode cloning and detection probes

Supplementary Table 1. Primer sequences for Gibson cloning of barcodes A, B, and C.

Primers	**Sequence**
A 1+2 forward	CCGGATCGTGACCAAACTCAGTCGTGGACACTCTGGGACGCTACTACACACACCTTCGAGACAAAT ATCGTCCGGGACACTCTGGGACGCTACTTTTTTG
A 1+2 reverse	AATTCAAAAAAGTAGCGTCCCAGAGTGTCCCGGACGATATTTGTCTCGAAGGTGTGTGTAGTAGCG TCCCAGAGTGTCCACGACTGAGTTTGGTCACGAT
B 2+3 forward	CCGGCTTCGAGACAAATATCGTCCGGGACACTCTGGGACGCTACTACACACACCTGCCAATGAAAG ATTCCTCGGGACACTCTGGGACGCTACTTTTTTG
B 2+3 reverse	AATTCAAAAAAGTAGCGTCCCAGAGTGTCCCGAGGAATCTTTCATTGGCAGGTGTGTGTAGTAGCG TCCCAGAGTGTCCCGGACGATATTTGTCTCGAAG
C 3+4 forward	CCGGCTGCCAATGAAAGATTCCTCGGGACACTCTGGGACGCTACTACACACACCAGGCTACTAAAC TACCTTGGGGACACTCTGGGACGCTACTTTTTTG
C 3+4 reverse	AATTCAAAAAAGTAGCGTCCCAGAGTGTCCCCAAGGTAGTTTAGTAGCCTGGTGTGTGTAGTAGCG TCCCAGAGTGTCCCGAGGAATCTTTCATTGGCAG

Supplementary Table 2. Barcode A, B, and C padlock and detection probes.

X-CODE	**Size (bp)**	**Barcode Unit Sequence**
BU1	56	/5Phos/ACATTAACGACTGAGTTTGGTCACGATAAATTATTACCGGTGCATACACTAAAGATA
BU2	56	/5Phos/ACATTACGGACGATATTTGTCTCGAAGAAATTATTACCGGTGCATACACTAAAGATA
BU3	56	/5Phos/ACATTACGAGGAATCTTTCATTGGCAGAAATTATTACCGGTGCATACACTAAAGATA
BU4	56	/5Phos/ACATTACCAAGGTAGTTTAGTAGCCTGAAATTATTACCGGTGCATACACTAAAGATA
Universal	31	AGTAGCGTCCCAGAGTGTCTAATGTTATC*T*T
Fluro1	21	/5Alex488N/ACGACTGAGTTTGGTCACGAT
Fluro2	21	/5ATTO550N/CGGACGATATTTGTCTCGAAG
Fluro3	21	/5CY5/CGAGGAATCTTTCATTGGCAG
Fluro4	21	/5Alex750N/CCAAGGTAGTTTAGTAGCCTG
Iso1	21	(163Dy) /5ThioMC6-D/ACGACTGAGTTTGGTCACGAT
Iso2	21	(Tb159) /5ThioMC6-D/CGGACGATATTTGTCTCGAAG
Iso3	21	(165Ho) /5ThioMC6-D/CGAGGAATCTTTCATTGGCAG
Iso4	21	(173Yb) /5ThioMC6-D/CCAAGGTAGTTTAGTAGCCTG

11.2 Probes for β-actin detection

Supplementary Table 3. Padlock and detection probes for β-actin.

β-actin detection sequences adapted from Wang et al. [118]

Probe	**Sequence**
Shell 1	/5Phos/ACATTAGCAGCGATATCGTCATCCATAATTATTACCGGTGCATACACTAAAGATA
Tail 1	AGCCGTTGTCGACGACCAGTAATGTTATCTT*T*T
Shell 2	/5Phos/ACATTAGGGCCGGACTCATCGTACTCAATTATTACCGGTGCATACACTAAAGATA
Tail 2	ACTTGCGGTGCACGATGGATAATGTTATCTT*T*T
Fluro1	/5CY5/GCAGCGATATCGTCATCCAT
Fluro2	/5CY5/GGGCCGGACTCATCGTACTC
Iso1	(Gd155) /5ThioMC6-D/GCAGCGATATCGTCATCCAT
Iso2	(Gd155) /5ThioMC6-D/GGGCCGGACTCATCGTACTC

11.3 X-CODE library cloning

Supplementary Table 4. Primer sequences for Gibson cloning of X-CODE library backbone.

Primers	**Sequence**
Filler_Fwd	CGGTGCCTGAACGCGTTAAGGAAACACCGGAGACGGTTG
Filler_Rev	TCTAGAGTGTGTGTGTCGACAGAGACGTACAAAAAAGAGCAAG
LTR_Fwd	GTCGACACACACACTCTAGAGGGTCTCTCTGGTTAGAC
LTR_Rev	GATGACATGAACTACTATACTGCTAGAGATTTTCCACAC

Supplementary Table 5. Primer sequences for add-ons of the X-CODE library construct.

Primers	**Sequence**
scaffold-gRNA_Fwd	ACACGAAGACGGGTTTTAGAGCTAGAAATAGCAAGT
scaffold-gRNA_Rev	ACACTCTAGAAGTACAAGCAAAAAGCAGTGTCTCAAG
WPRE-SplitLTR-TSO-U6 _Fwd	ACACGGATCCAATCAACCTCTGGATTACAA
WPRE-SplitLTR-TSO-U6 _Rev	ACACGAATTCTTTCGTCCTTTCCACAAGAT

11.4 Barcode unit sequences 1–27

Supplementary Table 6. Barcode sequences 1–27.

23 sequences adapted from PLAYR insert sequences. (*) indicates 4 additionally designed sequences, according to PLAYR design strategy.

X-CODE	Size (bp)	Barcode Unit Sequence
BU1	41	ATCGTGACCAAACTCAGTCGTGGACACTCTGGGACGCTACT
BU2	41	CGCAAGTCTAAACTGCTCATGGGACACTCTGGGACGCTACT
BU3	41	CTGCCAATGAAAGATTCCTCGGGACACTCTGGGACGCTACT
BU4	41	CAGGCTACTAAACTACCTTGGGGACACTCTGGGACGCTACT
BU5	41	CACCAGTTGAAACTCTTCGAGGGACACTCTGGGACGCTACT
*BU6**	41	CTCAATGCTCAATGCTCCAGGGGACACTCTGGGACGCTACT
BU7	41	CTACATGGCAAACCGCTTATGGGACACTCTGGGACGCTACT
BU8	41	CAACCTGGTAAACTCGATCTGGGACACTCTGGGACGCTACT
BU9	41	CATCCTGAGAAATACGTTCCGGGACACTCTGGGACGCTACT
BU10	41	CGCCATGATAAATCGCTACTGGGACACTCTGGGACGCTACT
BU11	41	CACACTTGGAAATACGCTCTGGGACACTCTGGGACGCTACT
BU12	41	CATGCGACTAAACGTCTTACGGGACACTCTGGGACGCTACT
*BU13**	41	CCGTAACGTAAATCTCTCAGGGGACACTCTGGGACGCTACT
*BU14**	41	CATCGACTGAAACCACTTGTGGGACACTCTGGGACGCTACT
BU15	41	CAACGCGTTAAACTCCGATTGGGACACTCTGGGACGCTACT
BU16	41	CAATTCCGGAAATTCGCACTGGGACACTCTGGGACGCTACT
BU17	41	CCGCTAAGTAAATCCTTCAGGGGACACTCTGGGACGCTACT
BU18	41	CTTAAGCGCAAACGTTACTCGGGACACTCTGGGACGCTACT
BU19	41	CTTCGAGACAAATATCGTCCGGGACACTCTGGGACGCTACT
BU20	41	CTCGGAATCAAATGACTCTCGGGACACTCTGGGACGCTACT
BU21	41	CACAGTGTCAAACTGTCTACGGGACACTCTGGGACGCTACT
*BU22**	41	CCGGCCTCTAAACTTATGACGGGACACTCTGGGACGCTACT
BU23	41	CAGATGCCTAAACACTTGTCGGGACACTCTGGGACGCTACT
*BU24**	41	CCATGCAGTAAACTAGTCTCGGGACACTCTGGGACGCTACT
BU25	41	CTGTAGACCAAATCTATCCGGGGACACTCTGGGACGCTACT
BU26	41	CTGGCACATAAACGCATCTTGGGACACTCTGGGACGCTACT
BU27	41	CCTCGAATGAAATCTCACGTGGGACACTCTGGGACGCTACT

11.5 Padlock probes 1–27 and universal probe

Supplementary Table 7. X-CODE padlock probes 1–27 and the universal probe.

