UTILITY OF A LIQUID BIOPSY TO IDENTIFY INTRA-TUMOURAL MOLECULAR HETEROGENEITY IN HEAD AND NECK SQUAMOUS CELL CARCINOMA: THE ROLE OF CIRCULATING TUMOUR CELLS

by

KARL FREDERICK BRAEKKAN PAYNE

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Supervisors:

Professor Hisham Mehanna

Dr Paul Nankivell

Dr Graham Taylor

Dr Chris Yau

Institute of Head and Neck Studies and Education

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

University of Birmingham 2022

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

BACKGROUND: Circulating tumour cells (CTCs) in the blood of cancer patients are a potentially rich source of biomarkers to guide cancer therapy, particularly when tumours cannot be sampled directly. The optimal method of CTC enrichment/isolation and subsequent downstream characterisation remains unclear. Current marker dependent platforms are biased towards specific CTC sub-groups, and CTC characterisation focuses on genomic characterisation, being unable to detect post-transcriptional modification.

AIMS: The aims of this project were: i) develop blood sampling methods able to preserve key CTC parameters with minimal sampling handling, ii) optimise a marker-independent CTC enrichment technique, the Parsortix microfluidic platform, for use in head and neck squamous cell carcinoma (HNSCC), iii) investigate methods for direct proteomic characterisation of CTCs to detect relevant post translational modifications. Collectively these aims would streamline sample collection thus simplifying and enhancing national/international multi-centre CTC research studies and trials while allowing deeper characterisation of CTC biology.

METHODS: A mock CTC model using the FaDu and SCC047 HNSCC cell lines spiked into healthy donor blood was developed and used to optimise the Parsortix platform and downstream immunofluorescence microscopy and flow cytometry characterisation. Subsequently, a mass cytometry antibody panel of 44 markers (including epithelial/EMT, proliferative, stemness, immune and phosphorylated signalling proteins) was optimised for staining quality on Transfix fixed cells, following Parsortix enrichment. Pre-treatment blood samples from HNSCC patients were prospectively recruited through the Accelerated2 sample collection platform.

RESULTS: In cell line spiking experiments, non-fixative EDTA blood collection tubes (BCTs) enriched with Parsortix demonstrated a mean capture rate of 53.5%, across a range of 9-129 cells/ml spiked concentrations (n=13). Transfix fixation BCTs demonstrated significantly improved capture rates of spiked cells at 0hr, 24hr and 72hr timepoints, when compared to EDTA BCTs. Parsortix enrichment significantly altered the gene expression of unfixed cells, causing downregulation of genes associated with RNA and ribosomal/protein processing genes and upregulation of genes associated with oxidative stress and cell injury/apoptosis.

Using immunofluorescence microscopy, CTCs were identified from HNSCC patient samples (enriched with Parsortix) with epithelial (EpCAM expression) and epithelial-mesenchymal transition (EMT, EpCAM and N-cadherin expression) characteristics. In a cohort of 20 patients, flow cytometry characterisation of the above markers was able to quantify and characterise CTCs. Sixty-five percent (13/20) demonstrated CTCs, at a mean count of 4 CTCs/ml (range 2-11.2 CTCs/ml). The presence of CTCs correlated with advanced stage of disease (p=0.0121), but not T or N stage. Neither the presence of epithelial nor EMT CTC sub-groups correlated with progression-free or overall survival. A mesenchymal gene expression profile in patient matched primary tumour tissue did not positively correlate with CTC EMT expression (p=0.347).

Mass cytometry analysis identified CTCs in 11 of 12 patient blood samples, with considerable CTC heterogeneity. Novel CTC sub-groups, common between patient samples, were defined based upon target marker expression patterns - for example EMT CTCs expressing proliferative markers and decreased immune-checkpoint markers. Mass cytometry outperformed bulk quantitative gene expression profiling of Parsortix enriched samples from the same cohort – both to identify the presence of CTCs and phenotypic sub-groups. A mesenchymal gene expression profile in primary tumour tissue was significantly associated with an increased proportion of proliferative and immune-checkpoint high CTCs.

DISCUSSION: We demonstrate for the first time that the combination of Parsortix microfluidic enrichment and mass cytometry can be successfully used to provide multi-parameter single-CTC characterisation of extracellular, intracellular and activated signalling proteins – a depth of information beyond that available from bulk CTC gene expression profiling. Even in a small cohort pilot study, we highlight an interesting trend that primary tumours with mesenchymal gene expression profiles exhibit increased levels of proliferative and immune-checkpoint high CTCs. The heterogeneity of CTC EMT and phenotypic sub-groups and their relationship to immune-checkpoint markers is notable – providing both an avenue for therapeutic targeting, but also a means of therapeutic escape in ‘immune low’ CTCs. We lay a platform for future proteomic based single-CTC studies in larger cohorts.

This thesis is dedicated to my grandfather, Dr Olaf Rikard Braekkan - recipient of Norway's highest civilian honour, *The Royal Norwegian Order of St. Olav*, for services to science.

Unfortunately, he didn't live long enough to see me follow in his footsteps,

but I think he would have enjoyed reading this thesis.

ACKNOWLEDGEMENTS

I acknowledge Birmingham Health Partners for funding year one of this PhD through an Institute of Translational Medicine starter fellowship. In addition, consumables during year one was funded by separate grants from the British Association of Oral and Maxillofacial Surgeons, British Association of Head and Neck Oncologists and Royal College of Surgeons of England, the latter via an honorary research fellowship. Years two and three of this PhD were fully funded by a Cancer Research UK Clinical PhD fellowship from the Birmingham Cancer Research UK Centre.

I acknowledge the assistance of Dr Naeem Kahn in mass cytometry panel design and Dr Tony Ly for mass spectrometry processing and data analysis. In addition, Dr Boris Noyvert, Dr Yi Pan and Dr Mohammed Elasrag for help with RNA sequencing data analysis. I also acknowledge the invaluable contribution from lab technician Gordon Taylor and research nurse June Jones. Finally, I sincerely thank Dr Jill Brooks for her continued sage advice throughout this project.

COVID-19 PANDEMIC IMPACT STATEMENT

The COVID-19 global pandemic had a significant impact upon the research conducted within this PhD. Firstly, University of Birmingham laboratories were closed for several months during the first UK lockdown in 2020, during which time all wet-lab research ceased. Secondly, recruitment of patients was halted during both lockdowns and once restrictions were lifted the reduction in elective operating significantly impacted patient recruitment (dropping to less than a third of pre-COVID recruitment levels).

The specific impact of these restrictions upon this project were as follows:

- Several planned cell line optimisation experiments in chapter 5 were halted due to lost laboratory time to enable progress analysing patient samples in chapters 6 and 7.
- Unfortunately, due to the first national lockdown, we were unable to collect follow-up samples in the patient cohort in chapter 6, thus missing the opportunity to investigate CTC temporal heterogeneity.
- The patient cohort in chapter 7 was significantly reduced in size due to slow recruitment after the lifting of lockdown restrictions. Our initial aim had been to process >30 samples, but the final pilot cohort was made up of 12 patients.
- With restrictions on time and sample availability we were unable to optimise a protocol for tumour tissue mass cytometry analysis, to allow direct comparison to CTC mass cytometry results reported in chapter 7.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS v
COVID-19 PANDEMIC IMPACT STATEMENT vi
TABLE OF CONTENTS vii
LIST OF FIGURES xii
LIST OF TABLES xiv

CHAPTER 1: INTRODUCTION 1
1.1 Head and neck squamous cell carcinoma 1
1.1.1 Incidence and aetiology 1
1.1.2 Staging and treatment 2
1.2 The cancer genome 6
1.2.1 DNA mutations 6
1.2.2 The genomic landscape of HNSCC 11
1.3 Intra-tumoural heterogeneity (ITH) 14
1.3.1 Spatial and temporal molecular heterogeneity 14
1.3.2 Single cell analysis to identify ITH 18
1.3.3 ITH in HNSCC and its clinical impact 20
1.4 Circulating tumour cells (CTCs) as a liquid biopsy 24
1.4.1 What is a liquid biopsy? 24
1.4.2 CTCs and their role in metastasis 26
1.4.3 CTC isolation/enrichment techniques 30
1.4.4 Storage and preservation of CTCs in blood samples 32
1.4.5 CTC molecular characterisation 36
1.4.6 Single-CTC sequencing and ITH 38
1.5 CTCs in HNSCC 40
1.5.1 CTC phenotypes in HNSCC and their use as diagnostic markers 40
1.5.2 CTC prognostic risk stratification in HNSCC 44
1.5.3 CTC derived predictive and therapeutic biomarkers in HNSCC 46
1.6 Summary – unanswered questions and future directions 51

CHAPTER 2: CTC EXPRESSION OF IMMUNE-MARKERS IN HNSCC:
A SYSTEMATIC REVIEW AND META-ANALYSIS 52
2.1 Introduction 52
2.2 Methods 54
2.3 Results 56
2.4 Discussion 65
2.5 Conclusion 70

CHAPTER 3: AIMS AND OBJECTIVES 71
3.1 Research question 71
3.2 Aims 71
3.3 Objectives 71

CHAPTER 4: METHODS 73
4.1 Reagents 73
4.2 HNSCC cell line culture and harvest 74
4.2.1 Cell line culture 74
4.2.2 Cell line harvest 74
4.3 Cell line spiking of healthy donor blood 75
4.4 Parsortix cell line and CTC enrichment, antibody staining and immunofluorescence microscopy 76
4.4.1 Cell line/CTC enrichment 76
4.4.2 Antibody staining 76
4.4.3 Elution of the Parsortix enriched cell population 77
4.5 Negative depletion protocol 77
4.6 Flow cytometry of Parsortix enriched cells 78
4.7 Immunofluorescence microscopy of Parsortix enriched and antibody stained cells 79
4.8 PBMC stimulation for phospho-antibody optimisation experiments 79
4.9 Mass cytometry sample protocol 80
4.10 Establishing a confidence threshold for cytometry based positive Identification of CTCs 83

4.11 RNA extraction, sequencing and analysis 84
4.11.1 RNA extraction from tissue, plasma and CTCs 84
4.11.2 RNA sequencing 84
4.11.3 Bioinformatics pipelines 84
4.12 Mass spectrometry protocol and data analysis 85
4.13 Patient sample recruitment and data capture 86

CHAPTER 5: OPTIMISATION OF THE PARSORTIX DEVICE USING A CELL LINE MODEL 87
5.1 The Parsortix CTC enrichment device 87
5.2. Aims and objectives 91
5.3 Capture rate of HNSCC spiked cells by the Parsortix device 92
5.3.1 The impact of spiked cell count on Parsortix capture rate 92
5.3.2 Investigating cell preserving blood collection tubes and sample storage time 94
5.4 Impact of Parsortix microfluidic enrichment on gene expression 97
5.5 Optimising RNA extraction from the Parsortix enriched cell population 101
5.6 Gene expression profiling of Parsortix enriched cells 103
5.7 Summary of key findings 106

CHAPTER 6: CHARACTERISING CTC SUB-POPULATIONS IN HNSCC USING IMMUNOFLOURESCENCE AND FLOW CYTOMETRY 107
6.1 Moving from *in-vitro* to *ex-vivo* CTC characterisation 107
6.2 Aims and objectives 107
6.3 Antibody optimisation 108
6.4 Visualisation of Parsortix enriched CTCs stained with fluorophore labelled antibodies 110
6.5 Flow cytometry characterisation of CTCs 117
6.5.1 Gating strategy 117
6.5.2 Patient cohort 119
6.5.3 Correlating CTC count to tumour pathological variables 123
6.5.4 Baseline CTC presence as a prognostic biomarker of clinical outcome 125
6.5.5 Longitudinal patient samples 129