X-CODE	Size (bp)	Probes
BU1	57	/5Phos/ACATTAACGACTGAGTTTGGTCACGATAAATTATTACCGGTGCATACACTAAAGATA
BU2	57	/5Phos/ACATTACATGAGCAGTTTAGACTTGCGAAATTATTACCGGTGCATACACTAAAGATA
BU3	57	/5Phos/ACATTACGAGGAATCTTTCATTGGCAGAAATTATTACCGGTGCATACACTAAAGATA
BU4	57	/5Phos/ACATTACCAAGGTAGTTTAGTAGCCTGAAATTATTACCGGTGCATACACTAAAGATA
BU5	57	/5Phos/ACATTACTCGAAGAGTTTCAACTGGTGAAATTATTACCGGTGCATACACTAAAGATA
BU6	57	/5Phos/ACATTACCTGGAGAATTTCGCATTGAGAAATTATTACCGGTGCATACACTAAAGATA
BU7	57	/5Phos/ACATTACATAAGCGGTTTGCCATGTAGAAATTATTACCGGTGCATACACTAAAGATA
BU8	57	/5Phos/ACATTACAGATCGAGTTTACCAGGTTGAAATTATTACCGGTGCATACACTAAAGATA
BU9	57	/5Phos/ACATTACGGAACGTATTTCTCAGGATGAAATTATTACCGGTGCATACACTAAAGATA
BU10	57	/5Phos/ACATTACAGTAGCGATTTATCATGGCGAAATTATTACCGGTGCATACACTAAAGATA
BU11	57	/5Phos/ACATTACAGAGCGTATTTCCAAGTGTGAAATTATTACCGGTGCATACACTAAAGATA
BU12	57	/5Phos/ACATTACGTAAGACGTTTAGTCGCATGAAATTATTACCGGTGCATACACTAAAGATA
BU13	57	/5Phos/ACATTACCTGAGAGATTTACGTTACGGAAATTATTACCGGTGCATACACTAAAGATA
BU14	57	/5Phos/ACATTACACAAGTGGTTTCAGTCGATGAAATTATTACCGGTGCATACACTAAAGATA
BU15	57	/5Phos/ACATTACAATCGGAGTTTAACGCGTTGAAATTATTACCGGTGCATACACTAAAGATA
BU16	57	/5Phos/ACATTACAGTGCGAATTTCCGGAATTGAAATTATTACCGGTGCATACACTAAAGATA
BU17	57	/5Phos/ACATTACCTGAAGGATTTACTTAGCGGAAATTATTACCGGTGCATACACTAAAGATA
BU18	57	/5Phos/ACATTACGAGTAACGTTTGCGCTTAAGAAATTATTACCGGTGCATACACTAAAGATA
BU19	57	/5Phos/ACATTACGGACGATATTTGTCTCGAAGAAATTATTACCGGTGCATACACTAAAGATA
BU20	57	/5Phos/ACATTACGAGAGTCATTTGATTCCGAGAAATTATTACCGGTGCATACACTAAAGATA
BU21	57	/5Phos/ACATTACGTAGACAGTTTGACACTGTGAAATTATTACCGGTGCATACACTAAAGATA
BU22	57	/5Phos/ACATTACGTCATAAGTTTAGAGGCCGGAAATTATTACCGGTGCATACACTAAAGATA
BU23	57	/5Phos/ACATTACGACAAGTGTTTAGGCATCTGAAATTATTACCGGTGCATACACTAAAGATA
BU24	57	/5Phos/ACATTACGAGACTAGTTTACTGCATGGAAATTATTACCGGTGCATACACTAAAGATA
BU25	57	/5Phos/ACATTACCGGATAGATTTGGTCTACAGAAATTATTACCGGTGCATACACTAAAGATA
BU26	57	/5Phos/ACATTACAAGATGCGTTTATGTGCCAGAAATTATTACCGGTGCATACACTAAAGATA
BU27	57	/5Phos/ACATTACACGTGAGATTTCATTCGAGGAAATTATTACCGGTGCATACACTAAAGATA
Universal	31	AGTAGCGTCCCAGAGTGTCTAATGTTATC*T*T

11.6 X-CODE isotope detection probes

Supplementary Table 8. Probes 1–18 for isotope detection of the X-CODE 15K library.

X-CODE	Size (bp)	Isotope conjugation	Probe sequence prior to conjugation
BU1	21	171Yb	/5ThioMC6-D/ACGACTGAGTTTGGTCACGAT
BU2	21	158Gd	/5ThioMC6-D/CATGAGCAGTTTAGACTTGCG
BU3	21	165Ho	/5ThioMC6-D/CGAGGAATCTTTCATTGGCAG
BU4	21	152Sm	/5ThioMC6-D/CCAAGGTAGTTTAGTAGCCTG
BU5	21	143Nd	/5ThioMC6-D/CTCGAAGAGTTTCAACTGGTG
BU6	21	147Sm	/5ThioMC6-D/CCTGGAGAATTTCGCATTGAG
BU7	21	145Nd	/5ThioMC6-D/CATAAGCGGTTTGCCATGTAG
BU8	21	144Nd	/5ThioMC6-D/CAGATCGAGTTTACCAGGTTG
BU9	21	146Nd	/5ThioMC6-D/CGGAACGTATTTCTCAGGATG
BU10	21	153Eu	/5ThioMC6-D/CAGTAGCGATTTATCATGGCG
BU11	21	168Er	/5ThioMC6-D/CAGAGCGTATTTCCAAGTGTG
BU12	21	154Sm	/5ThioMC6-D/CGTAAGACGTTTAGTCGCATG
BU13	21	167Er	/5ThioMC6-D/CCTGAGAGATTTACGTTACGG
BU14	21	161Dy	/5ThioMC6-D/CACAAGTGGTTTCAGTCGATG
BU15	21	172Yb	/5ThioMC6-D/CAATCGGAGTTTAACGCGTTG
BU16	21	166Er	/5ThioMC6-D/CAGTGCGAATTTCCGGAATTG
BU17	21	170Er	/5ThioMC6-D/CCTGAAGGATTTACTTAGCGG
BU18	21	174Yb	/5ThioMC6-D/CGAGTAACGTTTGCGCTTAAG

Supplementary Table 9. Probes 1–27 for isotope detection of the X-CODE 2M library.

X-CODE	Size (bp)	Isotope conjugation	Probe sequence prior to conjugation
BU1	21	163Dy	/5ThioMC6-D/ACGACTGAGTTTGGTCACGAT
BU2	21	158Gd	/5ThioMC6-D/CATGAGCAGTTTAGACTTGCG
BU3	21	165Ho	/5ThioMC6-D/CGAGGAATCTTTCATTGGCAG
BU4	21	173Yb	/5ThioMC6-D/CCAAGGTAGTTTAGTAGCCTG
BU5	21	143Nd	/5ThioMC6-D/CTCGAAGAGTTTCAACTGGTG
BU6	21	164Dy	/5ThioMC6-D/CCTGGAGAATTTCGCATTGAG
BU7	21	142Nd	/5ThioMC6-D/CATAAGCGGTTTGCCATGTAG
BU8	21	144Nd	/5ThioMC6-D/CAGATCGAGTTTACCAGGTTG
BU9	21	146Nd	/5ThioMC6-D/CGGAACGTATTTCTCAGGATG
BU10	21	162Dy	/5ThioMC6-D/CAGTAGCGATTTATCATGGCG
BU11	21	168Er	/5ThioMC6-D/CAGAGCGTATTTCCAAGTGTG
BU12	21	169Tm	/5ThioMC6-D/CGTAAGACGTTTAGTCGCATG
BU13	21	167Er	/5ThioMC6-D/CCTGAGAGATTTACGTTACGG
BU14	21	176Yb	/5ThioMC6-D/CACAAGTGGTTTCAGTCGATG
BU15	21	172Yb	/5ThioMC6-D/CAATCGGAGTTTAACGCGTTG
BU16	21	166Er	/5ThioMC6-D/CAGTGCGAATTTCCGGAATTG
BU17	21	170Er	/5ThioMC6-D/CCTGAAGGATTTACTTAGCGG
BU18	21	174Yb	/5ThioMC6-D/CGAGTAACGTTTGCGCTTAAG
BU19	21	159Tb	/5ThioMC6-D/CGGACGATATTTGTCTCGAAG
BU20	21	151Eu	/5ThioMC6-D/CGAGAGTCATTTGATTCCGAG
BU21	21	154Sm	/5ThioMC6-D/CGTAGACAGTTTGACACTGTG
BU22	21	150Nd	/5ThioMC6-D/CGTCATAAGTTTAGAGGCCGG
BU23	21	152Sm	/5ThioMC6-D/CGACAAGTGTTTAGGCATCTG
BU24	21	141Pr	/5ThioMC6-D/CGAGACTAGTTTACTGCATGG
BU25	21	149Sm	/5ThioMC6-D/CCGGATAGATTTGGTCTACAG
BU26	21	153Eu	/5ThioMC6-D/CAAGATGCGTTTATGTGCCAG
BU27	21	161Dy	/5ThioMC6-D/CACGTGAGATTTCATTCGAGG