6.5.6 Comparison of tumour and CTC EMT profiles 130
6.6 Summary of key findings 135

CHAPTER 7: MASS CYTOMETRY SINGLE-CELL CHARACTERISATION
OF CTCs IN HNSCC 136
7.1 Introduction to mass cytometry 136
7.2 Aims and objectives 137
7.3 Marker choice, antibody optimisation and panel design 138
7.3.1 Rationalising marker inclusion 138
7.3.2 Antibody optimisation 142
7.3.3 Mass cytometry panel design and antibody titration 148
7.3.4 Discussing the subjective and iterative process of antibody optimisation, titration and panel design in mass cytometry 153
7.3.5 Pilot test on Parsortix enriched blood sample 155
7.4 Mass cytometry characterisation of Parsortix enriched patient blood samples 158
7.4.1 Patient sample processing and data analysis 158
7.4.2 Clustering analysis identifies novel phenotypic sub-groups 165
7.4.3 Phosphorylated signalling proteins underpinning CTC sub-groups 175
7.5 Comparison of mass cytometry to RNAseq CTC and tumour characterisation 176
7.5.1 Gene expression profiling of Parsortix enriched samples 176
7.5.2 Correlating tumour EMT profile and CTC sub-groups 180
7.6 Exploring pan-proteomic CTC analysis 182
7.7 Summary of key findings 185

CHAPTER 8: DISCUSSION 186
8.1 Summary of thesis findings 186
8.2 Parsortix demonstrates clinical utility in a HNSCC cell line model 187
8.3 Flow cytometry characterises epithelial and EMT CTC sub-groups 192
8.4 Mass cytometry identifies previously unseen CTC heterogeneity in HNSCC 197
8.4.1 CTC EMT plasticity 197
8.4.2 Phenotypic diversity of CTC sub-groups 200
8.4.3 Signalling pathway activation in CTC sub-groups 201

8.4.4 Immune checkpoint expression 203
8.4.5 Comparison of mass cytometry to gene expression profiling of CTCs and tumour samples 204
8.5 Proteomic versus transcriptomics for CTC clinical translation 206
8.6 CTCs or circulating tumour DNA (ctDNA) in a liquid biopsy? 208
8.7 Challenges and lessons learnt 211
8.8 Conclusion 213

REFERENCES 215

APPENDIX

LIST OF FIGURES

1.1 Stage grouping for head and neck cancer TNM criteria 5

1.2 Mutant allele frequency (MAF) from bulk DNA sequencing makes estimation of clonal populations difficult 10

1.3 Multiple factors, both intrinsic and extrinsic, contribute to intra-tumoural heterogeneity (ITH) 15

1.4 Graphical representation of levels of ITH between different cancer types 17

1.5 Circulating tumour cells within the metastatic cascade 26

2.1 Systematic review study selection flowchart 57

2.2 No association between pre-treatment CTC immune-marker expression and progression-free survival (PFS) 63

2.3 No association between pre-treatment CTC immune-marker expression and overall survival (OS) 64

2.4 A mixed pattern of significance when post-treatment CTC expression of PD-L1 and IDO-1 is correlated to progression-free survival (PFS) 64

2.5 CTC expression of PD-L1 and IDO-1 post-treatment is associated with decreased overall survival (OS) 65

3.1 Thesis structure 72

4.1 Stimulation of PBMCs for 15 minutes produces optimum signalling pathway expression 80

4.2 Workflow for patient sample processing and mass cytometry antibody staining.. 80

5.1 The Parsortix CTC enrichment device 88

5.2 Spiked cell concentration does not impact capture rate 93

5.3 Parsortix is superior to CD45 MACS depletion to enrich spiked cells 94

5.4 Transfix BCTs demonstrate superior cell capture rates compared to EDTA and Cytodelics blood preservation 96

5.5 Differential gene expression analysis of FaDu cell line before and after Parsortix microfluidic enrichment demonstrates a cell stress response and downregulation of keratin proteins 99

5.6 Proof-of-principle that a rare cell population (epithelial CTC) can be identified from bulk sequencing of Transfix-preserved, Parsortix

enriched blood samples 105
6.1 Testing antibody compatibility with Transfix fixed HNSCC cell lines 109
6.2 Enriched HNSCC spiked cells are detectable but poorly visible within the enrichment cassette 111
6.3 Parsortix enriched healthy donor sample does not demonstrate EpCAM high cells 112
6.4 Parsortix enriched cells with characteristics of epithelial and EMT CTCs 115
6.5 High magnification images of CTCs reveals differing morphology and CTC clusters 116
6.6 Determining gating strategy to identify enriched epithelial cells 119
6.7 Gating strategy for identification of epithelial and mesenchymal CTCs 121
6.8 CTC count as cells per ml of whole blood and correlation to stage of disease 124
6.9 Survival outcomes correlated to tumour pathological variables 126
6.10 Correlation between CTC positivity and survival outcomes 128
6.11 Post treatment CTC levels decrease in patients responding to treatment 129
6.12 Tumour mesenchymal status does not correlate with blood CTC EMT profile 132
6.13 Mesenchymal expression in tumour and CTC results in significantly worse progression-free survival 134
7.1 Antibody optimisation experiments demonstrate range of binding efficacies on Transfix fixed cells 145
7.2 Assessing signal spillover for marker channels 149
7.3 Compensation bead antibody validation 152
7.4 Mass cytometry data clean-up strategy 156
7.5 Mass cytometry successfully identifies spiked HNSCC cell lines From Parsortix enriched whole blood 157
7.6 CTCs visualised on PCA plots 161
7.7 Antibody signal intensity is consistent between patient samples 163
7.8 CTCs cluster into sub-groups based upon epithelial-EMT status and functional phenotype marker expression patterns 166
7.9 Signal intensity heatmapping defines CTC sub-groups within patient samples 167
7.10 Relationship between epithelial-EMT status and phenotypic classification for clustered CTC sub-groups 171

7.11 Quantitative mRNA expression analysis identifies CTC specific transcripts in patient samples 178
7.12 Tumour samples cluster into mesenchymal and epithelial profiles 181
7.13 Mass spectrometry can identify proteins from Transfix fixed spiked cells enriched using Parsortix 184

LIST OF TABLES

1.1 Tumour (T) and lymph node (N) staging criteria 4
2.1 Characteristics of included studies 60
2.2 Extracted data from included studies 61
4.1 Antibody panel used for CTC detection 76
5.1 Downregulated differentially expressed genes in Parsortix enriched FaDu cells demonstrate enrichment of mRNA/RNA and ribosomal processing gene sets 100
5.2 The ThermoFisher PicoPure kit out performs the Qiagen RNeasy kit for RNA extraction from Transfix preserved samples 103
6.1 Immunofluorescence microscopy CTC characterisation test patients 113
6.2 Patient demographic and clinical data with CTC count 120
7.1 Reason and references for CTC marker inclusion in cytometry panel 139
7.2 Results of antibody clone optimisation experiments 145
7.3 Final mass cytometry panel 150
7.4 Demographic and clinicopathological data for the twelve patients in the mass cytometry cohort 159
7.5 Epithelial–EMT status and phenotype classification correlate together 172
7.6 Proportion of EMT, epithelial, proliferative and immune checkpoint high CTCs . 174

CHAPTER 1: INTRODUCTION

1.1 Head and neck squamous cell carcinoma

1.1.1 Incidence and aetiology

Head and neck cancer (HNC) is defined as a group of cancers that originate from the mucosal linings of the oral cavity, larynx, pharynx, paranasal sinuses and nasal cavity, also encompassing cancer of the salivary glands. Histologically, greater than 90% of these cancers are epithelial in origin and are termed head and neck squamous cell carcinomas (HNSCC). While salivary gland cancers and nasopharyngeal carcinoma are grouped together with HNSCC as HNC, they have different aetiology, pathology, clinical presentation and treatment pathways compared to HNSCC. Their incidence and survival outcome statistics are therefore mostly considered separately.

HNC is the 7th most common cancer worldwide, representing 650,000 cases and 330,000 deaths annually.[1] In the UK it is the 8th most common cancer with >12,000 cases diagnosed per year and a clear gender bias towards males – being 4th and 13th in cancer prevalence among males and females respectively.[2] More important is to understand the evolving incidence and pattern of this disease; over the past three decades HNC incidence rates have increased by a third in the UK. This increase is particularly marked in females compared to males, with a 43% increase during this time vs. 23% respectively.[2] Such results are partly attributed to a shifting aetiological profile. While the established risk factors of alcohol and tobacco consumption still predominate in the majority of HNSCC, a subset of cancers in the

oropharyngeal anatomical site have been shown to exhibit a viral aetiology, that of human papillomavirus (HPV). Contrary to HPV-negative(-ve) HNSCC, where the age distribution peaks between 65-75 years[2], HPV-positive(+ve) HNSCC affects a younger age group (<60 years age)[3]. While still more common in males, there is an increased female incidence of HPV+ve HNSCC, but a better prognosis than tobacco/alcohol driven disease.[4] Data from the USA has demonstrated that from 1984-2004 HPV+ve oropharyngeal carcinoma (OPC) cases increased by >200%. HPV+ OPC has recently overtaken cervical cancer as the commonest HPV driven cancer in the United States[5] and it is projected to be the most common cause of HNSCC by 2030.[3] The impact of this shifting aetiological pattern will greatly influence HNSCC treatment paradigms in the future.

1.1.2 Staging and treatment

As described, at the population level HNSCC rates are increasing. More important at an individual level is the stage at which these patients present to healthcare and the impact this has upon treatment and clinical outcomes. HNSCC is staged into groups I-IV, according to the American Joint Committee on Cancer TNM classification (T- tumour, N – nodal metastasis, M – distant metastasis) (table 1.1).[6] While specific staging criteria with regard spread of local disease in T4 cases are defined based upon the tumour site, the grading of tumour based upon size is consistent throughout (apart from glottic or paranasal sinus tumours). Together, the T and N grading form the stage grouping (stage I-IV, figure 1.1). Stage I/II are 'early' cancers with a small primary tumour and no evidence of metastasis, and stage III/IV are 'advanced' cancers with a larger primary mass and/or nodal/distant metastasis.

Unfortunately, the majority of patients presenting with HNSCC do so with advanced stage disease and nodal metastasis. Despite treatment, up to 60% can be expected to develop recurrent disease and 30% distant metastasis.[7] These late presentations contribute to poor overall 5-year survival in HNSCC of between 50-60% - dependent upon anatomical site, gender and HPV-status.[2] Despite advances in treatment, this figure has remained largely unchanged over the past few decades. HNSCC is also a complex and particularly devastating disease with significant functional morbidity, such as appearance, speech and swallowing from both the disease itself, but also the side effects of treatment. The resulting burden of care and impact upon health services is among the highest for any cancer type - requiring a large multi-disciplinary team.[8] Current curative intent regimens for early stage HNSCC usually involve a single treatment modality, for example surgical resection or primary radiotherapy. Advanced stage disease will often require a combined-modality approach of surgery, radiotherapy +/- chemotherapy depending on tumour site and stage.

Aside from efforts to improve techniques and pathways for earlier diagnosis, the key question facing head and neck oncology is that of delivering on the promise of precision medicine – *this being accurate treatment selection at the individual patient level*. Modern oncology is shifting from empirical treatment strategies to personalised treatment models based on biomarkers derived from the molecular profile of the tumour. To-date, the only prognostic biomarker in HNSCC is the aforementioned HPV-status of the tumour (commonly assessed by the surrogate marker of p16 overexpression via immunohistochemistry). Targeted biologicals and immunotherapy agents show significant promise in recurrent/metastatic (R/M) HNSCC patient groups, for example EFGR inhibition[9] or blockade of the PD-1 checkpoint.[10]

Table 1.1: Tumour (T) and lymph node (N) staging criteria. Table shows the description and criteria for grading T and N in head and neck cancer from the 7th edition of AJCC TNM staging. Tumour site specific criteria for T4a and T4b stages are not listed.