11.7 X-CODE fluorescent detection probes

Supplementary Table 10. Probes 1–27 for fluorescent detection of the X-CODE libraries.

X-CODE	Length (bp)	Probes
BU1	21	/5Alex488N/ACGACTGAGTTTGGTCACGAT
BU2	21	/5Alex488N/CATGAGCAGTTTAGACTTGCG
BU3	21	/5CY5/CGAGGAATCTTTCATTGGCAG
BU4	21	/5Alex488N/CCAAGGTAGTTTAGTAGCCTG
BU5	21	/5ATTO550N/CTCGAAGAGTTTCAACTGGTG
BU6	21	/5ATTO550N/CCTGGAGAATTTCGCATTGAG
BU7	21	/5Alex488N/CATAAGCGGTTTGCCATGTAG
BU8	21	/5ATTO550N/CAGATCGAGTTTACCAGGTTG
BU9	21	/5CY5/CGGAACGTATTTCTCAGGATG
BU10	21	/5Alex488N/CAGTAGCGATTTATCATGGCG
BU11	21	/5ATTO550N/CAGAGCGTATTTCCAAGTGTG
BU12	21	/5CY5/CGTAAGACGTTTAGTCGCATG
BU13	21	/5CY5/CCTGAGAGATTTACGTTACGG
BU14	21	/5ATTO550N/CACAAGTGGTTTCAGTCGATG
BU15	21	/5CY5/CAATCGGAGTTTAACGCGTTG
BU16	21	/5Alex488N/CAGTGCGAATTTCCGGAATTG
BU17	21	/5ATTO550N/CCTGAAGGATTTACTTAGCGG
BU18	21	/5CY5/CGAGTAACGTTTGCGCTTAAG
BU19	21	/5ATTO550N/CGGACGATATTTGTCTCGAAG
BU20	21	/5ATTO550N/CGAGAGTCATTTGATTCCGAG
BU21	21	/5CY5/CGTAGACAGTTTGACACTGTG
BU22	21	/5Alex488N/CGTCATAAGTTTAGAGGCCGG
BU23	21	/5ATTO550N/CGACAAGTGTTTAGGCATCTG
BU24	21	/5Alex488N/CGAGACTAGTTTACTGCATGG
BU25	21	/5Alex488N/CCGGATAGATTTGGTCTACAG
BU26	21	/5ATTO550N/CAAGATGCGTTTATGTGCCAG
BU27	21	/5CY5/CACGTGAGATTTCATTCGAGG

11.8 Akoya PhenoCycler-Fusion cycle information

Supplementary Table 11. Akoya PhenoCycler-Fusion cycles for X-CODE 15K library readout.

Cycle	**Alexa Fluor 488**	**Cy5 / Alexa Fluor 647**	**ATTO 550**
1	BU1	BU3	BU5
2	BU2	BU9	BU8
3	BU4	BU12	BU11
4	BU7	BU15	BU17
5	BU10	BU18	BU6
6	BU16	BU13	BU14
7		β-actin	

Supplementary Table 12. Akoya PhenoCycler-Fusion cycles for X-CODE 2M library readout.

Cycle	**Alexa Fluor 488**	**Cy5 / Alexa Fluor 647**	**ATTO 550**
1	BU1	BU3	BU5
2	BU2	BU9	BU8
3	BU4	BU12	BU11
4	BU7	BU15	BU17
5	BU10	BU18	BU6
6	BU16	BU13	BU14
7	BU25	BU21	BU19
8	BU22	BU27	BU20
9	BU24	β-actin	BU23
10			BU26

Rights and Permissions

Figure 1-2

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212750430530
License date	Feb 19, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Cell Biology
Licensed Content Title	Untangling the web of intratumour heterogeneity
Licensed Content Author	Zheqi Li et al
Licensed Content Date	Aug 8, 2022
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Fig. 1: Cell-autonomous and non-cell-autonomous sources of ITH.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, Not Applicable SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 1-3

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212750846229
License date	Feb 19, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	Spatial biology of cancer evolution
Licensed Content Author	Zaira Seferbekova et al
Licensed Content Date	Dec 9, 2022
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Fig. 3: Mapping the cancer ecosystem.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 1-4

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212751086158
License date	Feb 19, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	Building a lineage from single cells: genetic techniques for cell lineage tracking
Licensed Content Author	Mollie B. Woodworth et al
Licensed Content Date	Jan 23, 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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Figure 1: Prospective and retrospective lineage tracing.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 1-6

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212751243012
License date	Feb 19, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	Mastering the use of cellular barcoding to explore cancer heterogeneity
Licensed Content Author	Antonin Serrano et al
Licensed Content Date	Aug 18, 2022
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026

Portions	Fig. 1: Workflow of the multistep process involved in genetic and optical barcoding.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 1-7

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

Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212751419481
License date	Feb 19, 2026
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Molecular recorders to track cellular events
Licensed Content Author	Nanami Masuyama, Naoki Konno, Nozomu Yachie
Licensed Content Date	Jul 28, 2022
Licensed Content Volume	377
Licensed Content Issue	6605
Volume number	377
Issue number	6605
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	The DNA event recording concept
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road

	London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 1-9

ELSEVIER LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

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

License Number	6212760695185
License date	Feb 19, 2026
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Cancer
Licensed Content Title	Charting the tumor microenvironment with spatial profiling technologies
Licensed Content Author	Mattia Toninelli,Grazisa Rossetti,Massimiliano Pagani
Licensed Content Date	Dec 1, 2023
Licensed Content Volume	9
Licensed Content Issue	12
Licensed Content Pages	12
Start Page	1085
End Page	1096
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Figure 1. Detecting the spatial location of proteins.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

Figure 1-10

ELSEVIER LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

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

License Number	6212761024693
License date	Feb 19, 2026
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Pharmacological Sciences
Licensed Content Title	Spatial pharmacology using mass spectrometry imaging
Licensed Content Author	Presha Rajbhandari,Taruna V. Neelakantan,Noreen Hosny,Brent R. Stockwell
Licensed Content Date	Jan 1, 2024
Licensed Content Volume	45
Licensed Content Issue	1
Licensed Content Pages	14
Start Page	67
End Page	80
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Figure 1. DESI, MALDI, and SIMS mass spectrometry imaging (MSI) across spatial scales enable spatial pharmacology
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

Figure 1-12

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212761304305
License date	Feb 19, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	Spatial biology of cancer evolution
Licensed Content Author	Zaira Seferbekova et al
Licensed Content Date	Dec 9, 2022
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	no

your order?	
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Fig. 5: Translational opportunities.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 4-1

ELSEVIER LICENSE
TERMS AND CONDITIONS
Feb 19, 2026

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

License Number	6212731269568
License date	Feb 19, 2026
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cancer Cell
Licensed Content Title	Spatial profiling technologies illuminate the tumor microenvironment
Licensed Content Author	Ofer Elhanani,Raz Ben-Uri,Leeat Keren
Licensed Content Date	Mar 13, 2023
Licensed Content Volume	41
Licensed Content Issue	3
Licensed Content Pages	17
Start Page	404
End Page	420
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Figure 1. Technologies for spatial profiling of mRNA
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner

	Elizabeth Flittner 15 Cotswold Road
Requestor Location	
	London, Not Applicable SM2 5NG United Kingdom
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

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

Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212740935638
License date	Feb 19, 2026
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Three-dimensional intact-tissue sequencing of single-cell transcriptional states
Licensed Content Author	Xiao Wang, William E. Allen, Matthew A. Wright, Emily L. Sylwestrak, et al.
Licensed Content Date	Jul 27, 2018
Licensed Content Volume	361
Licensed Content Issue	6400
Volume number	361
Issue number	6400
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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Fig. S1. SNAIL probes for high-quality RNA imaging: mouse brain instantiation.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, Not Applicable SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

Figure 5-1

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Feb 19, 2026

This Agreement between Elizabeth Flittner ("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	6212741272657
License date	Feb 19, 2026
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Protocols
Licensed Content Title	Multiplexed profiling of RNA and protein expression signatures in individual cells using flow or mass cytometry
Licensed Content Author	Andrew D. Duckworth et al
Licensed Content Date	Feb 6, 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	Combinatorial RNA barcodes for spatial clonal tracing with multi-platform phenotyping to study cancer evolution
Institution name	Imperial College London
Expected presentation date	Apr 2026
Portions	Fig. 1: Schematic representation of the PLAYR assay.
The Requesting Person / Organization to Appear on the License	Elizabeth Flittner
Requestor Location	Ms. Elizabeth Flittner 15 Cotswold Road London, Not Applicable SM2 5NG United Kingdom
Payment Type	Invoice
Email Address	elizabeth.flittner@icr.ac.uk
Billing Address	Ms. Elizabeth Flittner 15 Cotswold Road London, United Kingdom SM2 5NG
Total	0.00 GBP

References

1 Zinner M, Lukonin I, Liberali P. Design principles of tissue organisation: How single cells coordinate across scales. *Curr Opin Cell Biol* 2020; **67**: 37–45.

2 Tabassum DP, Polyak K. Tumorigenesis: it takes a village. *Nat Rev Cancer* 2015; **15**: 473–483.

3 Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 2006; **6**: 924–935.

4 Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012; **481**: 306–313.

5 Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. *Embo J* 2021; **40**: e108389.

6 Greaves M. Evolutionary Determinants of Cancer. *Cancer Discov* 2015; **5**: 806–820.

7 Nowell PC. The Clonal Evolution of Tumor Cell Populations. *Science* 1976; **194**: 23–28.

8 Davis A, Gao R, Navin N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim Biophys Acta (BBA) - Rev Cancer* 2017; **1867**: 151–161.

9 Laplane L, Maley CC. The evolutionary theory of cancer: challenges and potential solutions. *Nat Rev Cancer* 2024; **24**: 718–733.

10 Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New Engl J Medicine* 2012; **366**: 883–892.

11 Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P *et al.* Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci* 2015; **112**: E6496–E6505.

12 Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* 2011; **144**: 27–40.

13 Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y *et al.* Punctuated Evolution of Prostate Cancer Genomes. *Cell* 2013; **153**: 666–677.

14 Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* 2016; **48**: 1119–1130.

15 Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J *et al.* A Big Bang model of human colorectal tumor growth. *Nat Genet* 2015; **47**: 209–216.

16 Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. *Nat Genet* 2016; **48**: 238–244.

17 Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 2014; **512**: 155–160.

18 McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* 2017; **168**: 613–628.

19 Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 2013; **501**: 338–345.

20 Network TCGAR, Abeshouse A, Ahn J, Akbani R, Ally A, Amin S *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 2015; **163**: 1011–1025.

21 Attard G, Parker C, Eeles RA, Schröder F, Tomlins SA, Tannock I *et al.* Prostate cancer. *Lancet* 2016; **387**: 70–82.

22 Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature* 2013; **501**: 355–364.

23 Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* 2020; **37**: 471–484.

24 Goyette M-A, Lipsyc-Sharf M, Polyak K. Clinical and translational relevance of intratumor heterogeneity. *Trends Cancer* 2023. doi:10.1016/j.trecan.2023.05.001.

25 Li Z, Seehawer M, Polyak K. Untangling the web of intratumour heterogeneity. *Nat Cell Biol* 2022; **24**: 1192–1201.

26 McGranahan N, Swanton C. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* 2015; **27**: 15–26.

27 Black JRM, McGranahan N. Genetic and non-genetic clonal diversity in cancer evolution. *Nat Rev Cancer* 2021; **21**: 379–392.

28 Torborg SR, Li Z, Chan JE, Tammela T. Cellular and molecular mechanisms of plasticity in cancer. *Trends Cancer* 2022. doi:10.1016/j.trecan.2022.04.007.

29 Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science* 2017; **357**. doi:10.1126/science.aal2380.

30 Beltran H, Hruszkewycz A, Scher HI, Hildesheim J, Isaacs J, Yu EY *et al.* The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. *Clin Cancer Res* 2019; **25**: 6916–6924.

31 Yamada Y, Beltran H. Clinical and Biological Features of Neuroendocrine Prostate Cancer. *Curr Oncol Rep* 2021; **23**: 15.

32 Pei G, Liu Y, Wang L. Spatially resolving cancer: from cell states to therapy. *Trends Cancer* 2025. doi:10.1016/j.trecan.2025.09.002.

33 Nishimura T, Kakiuchi N, Yoshida K, Sakurai T, Kataoka TR, Kondoh E *et al.* Evolutionary histories of breast cancer and related clones. *Nature* 2023; **620**: 607–614.

34 Savy T, Flanders L, Karpanasamy T, Sun M, Gerlinger M. Cancer evolution: from Darwin to the Extended Evolutionary Synthesis. *Trends Cancer* 2025; **11**: 204–215.

35 Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 2013; **19**: 1423–1437.

36 Visser KE de, Joyce JA. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell* 2023; **41**: 374–403.

37 Seferbekova Z, Lomakin A, Yates LR, Gerstung M. Spatial biology of cancer evolution. *Nat Rev Genet* 2022; : 1–19.

38 Dey P, Kimmelman AC, DePinho RA. Metabolic Codependencies in the Tumor MicroenvironmentMetabolic Codependencies in Cancer. *Cancer Discov* 2021; **11**: 1067–1081.

39 Faubert B, Li KY, Cai L, Hensley CT, Kim J, Zacharias LG *et al.* Lactate Metabolism in Human Lung Tumors. *Cell* 2017; **171**: 358-371.e9.

40 Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* 2014; **514**: 54–58.

41 Janiszewska M, Tabassum DP, Castaño Z, Cristea S, Yamamoto KN, Kingston NL *et al.* Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. *Nat Cell Biol* 2019; **21**: 879–888.

42 Sharma P, Goswami S, Raychaudhuri D, Siddiqui BA, Singh P, Nagarajan A *et al.* Immune checkpoint therapy—current perspectives and future directions. *Cell* 2023; **186**: 1652–1669.

43 Schürch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* 2020; **182**: 1341-1359.e19.

44 Swanton C. Embracing cancer complexity: Hallmarks of systemic disease. *Cell* 2024; **187**: 1589–1616.

45 Kretzschmar K, Watt FM. Lineage Tracing. *Cell* 2012; **148**: 33–45.

46 Baron CS, Oudenaarden A van. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat Rev Mol Cell Bio* 2019; **20**: 753–765.

47 VanHorn S, Morris SA. Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. *Dev Cell* 2020; **56**: 7–21.

48 Sankaran VG, Weissman JS, Zon LI. Cellular barcoding to decipher clonal dynamics in disease. *Science* 2022; **378**: eabm5874.

49 Woodworth MB, Girskis KM, Walsh CA. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat Rev Genet* 2017; **18**: 230–244.

50 Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat Rev Genet* 2020; **21**: 410–427.

51 Morgan D, Jost TA, Santiago CD, Brock A. Applications of high-resolution clone tracking technologies in cancer. *Curr Opin Biomed Eng* 2021; **19**: 100317.

52 Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 2007; **450**: 56–62.

53 Snippert HJ, Flier LG van der, Sato T, Es JH van, Born M van den, Kroon-Veenboer C *et al.* Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. *Cell* 2010; **143**: 134–144.