Primary tumour (T)		Regional lymph nodes (N)	
TX	Primary tumour cannot be assessed	NX	Regional lymph nodes cannot be assessed
T0	No evidence of primary tumour	N0	No regional node metastasis
Tis	Carcinoma in-situ	N1	Metastasis in a single ipsilateral lymph node, 3 cm or less in greatest dimension
T1	Tumour 2cm or less in greatest dimension	N2	Metastasis in a single ipsilateral lymph node, more than 3 cm but not more than 6 cm in greatest dimension; or in multiple ipsilateral lymph nodes, none more that 6 cm in greatest dimension; or in bilateral or contralateral lymph nodes, none greater than 6 cm in greatest dimension
T2	Tumour more than 2cm but not greater than 4cm in greatest dimension	N2a	Metastasis in a single ipsilateral lymph node, more than 3 cm but not more than 6 cm in greatest dimension
T3	Tumour more than 4cm in greatest dimension	N2b	Metastasis in multiple ipsilateral lymph nodes, none more that 6 cm in greatest dimension
T4a	Moderately advanced local disease (site specific criteria apply)	N2c	Metastasis in bilateral or contralateral lymph nodes, none more than 6 cm in greatest dimension
T4b	Very advanced local disease (site specific criteria apply)	N3	Metastasis in a lymph node more than 6 cm in greatest dimension.

Stage Grouping			
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage II	T2	N0	M0
Stage III	T3	N0	M0
	T1	N1	M0
	T2	N1	M0
	T3	N1	M0
Stage IVA	T4a	N0	M0
	T4a	N1	M0
	T1	N2	M0
	T2	N2	M0
	T3	N2	M0
	T4a	N2	M0
Stage IVB	Any T	N3	M
	T4b	Any N	M0
Stage IVC	Any T	Any N	M1

Clinical Stage Grouping by T and N Status					
N	**T1**	**T2**	**T3**	**T4a**	**T4b**
N0	I	II	III	IVa	IVb
N1	III	III	III	IVa	IVb
N2	IVa	IVa	IVa	IVa	IVb
N3	IVb	IVb	IVb	IVb	IVb

Figure 1.1: Stage grouping for head and neck cancer TNM criteria (adapted from AJCC 7th edition).

However, if targeted therapy, such as the anti-PD-1 agent nivolumab, produces a response rate of up to 30% in the R/M setting - being able to identify the majority of patients that will not benefit from such treatment would reduce the significant morbidity associated with drug toxicity while reducing the cost burden of these expensive treatments upon healthcare services. Unfortunately, robust predictive biomarkers to match patients to the appropriate treatment are lacking or have poor performance. As will be discussed in this thesis, one solution is the concept of a '*liquid biopsy*'- a minimally-invasive blood test allowing assessment of both tumour burden and the presence of druggable targets to derive biomarkers for treatment stratification.

1.2 The cancer genome

1.2.1 DNA mutations

It is long-established that cancer is a genetic disease and that genome instability is an underlying principle driving the biological features defining the 'hallmarks of cancer', as described by Hanahan and Weinberg.[11] Advances in next generation and high throughout sequencing technologies have heralded a new era in our understanding of cancer genomics. The ability to interrogate the cancer genome has provided a previously unseen picture of the complex heterogeneity of DNA abnormalities driving tumour growth and 'clonality', and their resultant impact upon treatment outcomes.

The starting point for cancer initiation is mutation of the DNA sequence. These may be germline (hereditary) or somatic (occurring after conception in somatic cells) mutations. Mutation of DNA can be caused by a variety of different mutagens – either external, such as tobacco smoke in HNSCC,[12] or internal mutagens generated in the intracellular environment, such as reactive oxygen species.[13] Alternatively a significant number of mutations occur spontaneously as errors in DNA replication and repair.[13] Broadly speaking, mutations can be classified as being small/simple mutations - involving single nucleotides, or large scale - affecting loss or gain of regions or entire chromosomes, such as copy number alterations (CNA). Single nucleotide polymorphisms (SNPs) include base substitutions, insertions or deletions and are responsible for a large portion of oncogenic mutations.[13] Synonymous mutations are 'silent' and do not cause a change in the transcribed codon, whereas a nonsynonymous mutation in the exome 'coding' region of DNA will cause an

altered gene product.[13] The resulting effect is either a loss or gain in function of the mutated gene. CNA (a somatic event, differentiated from germline *copy number variation*) is a type of structural chromosomal alteration whereby sections of the genome are duplicated or deleted - either small enough to be focal to a particular gene or broader/chromosome-arm level alterations.[13] The resulting deletion or amplification is correlated to transcriptional activity and differential expression of target genes, thought to occur via a 'gene dosage' effect.[14] CNA is shown to be strongly associated with several cancer types and implicated in oncogenesis.[15] Furthermore, focal CNAs in known oncogenic driver genes have been demonstrated as strong prognostic biomarkers, greater than equivalent gene SNPs.[16]

Research into the cancer genome has now identified that, on average, a tumour possess >10,000 somatic mutations, but mutations in a relatively small proportion of 'driver' genes (≈2-8 in each tumour) are responsible for the selective growth advantage that initiates tumour growth and progression.[17] Such driver genes can be common among a large proportion of different cancers, for example *TP53* mutation, or may be highly specific to individual cancers.[18] The cellular processes that control growth and replication leading to tumour formation are governed by 2 classifications of genes – oncogenes and tumour suppressor genes (TSG). Oncogenes, termed proto-oncogenes in a nonmutated format, are genes which promote cell proliferation. Therefore, a gain of function mutation in an oncogene causes uncontrolled cell growth. Conversely, TSGs and their products control inappropriate cell cycle progression. A loss of function mutation in a TSG leads to unchecked cell growth.

Ordinarily, cellular mechanisms are in place that defend against single mutations causing a malignant tumour. It is estimated that at least six mutations need to occur to cause malignant

progression from a normal cell.[13] However, the probability of further mutations may be increased by the impact of external mutagens and successive mutations. Oncogenic mutations that promote cell proliferation increase the likelihood of further mutations in a larger target cell population and mutations which affect genome stability will increase the overall mutation rate. For these reasons, tumourigenesis occurs on a pathway from normal cells, through dysplasia, to invasive malignancy (as seen in HNSCC). The accumulation of mutations that drive tumour growth creates groups of cells with a common mutational ancestry – termed 'clones'.[19] Mutations that promote this clonal expansion are termed 'driver mutations' - defined as *a mutation associated with tumour initiation, progression, metastasis or resistance to therapy.*[20] However, confirming specific mutations as 'drivers' in an individual tumour is complex (outside of in-vitro experimental modelling)[20]. To define a cancer driver mutation, it must be observed in multiple neoplasms, above the background mutation rate and be associated with clonal expansion, often in a role of a cellular process associated with oncogenesis.[19] In contrast, a 'passenger mutation' is a mutation that has no effect upon function or phenotype.[19] As tumour growth progresses these clones themselves undergo further mutation and expansion, a process known as clonal evolution. A useful analogy is that of a 'clonal evolutionary tree' – the dominant clone within the tumour makes up the trunk, while 'branching' sub-clones are observed, which in turn over time may become dominant trunks themselves during tumour growth.[21]

While the above is a theoretical model, in practice the identification and separation of clinically relevant clones and sub-clones, with their associated mutations and functional impact, is far more difficult. From these definitions of clonality and the underlying driver mutations, it can be seen that the method of tumour sampling and strategy for data analysis

can have a strong bearing on the conclusions and outcome. It is worth highlighting the role that computational methods can have on the conclusions derived from sequencing experiments. As the majority of studies are derived from sequencing of one tumour site only (DNA is extracted en-masse and bulk sequenced) inferring a mutation as being clonal or sub-clonal is complex and dependent on the chosen bioinformatic algorithm (which themselves are being constantly refined). In addition, this means of assessment does not consider the heterogeneity of clones within different parts of a tumour (efforts of multi-region tumour sequencing and advances in single-cell molecular profiling to more accurately assess genomic heterogeneity are discussed below in section 1.4). Put simply, assessment of clonality is made on the variant allele frequency (percentage of DNA reads of a variant compared to overall coverage at that locus) of a mutant gene.[22] After adjustment for tumour purity and ploidy those mutant variants with a higher frequency are classified as clonal. This data can then be phylogenetically reconstructed to estimate the longitudinal relationship of sub-clones to clones and thus the evolution of genomic heterogeneity within a tumour sample.[23]

However, the inferred clinical significance of mutant variants labelled as clonal driver events can only ever be as strong as the quality of the data. For example, a variant may not be a critical driver of tumour growth despite high frequency or appear falsely dominant due to sampling bias of a tumour. Subsequent therapeutic targeting may have a detrimental effect upon outcome due to 'competitive release' of more aggressive sub-clones or spatially heterogenous clones not sampled[19] - contributing to relapse and drug resistance. Figure 1.2 presents a theoretical experiment that demonstrates differing permutations of clonality from the same results of mutant variant allele frequency in bulk DNA sequencing. It is clear to see that the interpretation of this data will significantly impact the assessment of clonality

and the evolution of metastatic clone status. Therefore, as clinical oncology catches up with the rapid advances in sequencing technology, the accurate assessment of clonal status of driver mutations is a pressing issue in cancer genomic research. Single-cell sequencing presents a potential solution to this problem. Being able to perform serial genomic monitoring of tumours to accurately identify clonal drivers and the transition from sub-clone to clone for example those contributing to R/M, is clinically highly relevant. This concept will be explored in greater depth below when discussing advantages of liquid biopsies.

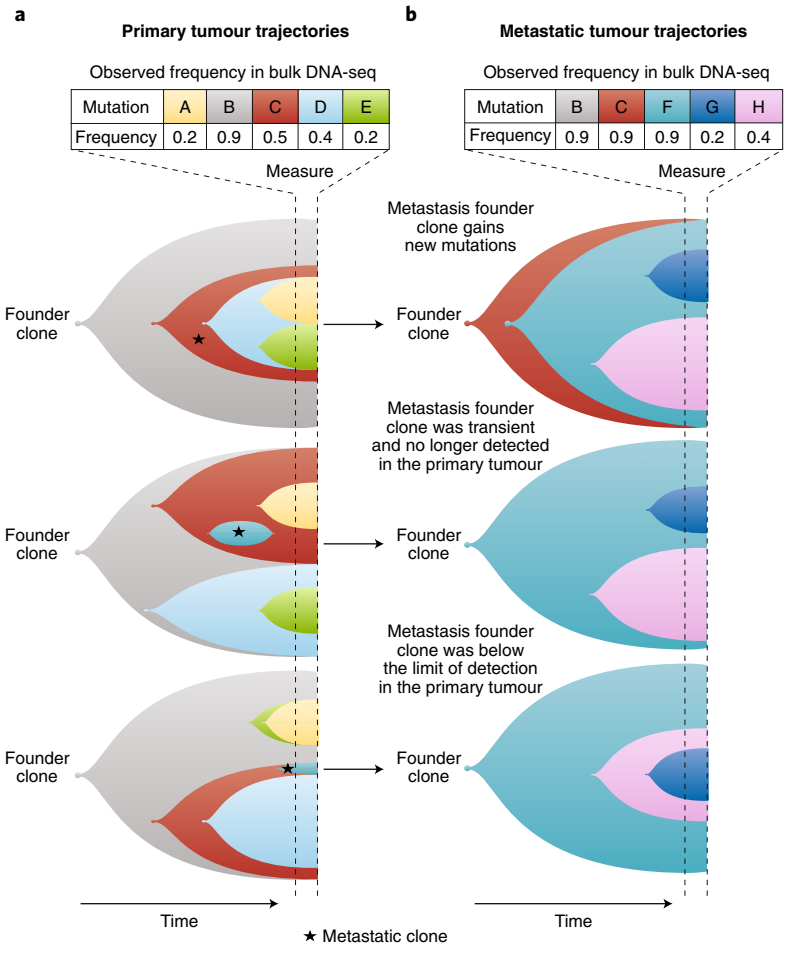


Figure 1.2: Mutant allele frequency (MAF) from bulk DNA sequencing makes estimation of clonal populations difficult. The above figure shows a theoretical model to demonstrate the potential inaccuracies of using mutant allele frequency for clonal assessment with bulk DNA sequencing datasets. The tables above show MAF's for 5 mutations at the primary tumour timepoint (a) and metastatic tumour timepoint (b). The 3 diagrams below each

timepoint are possible inferred clonal relationships from the MAF datasets above, specifically which mutations co-exist within the same cell group clones or sub-clones. This demonstrates how bulk DNA sequencing data is still prone to subjective analysis and single cell genomic (or proteomic) provides a more accurate estimation of clonality and tumour molecular evolution (adapted from Lawson *et al*.[17]).

Finally, it should be noted that stochastic clonal expansion is only one theory behind the genomic heterogeneity seen in cancer. The cancer stem cell model describes how malignant cells with stemness properties sit atop a hierarchy of self-renewing cells that sustain tumour growth and produce heterogeneity through differentiation.[24] This model has generated significant interest as it is thought these cancer stem cells may be responsible for recurrence after primary treatment failure. In reality these models are unlikely to be mutually exclusive, representing part of the complex ecosystem observed in carcinogenesis.[17] In addition, the tumour microenvironment, both stromal cells and infiltrating immune cells, has a complex and significant impact upon sustaining tumour growth and in turn driving genomic heterogeneity, in addition to epigenetic alterations.[25], [26]

1.2.2 The genomic landscape of HNSCC

Early small cohort sequencing experiments observed a highly variable and heterogenous mutational landscape in HNSCC.[27], [28] However, the work by The Cancer Genome Atlas (TCGA) Network, profiling 279 HNSCC tumours, provided the first comprehensive overview of the somatic genomic alterations driving oncogenesis in this cancer group.[29] Nineteen genes were found to be mutated at a frequency of >1% in this TCGA cohort, of which TSG mutations

were significantly more prevalent than oncogene mutations. Mutations in *TP53* (TSG involved in cell cycle regulation) were by far most common, present in 73% of all tumours. This proportion was even higher when only including HPV-ve tumours (86%). The next five most commonly mutated genes, in the range of 18-23% of cases, were *FAT1, CDKN2A, PIK3CA, NOTCH1* and *KMT2D* in order of mutation prevalence. A notable finding was that while HPV+ve tumours had similar overall mutational burden (mutations per mega-base of DNA) compared to HPV-ve tumours, they had a very different mutational profile – a finding corroborated in other studies.[30] In the HPV+ve group in TCGA data, *PIK3CA* mutation was seen in the majority of tumours (56%). In a separate cohort of HPV+ve tumours, in addition to *PIK3CA* mutations, Seiwert *et al*. demonstrated increased frequency of *KRAS, MLL2/3* and *NOTCH1* mutations. Thus, highlighting the differing mutational profile between these disease entities.

TCGA also presented CNA data from the HNSCC cohort. Numerous patterns of amplification and deletions were present within tumour sub sites and HPV status. Clustering analysis revealed a group of predominantly oral cancers with decreased CNAs, indicating a mutation rather than copy number tumourigenesis pathway for these cancers. In addition, HPV-ve tumours clustered around chromosome 7 amplification (containing the *EGFR* gene). Both HPV -ve and +ve cancers demonstrated frequent amplifications of *TP63* and *SOX2* (squamous lineage transcription factors) and also the oncogene *PIK3CA*. HPV-ve cancers demonstrated receptor tyrosine kinase amplifications (*EDGR, ERBB2* and *FGFR1)* and frequent deletions of TSGs (*FAT1, NOTCH1, SMAD4* and *CDKN2A).* While HPV+ve cancers exhibited *TRAF3* (involved in anti-viral immunity response) recurrent deletions and amplifications of *E2F1* and *CDKN2A* (p16).

Heterogeneity at the transcriptional level is also evident within HNSCC. Chung *et al*. profiled gene-expression of 60 HNSCC tumours and demonstrated that tumours clustered into 4 groups.[31] They noted dominant pathways within each tumour group, labelling them as EGFR-pathway enriched, mesenchymal-enriched, normal epithelium-like, and high levels of antioxidant enzymes subtypes for groups 1-4 respectively. Subsequently, Walter *et al*. validated these findings in a group of 140 HNSCC tumours to define the sub-groups as basal, mesenchymal, atypical and classical (groups 1-4 respectively) based upon functional gene-expression.[32] When considering HPV-ve cancers only, the atypical subtype had significantly worse recurrence-free survival (RFS) compared to the other 3 subtypes (p=0.01). TCGA confirmed these 4 gene-expression subtypes, noting significant associations between specific candidate gene expression and subtype groups.[29] For example, *CASP8* in the basal subtype (p<0.01), *NOTCH1* in the mesenchymal and basal subtypes (p<0.01) and *TRAF3* and *E2F1* in the atypical subtype (p=0.02 and <0.01 respectively). In addition, they demonstrated how somatic alterations correlated with expression subtypes, such as *TP53* mutations, *CDKN2A* loss of function and oxidative stress gene alterations in the classical subtype.

The above data serves to illustrate the considerable genomic heterogeneity within HNSCC. With numerous potential driver mutations and gene amplifications/deletions contributing to multiple pathways of oncogenesis, the task of providing personalised medicine that targets an individual tumours oncogenic driver(s) is complex. However, the correlation of tumour molecular subtype to treatment outcomes highlights the potential for tumour sequencing to provide prognostic and or predictive biomarkers. From a clinical standpoint of treatment stratification, it is important to highlight *both* the considerable inter-tumour *and* intra-tumoural heterogeneity (ITH) of HNSCC. Not only do patients have differing oncogenic

drivers, but these are heterogenous within the individual tumour itself. We will continue to build on this concept further – discussing ITH in HNSCC and methods to assess such genomic heterogeneity below. A final point to make is that, as highlighted above, loss of function of TSGs is more common than oncogene activation in HNSCC, but this is inherently more difficult to target. Therefore, targeting the resultant cell phenotype i.e. proteomic characterisation of (circulating) tumour cells, may hold increased clinical relevance compared to genomic assessment.

1.3 Intra-tumoural heterogeneity

1.3.1 Spatial and temporal molecular heterogeneity

Tumours are heterogenous groups of phenotypically diverse cells - encompassing malignant cells and non-malignant stromal cells and infiltrating immune cells. The interaction between these tumour cells, and also intrinsic and extrinsic factors, (figure 1.3) creates an ecosystem of pro- and anti-tumour activity that generates considerable intra-tumoural heterogeneity.[17] As highlighted, genomic instability and clonal expansion is a principal contributing factor. Henceforth we focus our discussion of ITH on the genetic profile of malignant cells.

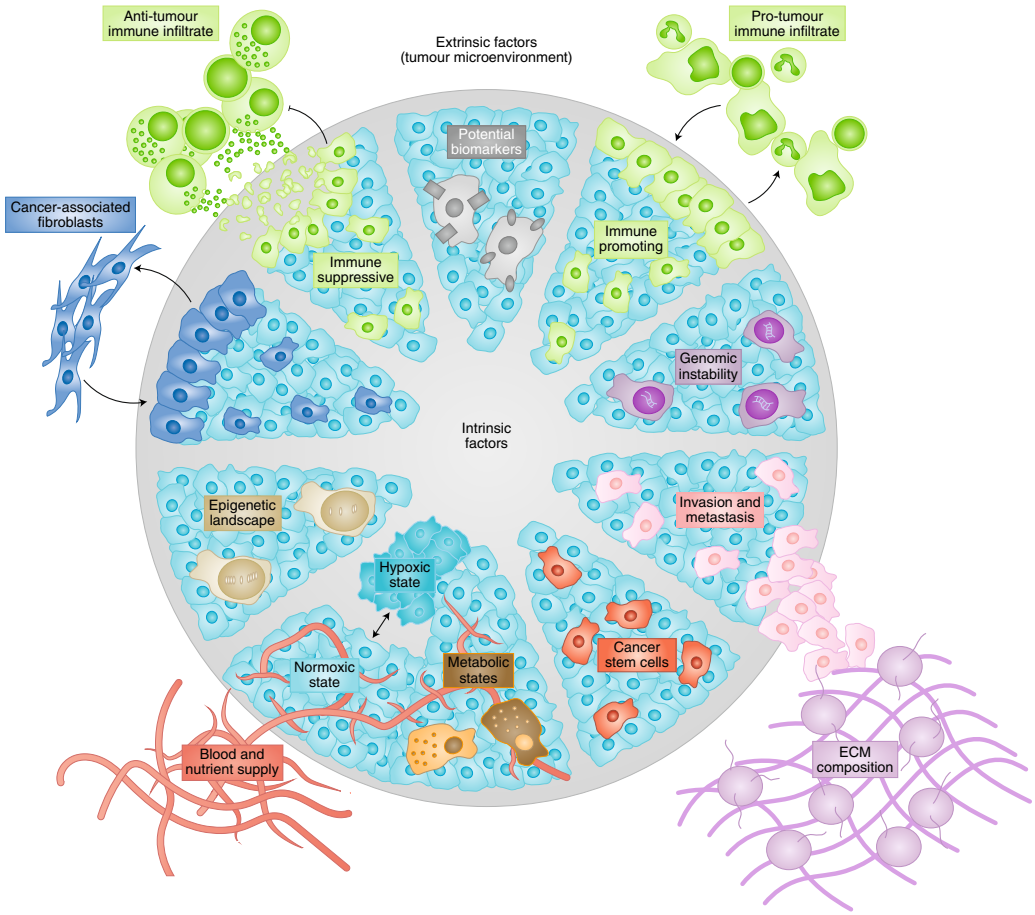


Figure 1.3: Multiple factors, both intrinsic and extrinsic, contribute to intra-tumoural heterogeneity (ITH). The above diagram provides a pictorial representation of the molecular factors and cellular compartments within a tumour that make-up and contribute to ITH. Of note is the relationship between extrinsic factors, such as immune cells, and the intrinsic tumour cell populations. While an oversimplification, it serves to highlight the multi-omic layers of complexity of ITH. While therapy is often targeted to one compartment, other omic factors must be considered (adapted from Lawson *et al.*[17])

ITH can be viewed across a spectrum of multi-omic datasets – considering variation of underlying mutation, methylation or transcriptional profiles and resulting protein marker expression. It is important to view ITH as a shifting continuum, appreciating both static and dynamic changes over time. Static sampling of tumour tissue, as is the current standard for genomic studies, provides a snapshot only. As discussed, this may be further compounded by

sampling bias – which brings into frame the concept of *spatial heterogeneity*. Put simply, different regions of a tumour may contain different clones that are inferring a particular advantage to the tumour in that location, or will go on to become dominant clones driving treatment resistance or recurrence.[33] Thus, single site tumour sampling, as is standard clinical practice to achieve a histological diagnosis, has a high potential to miss important driver mutations. Gerlinger *et al*. provided some of the most robust evidence of spatial ITH by performing multi-region sequencing on a clear-cell renal carcinoma.[33] In this study they sampled 9 sites from the primary tumour, together with 3 sites from 2 metastatic deposits. Among these sites, 128 mutations were found, of which 40 were ubiquitous across all regions, 59 mutations were shared by some but not all regions, and 29 were *unique* to a specific region. Of note, 31 mutations were seen to be shared by most of the primary tumour regions and 28 by most of the metastatic regions, indicating clonal expansion during the metastatic process. In the same study, to demonstrate the functional impact of observed ITH, interrogation of the mTOR pathway (a regulator of cell growth, proliferation and motility) revealed that primary tumour regions with mTOR mutations demonstrated increased phospho-antibody staining of downstream mTOR pathway targets, compared to those with wild type mTOR gene. For clinical context they were able to determine that a single biopsy site would only capture approximately 55% of all mutations in the tumour. In addition, when a known prognostic gene signature of clear-cell renal carcinoma (type A and B tumours with good and poor prognosis respectively) was applied to the primary tumour and metastasis, it resulted in metastatic regions and one primary tumour region being classified as type A 'good' prognosis and the remaining tumour regions classified as type B 'poor' prognosis. Thus, demonstrating how genomic based prognostic stratifiers may be flawed by tumour sampling bias.