54 Kebschull JM, Zador AM. Cellular barcoding: lineage tracing, screening and beyond. *Nat Methods* 2018; **15**: 871–879.

55 Mao S, Zhang C, Chen R, Tang S, Fan X, Hu J. Cell lineage tracing: Methods, applications, and challenges. *Quant Biol* 2025; **13**. doi:10.1002/qub2.70006.

56 Bhang HC, Ruddy DA, Radhakrishna VK, Caushi JX, Zhao R, Hims MM *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* 2015; **21**: 440–448.

57 McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 2016; **353**: aaf7907.

58 Frieda KL, Linton JM, Hormoz S, Choi J, Chow K-HK, Singer ZS *et al.* Synthetic recording and in situ readout of lineage information in single cells. *Nature* 2017; **541**: 107–111.

59 Pei W, Feyerabend TB, Rössler J, Wang X, Postrach D, Busch K *et al.* Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* 2017; **548**: 456–460.

60 Raj B, Gagnon JA, Schier AF. Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR–Cas9 barcodes by scGESTALT. *Nat Protoc* 2018; **13**: 2685–2713.

61 Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 2018; **360**: 981–987.

62 Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* 2018; **564**: 219–224.

63 Al'Khafaji AM, Deatherage D, Brock A. Control of Lineage-Specific Gene Expression by Functionalized gRNA Barcodes. *Acs Synth Biol* 2018; **7**: 2468–2474.

64 Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat Biotechnol* 2018; **36**: 469–473.

65 Alemany A, Florescu M, Baron CS, Peterson-Maduro J, Oudenaarden A van. Whole-organism clone tracing using single-cell sequencing. *Nature* 2018; **556**: 108–112.

66 Hwang B, Lee W, Yum S-Y, Jeon Y, Cho N, Jang G *et al.* Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. *Nat Commun* 2019; **10**: 1234.

67 Bowling S, Sritharan D, Osorio FG, Nguyen M, Cheung P, Rodriguez-Fraticelli A *et al.* An Engineered CRISPR–Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* 2020; **181**: 1410-1422.e27.

68 Umkehrer C, Holstein F, Formenti L, Jude J, Froussios K, Neumann T *et al.* Isolating live cell clones from barcoded populations using CRISPRa-inducible reporters. *Nat Biotechnol* 2020; : 1–5.

69 Feldman D, Tsai F, Garrity AJ, O'Rourke R, Brenan L, Ho P *et al.* CloneSifter: enrichment of rare clones from heterogeneous cell populations. *BMC Biol* 2020; **18**: 177.

70 Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* 2020; **367**. doi:10.1126/science.aaw3381.

71 Pei W, Shang F, Wang X, Fanti A-K, Greco A, Busch K *et al.* Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. *Cell Stem Cell* 2020. doi:10.1016/j.stem.2020.07.018.

72 Askary A, Sanchez-Guardado L, Linton JM, Chadly DM, Budde MW, Cai L *et al.* In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. *Nat Biotechnol* 2020; **38**: 66–75.

73 Chow K-HK, Budde MW, Granados AA, Cabrera M, Yoon S, Cho S *et al.* Imaging cell lineage with a synthetic digital recording system. *Science* 2021; **372**: eabb3099.

74 Emert BL, Cote CJ, Torre EA, Dardani IP, Jiang CL, Jain N *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat Biotechnol* 2021; **39**: 865–876.

75 Loveless TB, Grotts JH, Schechter MW, Forouzmand E, Carlson CK, Agahi BS *et al.* Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat Chem Biol* 2021; **17**: 739–747.

76 Gutierrez C, Al'Khafaji AM, Brenner E, Johnson KE, Gohil SH, Lin Z *et al.* Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. *Nat Cancer* 2021; **2**: 758–772.

77 Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ *et al.* Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* 2021; **39**: 1150-1162.e9.

78 Oren Y, Tsabar M, Cuoco MS, Amir-Zilberstein L, Cabanos HF, Hütter J-C *et al.* Cycling cancer persister cells arise from lineages with distinct programs. *Nature* 2021; **596**: 576–582.

79 Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S *et al.* Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature* 2022; **601**: 125–131.

80 Jindal K, Adil MT, Yamaguchi N, Yang X, Wang HC, Kamimoto K *et al.* Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. *Nat Biotechnol* 2023; : 1–14.

81 Li L, Bowling S, McGeary SE, Yu Q, Lemke B, Alcedo K *et al.* A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* 2023; **186**: 5183-5199.e22.

82 Li L, Bowling S, Lin H, Chen D, Wang S-W, Camargo FD. DARLIN mouse for in vivo lineage tracing at high efficiency and clonal diversity. *Nat Protoc* 2025; **20**: 2319–2344.

83 Liu Y, Ban Y, Gao D. Oligo-CALL: A next-generation barcoding platform for studying resistance to targeted therapy. *Sci Adv* 2025; **11**: eadw9990.

84 Weber TS, Biben C, Miles DC, Glaser SP, Tomei S, Lin C-Y *et al.* LoxCode in vivo barcoding reveals epiblast clonal fate bias to fetal organs. *Cell* 2025; **188**: 3882-3896.e19.

85 Koblan LW, Yost KE, Zheng P, Colgan WN, Jones MG, Yang D *et al.* High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer. *Science* 2025; : eadx3800.

86 Serrano A, Berthelet J, Naik SH, Merino D. Mastering the use of cellular barcoding to explore cancer heterogeneity. *Nat Rev Cancer* 2022; : 1–16.

87 Haas L, Elewaut A, Gerard CL, Umkehrer C, Leiendecker L, Pedersen M *et al.* Acquired resistance to anti-MAPK targeted therapy confers an immune-evasive tumor microenvironment and cross-resistance to immunotherapy in melanoma. *Nat Cancer* 2021; : 1–16.

88 Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE *et al.* Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med* 2016; **22**: 262–269.

89 Howland KK, Brock A. Cellular barcoding tracks heterogeneous clones through selective pressures and phenotypic transitions. *Trends Cancer* 2023. doi:10.1016/j.trecan.2023.03.008.

90 Wroblewska A, Dhainaut M, Ben-Zvi B, Rose SA, Park ES, Amir E-AD *et al.* Protein Barcodes Enable High-Dimensional Single-Cell CRISPR Screens. *Cell* 2018; **175**: 1141-1155.e16.

91 Dhainaut M, Rose SA, Akturk G, Wroblewska A, Nielsen SR, Park ES *et al.* Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell* 2022; **185**: 1223-1239.e20.

92 Rovira-Clavé X, Drainas AP, Jiang S, Bai Y, Baron M, Zhu B *et al.* Spatial epitope barcoding reveals clonal tumor patch behaviors. *Cancer Cell* 2022. doi:10.1016/j.ccell.2022.09.014.

93 Weber K, Bartsch U, Stocking C, Fehse B. A Multicolor Panel of Novel Lentiviral "Gene Ontology" (LeGO) Vectors for Functional Gene Analysis. *Mol Ther* 2008; **16**: 698–706.

94 Berthelet J, Wimmer VC, Whitfield HJ, Serrano A, Boudier T, Mangiola S *et al.* The site of breast cancer metastases dictates their clonal composition and reversible transcriptomic profile. *Sci Adv* 2021; **7**: eabf4408.

95 Masuyama N, Konno N, Yachie N. Molecular recorders to track cellular events. *Science* 2022; **377**: 469–470.

96 Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B *et al.* Molecular recording of mammalian embryogenesis. *Nature* 2019; **570**: 77–82.

97 Askary A, Chen W, Choi J, Du LY, Elowitz MB, Gagnon JA *et al.* The lives of cells, recorded. *Nat Rev Genet* 2025; **26**: 203–222.

98 Perli SD, Cui CH, Lu TK. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* 2016; **353**. doi:10.1126/science.aag0511.

99 Tang W, Liu DR. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* 2018; **360**. doi:10.1126/science.aap8992.

100 Farzadfard F, Gharaei N, Higashikuni Y, Jung G, Cao J, Lu TK. Single-Nucleotide-Resolution Computing and Memory in Living Cells. *Mol Cell* 2019; **75**: 769-780.e4.

101 Choi J, Chen W, Minkina A, Chardon FM, Suiter CC, Regalado SG *et al.* A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* 2022; **608**: 98–107.