In a recent review, McGranahan and Swanton collated multi-region sequencing data from 20 studies across 15 cancer types (not including HNSCC) (figure 1.4).[21] It was clear to see that the majority of cancers display considerable ITH, although this was variable between cancer types. Within this dataset was the observation that a large clonal burden was not directly correlated with a large sub-clonal burden, and vice versa. In addition, certain gene mutations appear more often as sub-clonal, while others are predominantly clonal – for example TP53 mutations, as seen in HNSCC.[29] The above evidence clearly highlights the impact that ITH can have upon tumour genetic assessment in a clinical setting and more importantly the methods we use to assess ITH. If treatment decisions are made on the basis of clonality and driver mutation status we must be sure we are not missing pieces of the puzzle through inadequate sampling or analysis.

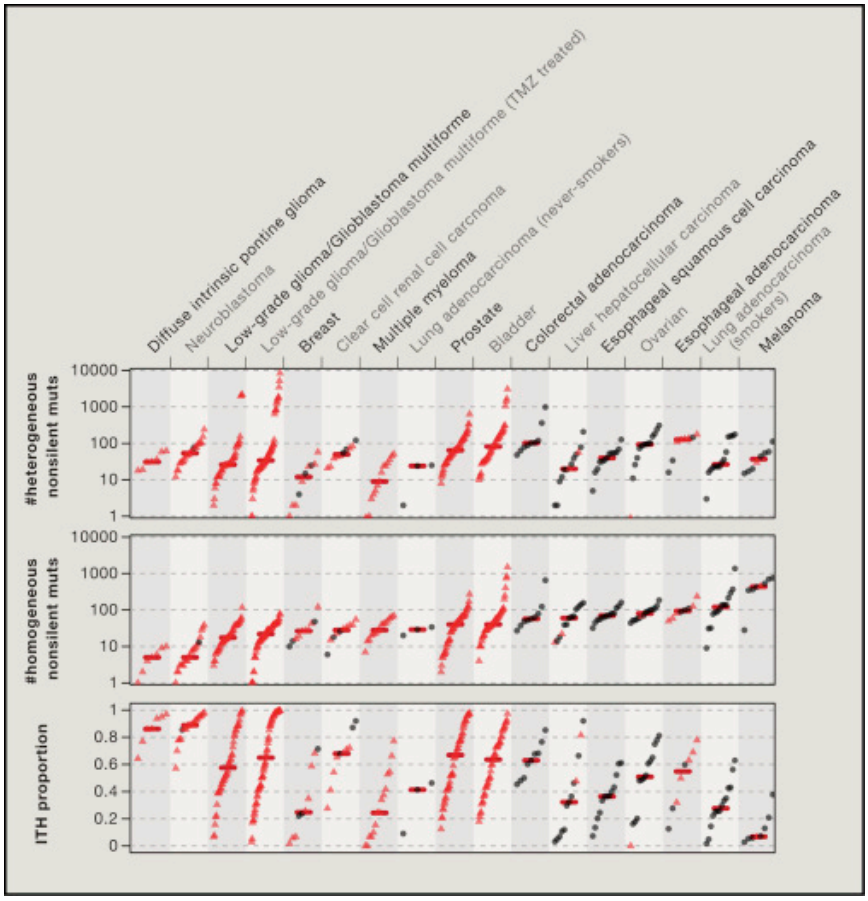


Figure 1.4: Graphical representation of differing levels of ITH between different cancer types. Each point represents a single tumour (black circles are treatment naïve and red triangles are post treatment). Notable how certain cancers, for example lung cancer and melanoma, have increased mutational burden but decreased ITH when compared to other cancers (adapted from McGranahan and Swanton *et al*.[21])

Temporal heterogeneity represents changes in the mutational landscape of the cancer over time, either within the primary tumour itself or in metastatic deposits occurring during disease progression. As discussed, tumourigenesis occurs on a pathway of mutagenesis and cellular dysplasia resulting in carcinoma, however the clonal mutational burden at each timepoint is not equal between all cancers. For example, lung and melanoma exhibit a high number of mutations acquired before tumour formation; in contrast to other cancer types, including HNSCC,[18] which have a high number of clonal somatic events occurring after tumourigenesis.[21] Such variance in temporal clonality between different cancer types has implications for the way in which tumour genetic testing is translated to the clinical setting. Hence, serial sampling may present a clearer picture of clonal expansion and heterogeneity for particular cancer types, with subsequent targeting of treatment based upon this patient-specific information.

1.3.2 Single-cell analysis to identify ITH

Most studies discussed so far have used bulk DNA or RNA sequencing. The disadvantage of this approach is a lack of clarity as to the cellular origin of a mutant variant, making it impossible to identify which cells (clones) have the same or *multiple* mutations and how this

functionally impacts the expression of genes and protein markers. In addition, the noise from bulk-sequencing may mask rare subclones that are clinically significant.[34] Evidence demonstrates that metastasis is often driven by expansion of a subclone within the primary tumour,[33], [35] therefore being able to accurately identify and map subclone evolution is crucial to directing therapy in this R/M group.

The advent of single-cell sequencing technologies has afforded researchers a previously unseen view of ITH. Since the first single-cell genome sequencing experiment of two breast cancers in 2011, demonstrating multiple clones with heterogenous CNAs,[36] multiple studies have achieved both single-cell exome and transcriptome sequencing in various cancers,[37] including HNSCC.[35] These studies have been able to demonstrate the genetic diversity and considerable ITH within tumours. Single-cell sequencing removes the contamination from stromal/immune cells, as malignant cells can be pre-sorted or identified from sequencing datasets. Therefore, rare malignant cell types can be identified and variants within cell clones more accurately quantified. Despite these advantages, single-cell analysis and identifying spatial ITH within tumours is somewhat contradictory. While technologies may differ, current protocols require a single cell suspension generated via tumour dissociation and then isolation of individual cells for DNA/RNA extraction, library preparation, barcoding and sequencing. Thus, the dissociation of tumour tissue removes the spatial context of malignant cells within the tumour and their relationship to other malignant or non-malignant cell groups, such as immune cells.

Technologies such as digital spatial profiling (DSP) or imaging mass cytometry are able to provide multi-parameter characterisation of cells in-situ within tumour sections, thus

retaining the spatial context of ITH. DSP allows whole transcriptome profiling to be performed for a region of interest on a tumour tissue section.[38] Gene or protein expression level data can be viewed at the near single-cell level, which can then be correlated to immunoflourescent images of the tissue to provide spatial resolution. Imaging mass cytometry uses a similar concept to (DSP), staining a tissue section with heavy metal labelled antibodies. The slide is then ablated in $1\mu m^2$ pixels to generate an image with layers for each marker that can be overlaid digitally. Therefore, multiple cell phenotypes with target marker expression profiles can be detected while being able to analyse the location within the tumour.[39] Mass cytometry is discussed further in chapter 7.

1.3.3 ITH in HNSCC and its clinical impact

As highlighted, HNSCC demonstrates considerable *inter-tumoural* and *intra-tumoural* molecular heterogeneity.[21],[36],[37] With regard to genomic ITH, few studies have examined spatial genomic heterogeneity in HNSCC. We focus on the proteo-genomic ITH of malignant cells within the tumour, while acknowledging that immune[42] and stromal[43] cell ITH can impact both treatment outcome and the clonal expansion contributing to genomic ITH.[26]

Zhang *et al*. performed whole-genome sequencing of 3 tumour sites from an oropharyngeal carcinoma (HPV-positive) in addition to 2 separate regions from one metastatic lymph node.[41] As expected, somatic point mutations (SPMs) were both homogenous and heterogenous between all 5 sites. The three tumour sites shared only 51.7% of genome wide

SPMs, and each tumour site contained exonic gene SPMs exclusive to that site (50, 57 and 149 gene SPMs respectively). Known HNSCC driver genes were exclusive to each tumour site and a set of genes were significantly mutated in the primary tumour samples only with a different set mutated in the lymph node samples. Phylogenetic reconstruction demonstrated a tumour clonal evolutionary pathway in relation to time. In addition, tumour samples spatially closer to each other were more homogenous, and lymph node metastasis samples clustered together likely forming from the same parental clone. While the results of this study are not surprising, they do highlight the considerable spatial genomic heterogeneity seen in HNSCC, that echoes other cancers as previously discussed.[20],[46] As will be discussed below, a liquid biopsy is one potential method to address this problem.

Correlating the results of multi-region sequencing to clinical outcomes in a large cohort is very difficult, given the considerable cost and resources needed to perform this testing. While it is fascinating from a mechanistic point of view to profile cancer progression, the clinical translation of this data requires a different approach. One solution has been to define a metric to estimate ITH from single site tissue biopsies, as described by Mroz *et al*. This group developed the '*MATH*' *(mutant allele tumour heterogeneity)* score to quantify ITH in HNSCC.[44] In brief, this method looks at mutant allele fractions of tumour-specific mutations and calculates the width of the distribution of these in relationship to a normal reference. Importantly, this definition is different to the mutation rate/load estimation as seen in the *tumour mutational burden* (TMB) score, as discussed below. Following this, the same group validated this score in a cohort of 74 patients[45] and then a larger cohort of 305 patients from TCGA database.[40] The MATH score was an independent prognostic biomarker of poor overall survival (OS) in all patients and was able to predict treatment response in patients

receiving chemoradiotherapy, in addition to distinguishing oral cavity and laryngeal cancers into 2 prognostic groups. The MATH score has subsequently been validated as a prognostic biomarker in other cancer types, such as colorectal cancer.[46]

High TMB (defined as the number of non-synonymous exonic somatic mutations per megabase of DNA) has been demonstrated as a marker of response to immune checkpoint inhibitor (ICI) therapy across a number of cancers,[47] including HNSCC.[48] The hypothesis behind this correlation is based on the assumption that an elevated TMB will result in increased tumour expression of immunogenic neoantigens and thus improved T-cell targeting.[49] While TMB is not a measure of ITH, we discuss it here because it is one of very few genomic biomarkers that it currently approved for clinical use and thus translation to a liquid biopsy platform is a logical progression.[50], [51] In addition, there appears to be an interesting negative correlation between TMB and ITH as predictive biomarkers. While a high neoantigen load from high TMB predicts response to ICI, increased clonal neoantigen ITH is seen in ICI non-responders.[52] Such findings are thought to be linked to mechanisms of immune evasion seen in tumours with high ITH.[53] This concept is important, because if ICI therapy and the immune-cell 'non-genetic' component of ITH is driving clonal selection,[26] then we need a dynamic and repeatable method to assess such evolving temporal genomic ITH, as discussed below using a liquid biopsy. True calculation of TMB from whole-exome sequencing is both costly and complex, however the discovery that using targeted gene panels can produce comparable results has led to the development of numerous commercially available TMB assays.[54] For example, the FDA authorised MSK-IMPACT TMB assay (identifying somatic mutations in 468 cancer-specific genes) was recently demonstrated to be a strong predictor of ICI response in a cohort of 138 HNSCC patients.[55]

Single-cell sequencing has also highlighted considerable transcriptional ITH in HNSCC. Puram *et al*. used single-cell RNA sequencing to analyse ITH in 18 oral cavity SCCs and 5 matching LN metastases.[35] When they examined gene expression of cells from individual tumours they were able to define clusters of cells with distinct gene signatures (up to 6 separate clusters in one patient, for example epithelial differentiation, cell cycling, stress response). What was interesting is that these signatures were *similar* between patients, so the pathways of ITH were common between patients. Such data is important if designing targeted panels to stratify patients based on ITH patterns. A key finding from this paper was that a sub-group of malignant cells (located at the advancing margin of the tumour as seen in tissue sections) exhibited a partial EMT. These cells had downregulated but preserved epithelial expression, with expression of some EMT markers such as Vimentin and SNAIL2 but *not* other classical EMT markers, thus demonstrating a partial EMT which was common among all the tumours. As previously discussed, EMT in carcinomas is linked to invasion, metastasis and drug resistance. When we discuss circulating tumour cells below, we will highlight how CTCs in HNSCC exhibit both a 'partial' EMT and 'full' EMT.