102 Chen W, Choi J. Molecular circuits for genomic recording of cellular events. *Trends Genet* 2025; **41**: 647–659.

103 Loveless TB, Carlson CK, Helmy CAD, Hu VJ, Ross SK, Demelo MC *et al.* Open-ended molecular recording of sequential cellular events into DNA. *Nat Chem Biol* 2025; **21**: 512–521.

104 Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 2023; **24**: 494–515.

105 Larson CR, Mandloi A, Acharyya S, Carstens JL. The tumor microenvironment across four dimensions: assessing space and time in cancer biology. *Front Immunol* 2025; **16**: 1554114.

106 Liu Y, Dai Y, Wang L. Spatial omics at the forefront: emerging technologies, analytical innovations, and clinical applications. *Cancer Cell* 2025. doi:10.1016/j.ccell.2025.12.009.

107 Gong D, Arbesfeld-Qiu JM, Perrault E, Bae JW, Hwang WL. Spatial oncology: Translating contextual biology to the clinic. *Cancer Cell* 2024. doi:10.1016/j.ccell.2024.09.001.

108 Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* 2022; **19**: 534–546.

109 Toninelli M, Rossetti G, Pagani M. Charting the tumor microenvironment with spatial profiling technologies. *Trends Cancer* 2023. doi:10.1016/j.trecan.2023.08.004.

110 Ren J, Luo S, Shi H, Wang X. Spatial omics advances for in situ RNA biology. *Mol Cell* 2024. doi:10.1016/j.molcel.2024.08.002.

111 Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015; **348**: aaa6090.

112 Shah S, Takei Y, Zhou W, Lubeck E, Yun J, Eng C-HL *et al.* Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* 2018; **174**: 363-376.e16.

113 Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019; **568**: 235–239.

114 He S, Bhatt R, Brown C, Brown EA, Buhr DL, Chantranuvatana K *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol* 2022; **40**: 1794–1806.

115 Borm LE, Albiach AM, Mannens CCA, Janusauskas J, Özgün C, Fernández-García D *et al.* Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat Biotechnol* 2023; **41**: 222–231.

116 Janesick A, Shelansky R, Gottscho AD, Wagner F, Williams SR, Rouault M *et al.* High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat Commun* 2023; **14**: 8353.

117 Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* 2014; **343**: 1360–1363.

118 Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018; **361**: eaat5691.

119 Chen X, Sun Y-C, Church GM, Lee JH, Zador AM. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res* 2018; **46**: e22–e22.

120 Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019; **363**: 1463–1467.

121 Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Bella DJD *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 2021; **39**: 313–319.

122 Liu Y, Yang M, Deng Y, Su G, Enninful A, Guo CC *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* 2020; **183**: 1665-1681.e18.

123 Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022; **185**: 1777-1792.e21.

124 Moffitt JR, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. *Nat Rev Genet* 2022; **23**: 741–759.

125 Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. *Nature* 2021; **596**: 211–220.

126 Khafizov R, Piazza E, Cui Y, Patrick M, Metzger E, McGuire D *et al.* Sub-cellular Imaging of the Entire Protein-Coding Human Transcriptome (18933-plex) on FFPE Tissue Using Spatial Molecular Imaging. *bioRxiv* 2024; : 2024.11.27.625536.

127 Lin J, Fallahi-Sichani M, Chen J, Sorger PK. Cyclic Immunofluorescence (CycIF), A Highly Multiplexed Method for Single-cell Imaging. *Curr Protoc Chem Biol* 2016; **8**: 251–264.

128 Radtke AJ, Kandov E, Lowekamp B, Speranza E, Chu CJ, Gola A *et al.* IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. *Proc Natl Acad Sci* 2020; **117**: 33455–33465.

129 Rivest F, Eroglu D, Pelz B, Kowal J, Kehren A, Navikas V *et al.* Fully automated sequential immunofluorescence (seqIF) for hyperplex spatial proteomics. *Sci Rep* 2023; **13**: 16994.

130 Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 2018; **174**: 968-981.e15.

131 Baharlou H, Canete NP, Cunningham AL, Harman AN, Patrick E. Mass Cytometry Imaging for the Study of Human Diseases—Applications and Data Analysis Strategies. *Front Immunol* 2019; **10**: 2657.

132 Rajbhandari P, Neelakantan TV, Hosny N, Stockwell BR. Spatial pharmacology using mass spectrometry imaging. *Trends Pharmacol Sci* 2024; **45**: 67–80.

133 Zhu X, Xu T, Peng C, Wu S. Advances in MALDI Mass Spectrometry Imaging Single Cell and Tissues. *Front Chem* 2022; **9**: 782432.

134 Leopold J, Prabutzki P, Engel KM, Schiller J. A Five-Year Update on Matrix Compounds for MALDI-MS Analysis of Lipids. *Biomolecules* 2023; **13**: 546.

135 He MJ, Pu W, Wang X, Zhang W, Tang D, Dai Y. Comparing DESI-MSI and MALDI-MSI Mediated Spatial Metabolomics and Their Applications in Cancer Studies. *Front Oncol* 2022; **12**: 891018.

136 Dreisewerd K, Bien T, Soltwisch J. Mass Spectrometry Imaging of Small Molecules, Methods and Protocols. *Methods Mol Biol* 2021; **2437**: 21–40.

137 Duckworth AD, Gherardini PF, Sykorova M, Yasin F, Nolan GP, Slupsky JR *et al.* Multiplexed profiling of RNA and protein expression signatures in individual cells using flow or mass cytometry. *Nat Protoc* 2019; **14**: 901–920.

138 Guruprasad KP, Vasudev V, Agrawal H, Thakur M, Krishan A, Sobti RC. Flow Cytometry, Principles, Applications, and Clinical Perspectives. 2024; : 1–25.

139 Lalli E, Gibellini D, Santi S, Facchini A. In situ hybridization in suspension and flow cytometry as a tool for the study of gene expression. *Anal Biochem* 1992; **207**: 298–303.

140 Porichis F, Hart MG, Griesbeck M, Everett HL, Hassan M, Baxter AE *et al.* High-throughput detection of miRNAs and gene-specific mRNA at the single-cell level by flow cytometry. *Nat Commun* 2014; **5**: 5641.

141 Frei AP, Bava FA, Zunder ER, Hsieh EW, Chen S-Y, Nolan GP *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat Methods* 2016; **13**: 269–275.

142 Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. *Cell* 2016; **165**: 780–791.

143 Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* 2023; **24**: 695–713.

144 Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* 2016; **13**: 229–232.

145 Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018; **361**: 1380–1385.

146 Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017; **14**: 865–868.

147 Zeng H, Huang J, Zhou H, Meilandt WJ, Dejanovic B, Zhou Y *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat Neurosci* 2023; **26**: 430–446.

148 Liu Y, DiStasio M, Su G, Asashima H, Enninful A, Qin X *et al.* High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat Biotechnol* 2023; : 1–5.

149 Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* 2020; **38**: 586–599.

150 Schulz D, Zanotelli VRT, Fischer JR, Schapiro D, Engler S, Lun X-K *et al.* Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Syst* 2018; **6**: 25-36.e5.

151 Cheng Y, Burrack RK, Li Q. Spatially Resolved and Highly Multiplexed Protein and RNA In Situ Detection by Combining CODEX With RNAscope In Situ Hybridization. *J Histochem Cytochem* 2022; **70**: 571–581.

152 Geier B, Sogin EM, Michellod D, Janda M, Kompauer M, Spengler B *et al.* Spatial metabolomics of in situ host–microbe interactions at the micrometre scale. *Nat Microbiol* 2020; **5**: 498–510.

153 Sun C, Wang A, Zhou Y, Chen P, Wang X, Huang J *et al.* Spatially resolved multi-omics highlights cell-specific metabolic remodeling and interactions in gastric cancer. *Nat Commun* 2023; **14**: 2692.

154 Vicari M, Mirzazadeh R, Nilsson A, Shariatgorji R, Bjärterot P, Larsson L *et al.* Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nat Biotechnol* 2024; **42**: 1046–1050.

155 Godfrey TM, Shanneik Y, Zhang W, Tran T, Verbeeck N, Patterson NH *et al.* Integrating Ambient Ionization Mass Spectrometry Imaging and Spatial Transcriptomics on the Same Cancer Tissues to Identify RNA–Metabolite Correlations. *Angew Chem Int Ed* 2025; **64**: e202502028.

156 Li X, Huang Y, Wang S, Li Y, Jiang F, Gao J *et al.* OpenFISH enables integrated high-resolution spatial transcriptomics and metabolomics on a single tissue section. *bioRxiv* 2025; : 2025.08.19.671030.

157 Nadalin F, Marzi MJ, Piscazzi MP, Fuentes-Bravo P, Procaccia S, Climent M *et al.* Multi-omic lineage tracing predicts the transcriptional, epigenetic and genetic determinants of cancer evolution. *Nat Commun* 2024; **15**: 7609.

158 Yang D, Jones MG, Naranjo S, Rideout WM, Min KH (Joseph), Ho R *et al.* Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* 2022; **185**: 1905-1923.e25.

159 Jones MG, Sun D, Min KH (Joseph), Colgan WN, Tian L, Weir JA *et al.* Spatiotemporal lineage tracing reveals the dynamic spatial architecture of tumor growth and metastasis. 2024. doi:10.1101/2024.10.21.619529.

160 Koblan LW, Yost KE, Zheng P, Colgan WN, Jones MG, Yang D *et al.* High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer. *Science* 2025; : eadx3800.

161 Liao H, Choi J, Shendure J. Molecular recording using DNA Typewriter. *Nat Protoc* 2024; **19**: 2833–2862.

162 NEBuilder. https://nebuilder.neb.com/#!/ (accessed 1 Feb2026).

163 Wong ASL, Choi GCG, Cui CH, Pregernig G, Milani P, Adam M *et al.* Multiplexed barcoded CRISPR–Cas9 screening enabled by CombiGEM. *Proc National Acad Sci* 2016; **113**: 2544–2549.

164 Dannhorn A, Kazanc E, Ling S, Nikula C, Karali E, Serra MP *et al.* Universal Sample Preparation Unlocking Multimodal Molecular Tissue Imaging. *Anal Chem* 2020; **92**: 11080–11088.

165 Dannhorn A, Kazanc E, Flint L, Guo F, Carter A, Hall AR *et al.* Morphological and molecular preservation through universal preparation of fresh-frozen tissue samples for multimodal imaging workflows. *Nat Protoc* 2024; : 1–27.

166 Soja E, Sivajothi S, Flynn W, Courtois ET. Hematoxylin and Eosin (H&E) Staining of Tissues following Multiplexed Imaging on Phenocycler-Fusion. 2024.https://www.protocols.io/view/hematoxylin-and-eosin-h-amp-e-staining-of-tissues-36wgqn55xgk5/v1 (accessed 1 Feb2026).

167 Pitino E, Pascual-Reguant A, Segato-Dezem F, Wise K, Salvador-Martinez I, Crowell HL *et al.* STAMP: Single-cell transcriptomics analysis and multimodal profiling through imaging. *Cell* 2025. doi:10.1016/j.cell.2025.05.027.

168 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ* 2011; **17**: 10–12.

169 Gaspar JM. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinform* 2018; **19**: 536.

170 Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 2012; **30**: 918–920.

171 Conroy MJ, Andrews RM, Andrews S, Cockayne L, Dennis EA, Fahy E *et al.* LIPID MAPS: update to databases and tools for the lipidomics community. *Nucleic Acids Res* 2023; **52**: D1677–D1682.

172 Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res* 2007; **35**: D527–D532.

173 Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017; **7**: 16878.

174 Schmidt U, Weigert M, Broaddus C, Myers G. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II. *Lect Notes Comput Sci* 2018; : 265–273.

175 Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; **9**: 676–682.

176 Wickham H. ggplot2, Elegant Graphics for Data Analysis. *R* 2016. doi:10.1007/978-3-319-24277-4.

177 Young AP, Jackson DJ, Wyeth RC. A technical review and guide to RNA fluorescence in situ hybridization. *Peerj* 2020; **8**: e8806.

178 Elhanani O, Ben-Uri R, Keren L. Spatial profiling technologies illuminate the tumor microenvironment. *Cancer Cell* 2023. doi:10.1016/j.ccell.2023.01.010.

179 Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 2018; **145**: dev165753.

180 Cheng Y, Dang S, Zhang Y, Chen Y, Yu R, Liu M *et al.* Sequencing-free whole-genome spatial transcriptomics at single-molecule resolution. *Cell* 2025. doi:10.1016/j.cell.2025.09.006.

181 Sountoulidis A, Liontos A, Nguyen HP, Firsova AB, Fysikopoulos A, Qian X *et al.* SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLoS Biol* 2020; **18**: e3000675.

182 Sari AIP, Copeland K, Nuwongsri P, Pipatsakulroj W, Jinawath A, Israsena N *et al.* RNAscope Multiplex FISH Signal Assessment in FFPE and Fresh Frozen Tissues: The Effect of Archival Duration on RNA Expression. *J Histochem Cytochem* 2024; **73**: 9–28.

183 Saunders RA, Allen WE, Pan X, Sandhu J, Lu J, Lau TK *et al.* Perturb-Multimodal: A platform for pooled genetic screens with imaging and sequencing in intact mammalian tissue. *Cell* 2025. doi:10.1016/j.cell.2025.05.022.

184 Cheng Y, Hu M, Yang B, Jensen TB, Zhang Y, Yang T *et al.* Perturb-tracing enables high-content screening of multi-scale 3D genome regulators. *Nat Methods* 2025; **22**: 950–961.

185 Liu S, Punthambaker S, Iyer EPR, Ferrante T, Goodwin D, Fürth D *et al.* Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel in situ analyses. *Nucleic Acids Res* 2021; **49**: e58–e58.

186 Foster BA, Gingrich JR, Kwon ED, Madias C, Greenberg NM. Characterization of prostatic epithelial cell lines derived from transgenic adenocarcinoma of the mouse prostate (TRAMP) model. *Cancer Res* 1997; **57**: 3325–30.

187 Deng R, Zhang K, Sun Y, Ren X, Li J. Highly specific imaging of mRNA in single cells by target RNA-initiated rolling circle amplification. *Chem Sci* 2017; **8**: 3668–3675.

188 Kim EN, Chen PZ, Bressan D, Tripathi M, Miremadi A, Pietro M di *et al.* Dual-modality imaging of immunofluorescence and imaging mass cytometry for whole-slide imaging and accurate segmentation. *Cell Rep Methods* 2023; **3**: 100595.

189 Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 2017; **14**: 297–301.

190 Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X *et al.* Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal Chem* 2009; **81**: 6813–6822.

191 Ornatsky O, Bandura D, Baranov V, Nitz M, Winnik MA, Tanner S. Highly multiparametric analysis by mass cytometry. *J Immunol Methods* 2010; **361**: 1–20.

192 Zunder ER, Finck R, Behbehani GK, Amir ED, Krishnaswamy S, Gonzalez VD *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat Protoc* 2015; **10**: 316–333.

193 Qin X, Sufi J, Vlckova P, Kyriakidou P, Acton SE, Li VSW *et al.* Cell-type-specific signaling networks in heterocellular organoids. *Nat Methods* 2020; **17**: 335–342.

194 Sufi J, Qin X, Rodriguez FC, Bu YJ, Vlckova P, Zapatero MR *et al.* Multiplexed single-cell analysis of organoid signaling networks. *Nat Protoc* 2021; **16**: 4897–4918.

195 Bai Z, Zhang D, Gao Y, Tao B, Zhang D, Bao S *et al.* Spatially exploring RNA biology in archival formalin-fixed paraffin-embedded tissues. *Cell* 2024; **187**: 6760-6779.e24.

196 Matsubara T, Soh J, Morita M, Uwabo T, Tomida S, Fujiwara T *et al.* DV200 Index for Assessing RNA Integrity in Next-Generation Sequencing. *BioMed Res Int* 2020; **2020**: 9349132.

197 Jacobsen SB, Tfelt-Hansen J, Smerup MH, Andersen JD, Morling N. Comparison of whole transcriptome sequencing of fresh, frozen, and formalin-fixed, paraffin-embedded cardiac tissue. *PLOS ONE* 2023; **18**: e0283159.

198 Grases D, Porta-Pardo E. A practical guide to spatial transcriptomics: lessons from over 1000 samples. *Trends Biotechnol* 2025. doi:10.1016/j.tibtech.2025.08.020.