1.4 Circulating tumour cells (CTCs) as a liquid biopsy

1.4.1 What is a liquid biopsy?

The term 'liquid biopsy' refers to the identification of tumour-specific material in any bodily fluid – including blood, saliva and urine. The majority of research has focused on plasma (blood) derived material. Blood compartments include circulating tumour DNA (ctDNA), tumour cells (CTCs), exosomes, vesicles, and microRNA.

CTCs were discovered as early as 1869[56] and levels of cell-free circulating DNA observed in higher quantities in cancer patients in 1977.[57] However, it was the confirmation that cell-free circulating DNA harboured tumour specific genomic alterations, by Stroun *et al*. in 1989, that was arguably the birth of liquid biopsy research and the realisation of the clinical potential this sampling technique could have.[58] Since this early research, the benefit of a liquid biopsy to assess tumour burden and driver mutations has been demonstrated, using both ctDNA[59]–[61] and CTCs[62][63] – with significant progress in head and neck cancer.[42][43]

A liquid biopsy is proposed to have several advantages over the current standard of care. Firstly, a blood test is a minimally invasive procedure that can be performed in an outpatient or even community setting, with less morbidity than a solid tissue biopsy which requires a surgical procedure often under a general anaesthetic. In addition, once analysis protocols are standardised and sequencing costs are further reduced it has the potential for significant cost savings. However, most notable is the ability to perform serial tests at multiple treatment

time-points - thus being able to closely monitor treatment response, detect residual disease and catch recurrence early.[66] Obtaining tumour-specific biomarkers from a liquid biopsy is particularly advantageous in the R/M setting, when a tissue biopsy is often unobtainable or high risk. Currently R/M treatment regimens are often empirical and not targeted to molecular drivers specific to the individual's tumour. Therefore, the ability of liquid biopsies to provide therapeutic biomarkers to direct targeted/immunotherapies, which are increasingly becoming the mainstay of treatment in R/M patients, is particularly promising. Such examples include *EGFR* mutation assessment of ctDNA to direct EGFR-inhibitor treatment,[67] or expression of PD-L1 on CTCs as a biomarker for checkpoint immunotherapy.[68]

Over the past decade the question being asked in liquid biopsy research has shifted, from proof-of-principle that a liquid biopsy (ctDNA or CTC) can demonstrate tissue matched tumour-specific genomic alterations, to 'can a liquid biopsy tell us *more* than a single tissue biopsy?' i.e. can a liquid biopsy illuminate both spatial and evolving temporal ITH. As discussed, a single tissue biopsy is inadequate to identify the clonal landscape within a tumour[33] and often cannot be repeated over multiple time-points. Analysis of both ctDNA and CTCs has provided strong evidence that a liquid biopsy may be superior to a tissue biopsy in assessing ITH.[69] Henceforth, we focus our discussion on recent advances in CTC research, as is the topic of this thesis.

1.4.2 CTCs - their role in metastasis and the importance of an epithelial-mesenchymal transition

Metastasis is best viewed as a multi-stage pathway (figure 1.5). Invasion and intravasation of tumour cells at the primary site constitutes the first stage, the second stage involves survival of CTCs in the circulation and dissemination to distant sites, and the third stage involves extravasation, tumour initiation and metastatic growth.

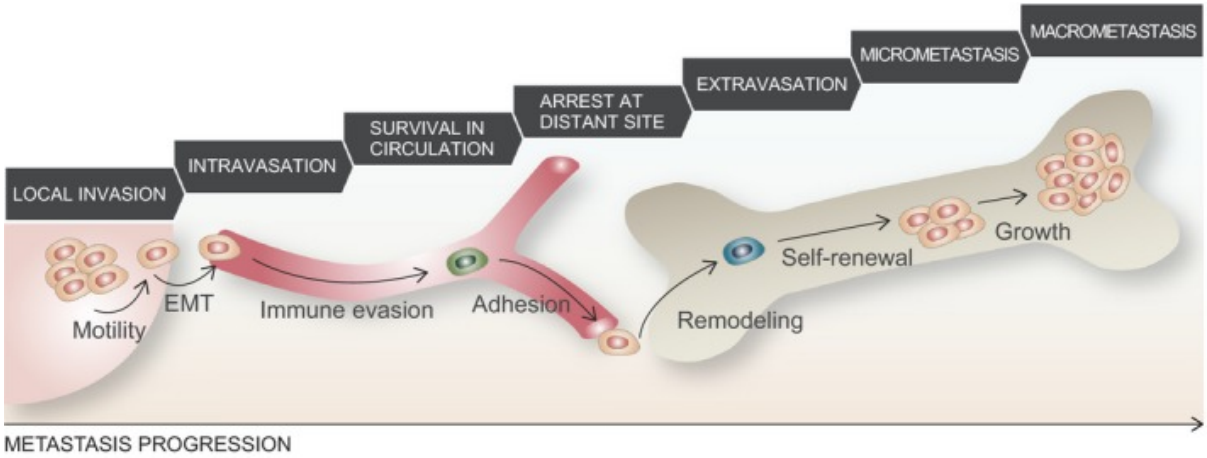


Figure 1.5: Circulating tumour cells within the metastatic cascade. The above figure demonstrates the metastatic cascade and the proposed life-cycle of a CTC – from intravasation and an epithelial-mesenchymal transition, to extravasation with a mesenchymal-epithelial transition and metastasis formation. Time spent in the circulation is crucial for immune surveillance and provides an opportunity for detection and/or therapeutic targeting.

CTCs are cells derived from a tumour mass (primary or metastatic) that have entered the vascular circulation. The exact mechanism and regulation by which CTCs gain entry into the circulation remains unknown - likely a combination of passive and active invasion and migration.[70] As a tumour mass grows and stimulates angiogenesis, tumour cells may be

shed passively into 'leaky' vasculature or lymphatic vessels.[71] However, increasing evidence points towards CTCs having an invasive phenotypic 'plasticity', as required for intravasation – the active migration across the basement membrane and vessel wall, and into the circulation.[72] These cells have a loss of cell-cell epithelial-like adhesion and increased cell motility, demonstrating an epithelial-to-mesenchymal transition (EMT).[72]

While EMT is a normal physiological pathway associated with processes such as embryogenesis or tissue regeneration and repair, in the setting of cancer it is seen as a critical step for epithelial cells to acquire a malignant phenotype.[73] Gene expression studies of tissue from multiple cancer types, including HNSCC, have demonstrated that malignant cells express a number of upregulated and downregulated genes associated with an EMT and that these are associated with worse clinical outcomes,[74][75] Jung *et al*. analysed gene expression data of HNSCC patients in TCGA, demonstrating that an EMT signature of 82 genes separated tumours into 'epithelial' and 'mesenchymal (EMT)' groups.[75] The EMT group had significantly worse disease-free survival (DFS) and OS, in addition to demonstrating increased expression of immune evasion markers PD-1 and CTLA4 and increased frequency of *TP53* and *EGFR* somatic mutations.

Several groups have investigated EMT-related gene expression in CTCs. In a study of 8 patients with prostate cancer, Chen *et al*. used RT-PCR to investigate expression of 84 EMT-related genes in 38 single-CTCs.[76] They were able to show increased CTC metastatic potential through upregulation of several key EMT-related genes, with a heterogeneous expression pattern between CTC populations. Of clinical note from this study was a defined subset of 18 EMT-related genes (e.g. *PTPRN2*, *ALDH1*, *ESR2*, and *WNT5A)* that separated castration-

resistant patients from castration-sensitive patients ($p<0.001$), demonstrating a potential prognostic biomarker signature. EMT marker expression in HNSCC CTCs is discussed below.

If current estimates of CTC numbers being in the region of tens of cells per ml of whole blood are considered accurate, then it would seem the majority of CTCs do not cause overt metastases. Experimental models (injecting cancer cells into mice) have estimated that the proportion of circulating cancer cells forming metastases to be around 0.02%.[77] Within the circulation CTCs are subjected to harsh survival conditions, including the shear forces of turbulent blood flow (CTCs are larger than red blood cells and most leukocytes) as well as immunological surveillance. Therefore, the half-life of CTCs is predicted to be very low, in the region of several hours and almost certainly no longer than 24 hours.[78], [79] The concept of organ specific patterns of metastases in different cancer groups has long been debated – dating back to the 'seed and soil' theory of Paget published in 1889.[80] Lymphatic and blood circulatory patterns cannot account for all cancer specific metastatic patterns and it is known that tumour cells express markers that facilitate growth and proliferation in specific organ tissues.[72] Following this, recent evidence has demonstrated that CTCs themselves may carry particular traits that infer patterns of metastasis. For example, analysis of CTCs from metastatic breast cancer patients has revealed a gene expression signature in a sub-population of CTCs strongly associated with brain metastasis. [81][82] As will be discussed, therapeutically targeting such CTC sub-populations could greatly improve outcomes in cancers with such aggressive metastatic properties.

There is most certainly a multitude of CTC-immune cell interactions contributing to cell death or survival, that we are yet to discover. Likely mediated via a combination of immune

regulatory marker expression on CTCs and direct interactions with immune cells. The immune checkpoint marker programmed cell death ligand 1 (PD-L1) binds to the PD-1 receptor on T-cells, acting to suppress inflammatory activity.[83] PD-L1 is expressed by several cancers, playing a role in immune evasion and is a recognised druggable target with several therapeutic agents approved for clinical use.[83] Thus, the expression of PD-L1 on CTCs is one potential route of CTC immune evasion, as observed in several cancer types, including HNSCC. [84], [85] In a study of 114 inoperable HNSCC patients, Strati *et al*. observed that CTC PD-L1 expression after chemo-radiotherapy treatment (12/54 patients) was a significant negative prognostic marker of PFS and OS.[84] In breast cancer patients, Szczerba *et al*. demonstrated how CTCs bound to neutrophils, in cell-clusters, exhibited upregulation of several genes involved in cell cycle regulation and DNA replication when compared to single (non-cluster) CTCs.[86] In addition, the tumour cell proliferation marker Ki67 was highly expressed in these neutrophil-clustered CTCs, in comparison to single CTCs and primary tumour cells. The presence of CTC-neutrophil clusters was significantly associated with worse PFS in these patients. From this data, it appears that in some situations immune cells may actually provide a protective role to CTCs, thus being pro-metastatic.

CTC-clusters are defined as a group of >3 cells (as observed via separate nuclei) and held together in close proximity.[87], [88] Across numerous cancer types, including HNSCC,[88], [89] evidence is growing that they have increased metastatic potential compared to single CTCs, and their presence is an adverse prognostic indicator.[90] In a mouse model, Aceto *et al*. demonstrated CTC-clusters as being oligoclonal cell clusters from a primary tumour mass and *not* CTCs aggregated within the circulation.[91] In addition, they calculated a CTC-cluster as being ≈50 fold more likely to cause a metastatic deposit than a single-CTC. When compared

to single-CTCs, CTC-clusters are detected in much fewer numbers, but this data should be judged with caution. Firstly, there is the potential for bias within current CTC enrichment/isolation methods that were developed for single cell detection. Secondly, CTC-clusters may exhibit a higher clearance rate and shorter half-life than single-CTCs, given their propensity to become entrapped within smaller capillary vessels.[91] Nevertheless, if these entities have such high metastatic potential, targeting them therapeutically has the potential for profound clinical benefit. Gkountela *et al*. identified 39 FDA approved compounds (several of them Na^+/K^+ ATPase inhibitors) that caused CTC-cluster dissociation in samples taken from breast cancer patients.[92] Of note, once dissociated into single cells, the previously clustered CTCs demonstrated hypermethylation of stemness related transcription factors (*OCT4, NANOG, SOX2* and *SIN3A*). RNA-sequencing confirmed downregulation of these stemness related genes involved in self-renewal and proliferation. Testing these findings in a mouse model, they observed that treating tumour bearing mice with the Na^+/K^+ ATPase inhibitor ouabain decreased CTC-cluster count and resulted in an 80-fold decrease in the metastatic burden. Thereby, the therapeutic targeting of CTC-clusters caused a phenotypic switch in CTC behaviour and tumour metastatic potential.

1.4.3 CTC isolation/enrichment techniques

The rarity of CTCs poses a significant challenge for strategies to enrich, isolate, quantify and characterise these cells. CTCs are estimated to account for <0.004% of all mononucleated cells,[79] or approximately one in a billion circulating cells.[93] CTC detection can be classified into two broad categories: those assessing biological characteristics of CTCs, such as surface

protein, or mutation/gene expression markers, referred to as 'marker-dependent' and those based upon physical properties of CTCs, such as size, deformability and density, referred to as 'marker-independent'. The CellSearch platform (currently the only FDA approved CTC detection platform) is an example of a marker-dependent platform that adopts principles of 'positive selection' to isolate EpCAM+ve cells using immuno-magnetic cell sorting. With a standardised and reproducible protocol, early research was quick to endorse this method of CTC detection. However, this method is inherently biased to miss CTCs that have low or negative EpCAM expression, for example those cells with an acquired invasive and migratory phenotype having undergone an EMT.[94] An alternative approach is to negatively deplete non-tumour PBMCs from blood, thus leaving CTCs remaining. In theory, it is experimentally simpler to remove a large homogenous population of peripheral blood mononuclear cells (PBMCs, all expressing the CD45 marker) than to try and detect CTCs. Several protocols exist to achieve this outcome. In general, most negative-depletion protocols require a first step of removing red blood cells, either through a red cell lysis protocol or a Ficoll/lymphoprep density gradient centrifugation to isolate PBMCs (including CTCs). Following this, immunomagnetic separation removes CD45+ve white blood cells (WBCs) leaving a CD45-ve cell population. Confirmation of CTC presence/phenotype can then be achieved through immunofluorescent (IF) antibody staining. While this technique is relatively simple and low cost compared to other automated platforms, particular disadvantages are a potentially high false negative rate due to CTC cell loss during the density gradient centrifugation step and an enriched cell fraction that is impure and contaminated with WBCs. Such an approach has proved successful when bulk sequencing the entire 'pooled' enriched fraction as, for example, the presence of RNA transcripts can still be detected from a low number of cells using PCR methods despite the WBC contamination.

Microfluidic CTC enrichment is one example of a marker-independent approach that uses the physical properties of CTCs being larger and less deformable than other circulating cells. Microfluidic devices capture CTCs by pumping whole blood, at pressure, through a plastic chamber or cassette with specifically designed micro-channels. These devices can be further classified into those using size-based filtration, streamline size sorting, cell polarizability (dielectrophoresis) or a combination of these. In addition, the microfluidic chamber/cassette can contain antibodies coated onto membranes or micro-posts, to incorporate an element of positive selection to retain CTCs within the device. In streamline sorting the pressurised flow of cells through microchannels of changing dimensions/shapes produces laminar flow of different sized cells and hydrodynamic separation of cells based upon size and deformability characteristics. An example of a device using streamline sorting is the Parsortix device, as discussed further in chapter 5.

1.4.4 Storage and preservation of CTCs in blood samples

A review of current CTC research quickly reveals not only a lack of consensus on enrichment technique, but also significant variability in the protocols used for collecting and storing blood samples prior to CTC analysis.[95] Broadly speaking, variability in pre-enrichment blood processing can be grouped into:

i) Blood collection tube (BCT) choice

ii) Time from blood collection to CTC enrichment

iii) Temperature for BCT storage.

Numerous BCTs are available and have been trialled in CTC research. These can be divided into non-CTC specific anticoagulant containing BCTs, such as EDTA or Citrate BCTs, or CTC/cell-free DNA specific BCTs. For example, the cell-free DNA BCT™ (Streck), Transfix® CTC BCT (Cytomark) or, CellSave® (Cellsearch). Such variety in current practice is partly driven by the fact that the output metric of CTC research may govern the method of CTC collection and stabilisation. In cases when enumeration of CTCs is the primary goal then cell membrane stabilisation is the priority to preserve whole intact cells. Several studies have reported increased recovery rates of spiked cells using preservative containing BCTs, such as the cell-free DNA BCT (Streck), compared to EDTA BCTs when stored for up to 4 days.[96] Several of these BCTs are designed for the stabilisation of both CTCs and cell-free DNA to enable a complete liquid biopsy approach from one blood sample. However, if molecular characterisation of CTCs is required, such as RNA sequencing, then these CTC-specific BCTs may impose limitations. While the exact make-up of the stabilising agent in each of these BCTs is unknown, they all contain compounds designed to stabilise proteins and/or a fixative agent (usually formaldehyde based). Formaldehyde is known to cause cross linkages across nucleic acid molecules and thus reduce the RNA/DNA yield and cause potential sequencing errors.[97] However, in an investigation of NGS compatibility with preservative BCTs, Yee *et al*. used the Cellsearch® platform to enrich breast cancer cell lines spiked in EDTA and cell-free DNA (Streck) BCTs.[98] The preservative BCT did not impact DNA extraction from enriched cells and they were able to successfully perform whole genome amplification and NGS to detect 3 known mutations (PIK3CA, TP53 and HER2) within the cell lines.

RNA degradation within CTCs has been shown to occur within a matter of hours.[99] Hypothermia of cells via cold storage aims to slow down cellular metabolism, thereby improving RNA stability and yield and enable more accurate assessment of CTC gene

expression. In a prostate cancer cell line model, Wong *et al*. compared RNA extraction from cells stored 'cold' (4°C) and at room temperature (RT) in EDTA BCTs for 48hrs and 72hrs prior to microfluidic enrichment (using the iCHIP device).[99] Cold stored cells demonstrated significantly less RNA degradation compared to RT storage ($p<0.0001$), with evidence of highly fragmented RNA in RT stored blood. Cold stored BCTs displayed near comparable RNA quality (fragmentation) from immediately (0hrs) enriched spiked cells. When they examined the expression of 26 prostate cancer specific genes in cold stored enriched cells there was no significant change in gene expression at 0, 48 and 72 hrs. In the same study, the authors compared RNA quality and expression of 40 genes in CTCs enriched from 8 prostate cancer patients, using multiple samples from each patient stored at 4°C for 0, 24 and 48hrs. Interestingly there was no difference in gene expression in CTCs enriched at each timepoint, even those blood samples stored for 48hrs. These findings were replicated by Apostolou *et al.,* who evaluated RNA yield from CTCs enriched from 2 cancer patients (breast and colorectal cancer) at multiple storage time points.[100] Blood samples were taken in EDTA BCTs and stored at 2-8°C, with CTC enrichment performed at 24hr intervals up to 96hrs. In this study they demonstrated no reduction in RNA yield (via difference in expression of 7 genes) up to 72hrs post sample collection when cold storage was used. Therefore, cold storage appears to be of benefit, particularly in non-preservative containing BCTs, when transcriptomic profiling of CTCs is being performed.

The obvious question then arises - *what effect will the enrichment technique have upon the gene expression of CTCs?* Physiological stress is known to alter gene expression of cells and what is thought to be a snapshot of a CTC expression profile may actually be a biased result due to sample handling protocol. Given that CTCs are either being magnetically pulled from

the blood, spun at high velocity or pushed through a microfluidic channel at pressure, it is reasonable to assume this will induce a cellular stress response. However, very few studies have investigated this possibility. A study in a breast cancer cell line model found no variance in the expression of 15 breast cancer associated genes before and after microfluidic enrichment of spiked cells, [101], although further investigation is required (as discussed in chapter 5). To add another layer of complexity, the site of blood draw may also impact CTC yield. Sun *et al.* demonstrated that CTC count and phenotype differed in blood samples from peripheral arteries/veins and those captured in vessels in close proximity to the primary tumour (hepatic and portal veins in hepatocellular carcinoma patients).[102] Those CTCs captured in peripheral vessels further from the primary tumour demonstrated an EMT-gene expression pattern, and higher ratio of mesenchymal vs epithelial CTCs. Of note, no difference was found between peripheral artery and vein samples. From a practical stand-point the majority of blood samples in clinical practice will be taken from peripheral vessels.

Taken together, the above data demonstrates the heterogeneity of sample collection protocols and the difficulty in drawing firm conclusions from studies evaluating multiple variables together, for example storage time and temperature. The key message to highlight is the need for a standardised protocol of sample collection and storage, that is specific and validated for the CTC enrichment method of choice.

1.4.5 CTC molecular characterisation

If stage one of a CTC detection protocol is the enrichment/isolation from whole blood; stage two involves characterisation of the enriched cells to a) demonstrate they are of tumour origin, and b) allow in-depth analysis of CTC multi-omic biomarkers, including proteomic, transcriptomic, genomic and methylomic markers. As with the variability seen in CTC detection, characterisation protocols also vary greatly and often this will be aligned to the enrichment isolation method used. The optimum strategy is to visualise the enriched cells using IF antibody staining and then perform subsequent downstream analysis once CTC presence has been confirmed. IF microscopy allows quantification of CTCs and interrogation of a small number of cell markers (dependent upon the microscope configuration but normally a maximum of 4 markers due to spectral overlap of available fluorophores). The accepted definition of a CTC from a carcinoma is a nucleated cell staining positive for an epithelial marker and DAPI (nuclear dye) and negative for CD45 (a WBC marker).[31][34] EpCAM and pan-cytokeratin antibodies (cytokeratin's 8, 18, 19 and 20) have both been used as epithelial markers in HNSCC.[104]–[106] Additional markers can be included to further define CTC phenotypic sub-groups, such as mesenchymal, stemness, invasive or immune checkpoint markers.

Flow cytometry is an alternative method of analysing the enriched cell population following negative depletion, providing both CTC quantification and a means for increased marker interrogation. However, unless flow assisted cell sorting is used to isolate predetermined populations, the cells are not able to be retained for further downstream analysis. A systematic review of 22 studies investigating CTCs in HNSCC reported that the majority of

these (15/22) used an immunomagnetic positive selection technique to identify CTCs.[107] Within this group the CellSearch® system was most frequently utilised.[107] The considerable diversity between CTC enrichment/isolation protocols creates significant problems when interpreting data from studies using differing techniques, in particular when CTC count is used as a surrogate marker for survival outcomes.

Options for genome or transcriptome characterisation of CTCs include DNA/RNA fluorescence in-situ hybridization (FISH), bulk sequencing of the enriched CTC population or single-CTC DNA/RNA sequencing. The majority of early research that sought to achieve RNA or DNA sequencing of CTCs, extracted RNA/DNA from enriched samples of pooled CTCs. While this method is successful in assessing the underlying genotypic or transcriptional landscape of CTCs, the enriched population is often 'impure' containing leukocytes. Furthermore, it does not allow an in-depth analysis of the contribution of tumour sub-clones to CTC heterogeneity or defining CTC sub-groups as clinically/therapeutically relevant biomarkers. In response to these needs, sequencing at the single-CTC level has become the goal of recent research.

1.4.6 Single-CTC sequencing and ITH

Several cancers, including lung, breast, colorectal and prostate, among others, have achieved CTC genetic sequencing at a single-cell level.[78] Heitzer *et al*. sequenced 37 CTCs from six patients with colorectal cancer, identifying concordant CNAs between matched tumour and CTC samples.[108] Single-CTC sequencing performed in 2 patient samples revealed 20 target gene mutations found exclusively within the CTCs. After undertaking follow-up in-depth

sequencing of the tissue samples, the majority (17/20) of mutations were identified as low frequency tumour sub-clones - thus CTCs were able to sensitively detect tumour clonal heterogeneity. In a similar study, Shaw *et al.* demonstrated significant heterogeneity (*PIK3CA, TP53, ESR1* and *KRAS* mutations) within 40 single CTCs from 5 metastatic breast cancer patients. Of note, CTCs demonstrated *ESR1* and *KRAS* mutations not found in matched tumour samples. Demonstrating driver mutations within CTCs and notably rare sub-clones is clinically relevant. Low frequency metastatic clones are often not detectable within the primary tumour.[17] Therefore, if a tissue biopsy of the metastatic lesion is unobtainable, the ability of CTCs to detect metastatic clonal expansion could be a valuable tool to delivering personalised treatment in this patient group.