199 Aichler M, Walch A. MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab Investig* 2015; **95**: 422–431.

200 Tuck M, Grélard F, Blanc L, Desbenoit N. MALDI-MSI Towards Multimodal Imaging: Challenges and Perspectives. *Front Chem* 2022; **10**: 904688.

201 Kennedy-Darling J, Bhate SS, Hickey JW, Black S, Barlow GL, Vazquez G *et al.* Highly multiplexed tissue imaging using repeated oligonucleotide exchange reaction. *Eur J Immunol* 2021; **51**: 1262–1277.

202 Black S, Phillips D, Hickey JW, Kennedy-Darling J, Venkataraaman VG, Samusik N *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat Protoc* 2021; **16**: 3802–3835.

203 Sadeghipour A, Babaheidarian P. Biobanking, Methods and Protocols. *Methods Mol Biol* 2018; **1897**: 253–268.

204 Qiu Y, Wei K, Lin H, Liu Y, Lin C, Ke R. Combined amplification-based single-molecule fluorescence in situ hybridization with immunofluorescence for simultaneous in situ detection of RNAs and proteins. *Biochem Biophys Res Commun* 2024; **696**: 149508.

205 Chaudhuri AD, Yelamanchili SV, Fox HS. Combined fluorescent in situ hybridization for detection of microRNAs and immunofluorescent labeling for cell-type markers. *Front Cell Neurosci* 2013; **7**: 160.

206 Bourceau P, Geier B, Suerdieck V, Bien T, Soltwisch J, Dreisewerd K *et al.* Visualization of metabolites and microbes at high spatial resolution using MALDI mass spectrometry imaging and in situ fluorescence labeling. *Nat Protoc* 2023; : 1–30.

207 Sans M, Chen Y, Thege FI, Dou R, Min J, Yip-Schneider M *et al.* Integrated spatial transcriptomics and lipidomics of precursor lesions of pancreatic cancer identifies enrichment of long chain sulfatide biosynthesis as an early metabolic alteration. *bioRxiv : Prepr Serv Biol* 2023. doi:10.1101/2023.08.14.553002.

208 Ben-Chetrit N, Niu X, Swett AD, Sotelo J, Jiao MS, Stewart CM *et al.* Integration of whole transcriptome spatial profiling with protein markers. *Nat Biotechnol* 2023; : 1–6.

209 Truong JXM, Spotbeen X, White J, Swinnen JV, Butler LM, Snel MF *et al.* Removal of optimal cutting temperature (O.C.T.) compound from embedded tissue for MALDI imaging of lipids. *Anal Bioanal Chem* 2021; **413**: 2695–2708.

210 Monaghan J, Woytowich N, Zhao T, Nguyen K, Mahony E, Lum JJ *et al.* Ambient mass spectrometry imaging enables spatial metabolomics of optimal cutting temperature compound (OCT)-embedded tumors. 2025. doi:10.1101/2025.11.05.686836.

211 Körber A, Anthony IGM, Heeren RMA. Mass Spectrometry Imaging. *Anal Chem* 2025; **97**: 15517–15549.

212 Strupat K, Karas M, Hillenkamp F. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry. *Int J Mass Spectrom Ion Process* 1991; **111**: 89–102.

213 Teearu A, Vahur S, Haljasorg U, Leito I, Haljasorg T, Toom L. 2,5-Dihydroxybenzoic acid solution in MALDI-MS: ageing and use for mass calibration. *J Mass Spectrom* 2014; **49**: 970–979.

214 Jacksén J, Emmer Å. Evaluation of 2,6-dihydroxyacetophenone as matrix-assisted laser desorption/ionization matrix for analysis of hydrophobic proteins and peptides. *Anal Biochem* 2012; **425**: 18–20.

215 Kaya I, Michno W, Brinet D, Iacone Y, Zanni G, Blennow K *et al.* Histology-Compatible MALDI Mass Spectrometry Based Imaging of Neuronal Lipids for Subsequent Immunofluorescent Staining. *Anal Chem* 2017; **89**: 4685–4694.

216 Strittmatter N, Richards FM, Race AM, Ling S, Sutton D, Nilsson A *et al.* Method To Visualize the Intratumor Distribution and Impact of Gemcitabine in Pancreatic Ductal Adenocarcinoma by Multimodal Imaging. *Anal Chem* 2022; **94**: 1795–1803.

217 Nunes JB, Ijsselsteijn ME, Abdelaal T, Ursem R, Ploeg M van der, Giera M *et al.* Integration of mass cytometry and mass spectrometry imaging for spatially resolved single-cell metabolic profiling. *Nat Methods* 2024; **21**: 1796–1800.

218 Zhou Q, Fülöp A, Hopf C. Recent developments of novel matrices and on-tissue chemical derivatization reagents for MALDI-MSI. *Anal Bioanal Chem* 2021; **413**: 2599–2617.

219 Perry WJ, Patterson NH, Prentice BM, Neumann EK, Caprioli RM, Spraggins JM. Uncovering matrix effects on lipid analyses in MALDI imaging mass spectrometry experiments. *J Mass Spectrom* 2020; **55**: e4491.

220 Baker TC, Han J, Borchers CH. Recent advancements in matrix-assisted laser desorption/ionization mass spectrometry imaging. *Curr Opin Biotechnol* 2017; **43**: 62–69.

221 Hermann J, Noels H, Theelen W, Lellig M, Orth-Alampour S, Boor P *et al.* Sample preparation of formalin-fixed paraffin-embedded tissue sections for MALDI-mass spectrometry imaging. *Anal Bioanal Chem* 2020; **412**: 1263–1275.

222 Ma X, Fernández FM. Advances in mass spectrometry imaging for spatial cancer metabolomics. *Mass Spectrom Rev* 2022; : e21804.

223 Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer* 2015; **15**: 701–711.

224 Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI *et al.* Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol* 2021; **18**: 79–92.

225 Watson PA, Ellwood-Yen K, King JC, Wongvipat J, LeBeau MM, Sawyers CL. Context-Dependent Hormone-Refractory Progression Revealed through Characterization of a Novel Murine Prostate Cancer Cell Line. *Cancer Res* 2005; **65**: 11565–11571.

226 Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R *et al.* Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* 2003; **4**: 223–238.

227 Ammirante M, Luo J-L, Grivennikov S, Nedospasov S, Karin M. B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. *Nature* 2010; **464**: 302–305.

228 Shen Y-C, Ghasemzadeh A, Kochel CM, Nirschl TR, Francica BJ, Lopez-Bujanda ZA *et al.* Combining intratumoral Treg depletion with androgen deprivation therapy (ADT): preclinical activity in the Myc-CaP model. *Prostate Cancer Prostatic Dis* 2018; **21**: 113–125.

229 Lyu A, Fan Z, Clark M, Lea A, Luong D, Setayesh A *et al.* Evolution of myeloid-mediated immunotherapy resistance in prostate cancer. *Nature* 2024; : 1–11.

230 Al-janabi H, Moyes K, Allen R, Fisher M, Crespo M, Gurel B *et al.* Targeting a STING agonist to perivascular macrophages in prostate tumors delays resistance to androgen deprivation therapy. *J Immunother Cancer* 2024; **12**: e009368.

231 Wang Y, Herroon MK, Zielske SP, Ellis L, Podgorski I, Taichman RS *et al.* Use of FVB Myc-CaP cells as an immune competent, androgen receptor positive, mouse model of prostate cancer bone metastasis. *J Bone Oncol* 2021; **30**: 100386.

232 Lee E, Zhang Z, Chen C-C, Choi D, Rivera ACA, Linton E *et al.* Timing of treatment shapes the path to androgen receptor signaling inhibitor resistance in prostate cancer. *bioRxiv* 2024; : 2024.03.18.585532.

233 Rojas F, Hernandez S, Lazcano R, Laberiano-Fernandez C, Parra ER. Multiplex Immunofluorescence and the Digital Image Analysis Workflow for Evaluation of the Tumor Immune Environment in Translational Research. *Front Oncol* 2022; **12**: 889886.

234 Souza N de, Zhao S, Bodenmiller B. Multiplex protein imaging in tumour biology. *Nat Rev Cancer* 2024; **24**: 171–191.

235 Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD *et al.* Multiplexed ion beam imaging of human breast tumors. *Nat Med* 2014; **20**: 436–442.