In addition to the aforementioned identification of tumour genomic heterogeneity in CTCs, the true advantage of single-CTC sequencing potentially lies in the identification of gene-expression variation. Transcriptomic analysis of CTCs allows the elucidation of signalling pathways responsible for heterogeneity in CTC phenotypic sub-populations; in turn, providing a picture of disseminated cancer cell biology and potential therapeutic targets.[78], [109] Powell *et al*. profiled expression of 87 genes within 105 single-CTCs from 35 breast cancer patients (14 primary and 21 metastatic).[101] For the first time this group were able to define signatures of gene-expression within single-CTCs that correlated to EMT and clinical metastasis. These findings were echoed by Gorges *et al*. who performed single-cell RNA qPCR in a cohort of 55 CTCs from 5 breast and prostate cancer patients. In each patient CTCs were able to be grouped into at least two sub-population clusters based on differential gene-expression related to multiple cancer pathways – including EMT, DNA repair, treatment resistance, stemness and tumour progression.[109] These authors suggest that identification

of such metastatic cell diversity could effectively guide treatment selection, particularly in advanced stage and R/M cancer patients. Single-CTC methylation has also been investigated, demonstrating distinct differences in methylation profiles of single-CTCs and CTC clusters in breast cancer patients.[92]

CTCs have been investigated as markers of both spatial and temporal ITH. For example, CTCs have been shown to harbour druggable targets not evident in tissue biopsies. Fehm *et al*. demonstrated that in a cohort of 62 patients with HER2-negative breast cancer, up to 49% displayed HER2-positive CTCs and in only 50% of cases did HER2 positivity correlate between CTCs and matched tissue.[110] Similar findings of heterogenous tumour and CTC marker expression have also been demonstrated in androgen receptor (AR) expression in prostate cancer [111] and PD-L1 expression in HNSCC.[112]

The ability of liquid biopsies to identify acquired resistance mechanisms during or after therapy is a key strategy for clinical translation of this technology. Thus, CTCs being able to detect temporal ITH directly impacting clinical outcome is very powerful, particularly in metastasis. Single-cell RNA-sequencing of CTCs from metastatic prostate cancer patients demonstrated that non-canonical Wnt signalling pathway activation in CTCs was strongly associated with those patients who progressed while on AR inhibitor treatment.[113] In addition, serial sampling of CTCs has been shown to demonstrate a shifting mutation status in response to treatment. In a small cohort of metastatic colorectal patients, Kalikaki *et al.* tracked *KRAS* mutations in CTCs before, during and/or after chemotherapy +/- targeted therapy.[114] In several patients, tumour matched *KRAS* mutations in CTCs remained after

treatment, and the emergence of *KRAS* mutations in CTCs not seen in the tumour prior to treatment was correlated with progression or relapse on targeted therapy.

1.5 CTCs in HNSCC

1.5.1 CTC phenotypes in HNSCC and their use as diagnostic markers

As technology to detect and characterise CTCs has improved, so has the quality and depth of data output from research in HNSCC. In addition, research has shifted focus. Early research sought to optimise methods to confirm the presence of CTCs for diagnostic purposes. Following this, techniques that allowed enumeration of CTCs have provided a metric for comparison to survival outcomes and prognostic stratification. More recently, multi-parameter molecular CTC characterisation is the target of current efforts.

As discussed, the definition of a CTC in HNSCC is an epithelial marker positive (EpCAM or cytokeratin), nucleated DAPI positive, CD45 (leukocyte) negative cell. The first studies demonstrating the presence of CTCs in HNSCC were published in 2009. Winter *et al*. used a negative depletion technique to enrich CTCs and then performed RT-PCR to detect transcripts of four epithelial/tumour markers (ELF3, CK19, EGFR, EphB4) in a cohort of 16 patients (15/16 advanced stage II/IV disease).[115] Eleven patients (69%) had CTC transcripts detected before treatment, and in 10 patients CTCs were detected after treatment, of which 7 had detectable CTCs at both time-points. In the same year, Balasubramanian *et al*. were able to obtain IF confocal microscopy images of CTCs, enriched using a negative depletion technique and

stained with a pan-cytokeratin antibody.[116] In a cohort of 32 patients, 63% demonstrated detectable CTCs (treatment time-point not specified) and of those with detectable CTCs, the majority (75%) had <10 CTCs per ml.

The addition of non-epithelial surface protein markers to CTC detection antibody panels has enabled numerous CTC sub-groups to be identified in HNSCC. The aforementioned mesenchymal transition of a tumour cell to facilitate intravasation led several groups to investigate if mesenchymal phenotypic CTCs could be identified. Using N-cadherin and Vimentin as mesenchymal markers, Balasubramanian *et al*. enriched CTCs using negative depletion from 8 HNSCC patients.[117] By using multiple aliquots from each patient's blood sample they were able to assess combinations of more than 4 markers using IF microscopy for each patient – including EpCAM, pan-CK and EGFR as epithelial markers and CD44 as a marker of invasive potential. For the first time, this group demonstrated that CTCs were present with both epithelial *and* mesenchymal expression, mesenchymal only expression and variable CD44 expression. All patients demonstrated CTCs with epithelial and mesenchymal marker expression and 6 out of 8 patients had epithelial negative but mesenchymal positive CTCs present. This represented definitive evidence that marker-dependent epithelial only CTC detection approaches would miss CTC phenotypes with potential clinical significance. In a similar study, Weller *et al*. enriched CTCs via density centrifugation in 10 HNSCC patients, using N-cadherin as a mesenchymal marker and including CD133 as a stemness marker. Their findings echoed that of the above study, in demonstrating CTCs with both epithelial only, mesenchymal only and mixed phenotypes, in addition to CTCs with apparent stemness properties. However, they went further by comparing these sub-groups to clinical outcomes. CTC count for all of the above phenotypic sub-groups decreased significantly after surgical

resection (p=0.002). Of note, the presence of mesenchymal positive CTCs after resection was significantly associated with shorter OS (p=0.04). Hsieh *et al*. enriched CTCs based upon EpCAM marker positivity and then performed RT-PCR to assess podoplanin expression.[118] Podoplanin has been shown to be a marker of tumour invasion, nodal metastasis and poor prognosis in HNSCC. In 53 patients with R/M HNSCC undergoing chemotherapy, podoplanin expression was demonstrated in a subset of EpCAM+ve CTCs enriched prior to treatment. Absolute CTC count of these podoplain/EpCAM+ve cells was not correlated to PFS or OS. However, when the ratio of podoplanin+ve to -ve CTCs was calculated, a cut-off of >20% podoplanin+ve CTCs was significantly correlated to decreased PFS and OS (p=0.006 and 0.008 respectively). Such multi-parameter analysis of CTCs demonstrates that a sub-group(s) of CTCs possess an invasive phenotype that appears to correlate to a poorer prognosis.

It is interesting to observe a wide range of both the number of patients with detectable CTCs and CTC count between different studies. CTCs have been detected in 15% - 91% of HNSCC patients,[104], [105] with a CTC count in the tens to hundreds - ranging from 1 -282 CTCs/ml blood.[103] Again, this points to a fundamental lack of standardisation and comparability between different detection methods. However, despite this there is no obvious trend of each method producing consistently similar results, although the numbers of studies remains small. One possible explanation may be the lack of homogeneity between study cohorts with regard to stage of disease. The majority of studies have patient cohorts made up with advanced disease or R/M patients as these are often easily accessible for sample collection through clinical trials and offer the highest probability of CTCs being present for detection.

One would expect patients with a high burden of disease (T stage) or with active metastasis (N stage) and tumour dissemination to have CTCs present and/or an increased CTC count. In a study of 144 advanced stage (III/IV) HNSCC patients, Tinhofer *et al.* used a negative depletion method and flow cytometry assessment of EpCAM and/or pan-cytokeratin marker positivity to detect CTCs.[119] Overall 29% (44/144) were positive for CTCs. The presence of CTCs was not correlated with T stage, but nodal metastasis positive patients were more likely to be positive for CTCs. Using a microfiltration assay to detect EpCAM and pan-CK positive CTCs in 32 early and advanced stage HNSCC patients, Kawada *et al*. compared CTC count to tumour site, T and N stage.[105] CTCs were detected in 91% (29/32) of patients, with a mean count of 4 CTCs/ml and a range of 0-17.7 CTCs/ml. Tumour site and nodal metastasis displayed no relationship to CTC count. Patients with T3/4 stage tumours had a significantly greater CTC count than T1/2 stage patients (p=0.0046), as did advanced compared to early stage patients(p=0.031). With such contrasting evidence, it remains unclear as to the relationship between T and N stage and CTC presence or burden. Potentially due to small patient cohorts, or demonstrating an underlying aetiology entirely separate to tumour burden, for example individual patient CTC immune response and surveillance.

1.5.2 CTC prognostic risk stratification in HNSCC

While several groups have attempted to correlate CTC presence or count with survival outcomes, these findings should be treated with caution. Firstly, the patient cohorts are generally small proof-of-concept studies and not sufficiently powered for this purpose, which makes identifying CTC presence/count as an independent prognostic factor very difficult. Patients with advanced stage disease, larger and more invasive tumours with nodal

metastasis are known to have poorer survival outcomes. Given the above equivocal observations that CTC presence/count may or may not correlate with stage of disease, it is therefore difficult to mitigate for these risk factors in statistical analysis.

Despite this, the benefit of CTC count as a metric of disease burden or treatment response does demonstrate potential. The attempt to define cut-off values for CTC count and prognostic significance may hold some clinical merit, but only if methods of CTC enumeration are standardised. Jatana *et al*. reported that the presence of CTCs at the time of surgery was significantly associated with worse DFS (p=0.01), using a negative depletion and pan-CK antibody CTC detection strategy in 48 HNSCC patients.[120] When grouped based upon CTC count (0, <25, >25 CTCs/ml), patients with a count >25/ml had a significantly worse clinical outcome (DFS, p=0.04). Of note, no patient negative for CTCs went on to develop recurrence. Grisanti *et al*. used the CellSearch platform to identify EpCAM+ve CTCs in 53 R/M HNSCC patients and correlate this to PFS and OS.[121] They reported that a count of $\geq$2 CTCs/7.5ml of blood achieved significance to group patients into worse outcomes for PFS and OS (p<0.0005). Multivariate regression analysis identified CTC presence at baseline as an independent prognostic risk factor for both PFS and OS (p=0.002), with a mean CTC count of 5 cells/7.5ml and range 1-43. Such data clearly demonstrates how different CTC detection methods produce very different results of CTC count, further compounding problems in the clinical translation of this technology.

Aside from a static assessment taken pre-treatment for outcome prognostication, the ability to view a dynamic picture of tumour burden at multiple timepoints - before, during and after treatment - is arguably more clinically beneficial. Especially given the lack of a sensitive means

to detect recurrence or persistent disease in HNSCC. Several studies have investigated CTC count pre- and post-treatment, seeking to use CTC count as a biomarker of treatment response. In the largest study to date, Liu *et al.* analysed pre- and post-treatment blood samples from 178 patients with advanced stage nasopharyngeal/hypopharyngeal SCC, using a negative depletion enrichment method and epithelial marker FISH to identify and quantify CTCs.[122] Seventy four percent of patients were CTC positive at baseline, with a CTC count range of 1-22 CTCs/ml. In 140 patients the CTC count remained constant or decreased after treatment, with 90 patients CTC positive both pre- and post- treatment. However, in patients who exhibited an increase in CTC count post-treatment (38/178), both PFS and OS were significantly decreased ($p<0.001$ and $p=0.03$ respectively). In addition, patients who were negative for CTCs post-treatment had increased OS ($p<0.05$), and PFS was 1.2-fold higher than those with detectable CTCs.

The recently published CIRCUTEC trial is one of few studies to perform CTC count analysis at multiple time points during treatment.[123] In this trial 65 R/M HNSCC patients undergoing chemotherapy and cetuximab treatment were assessed at 3 time points (day 0, day 7 and day 21 of treatment).[123] Using the EPISPOT assay, CTCs were detected in 69% (45/65) of patients at day 0. An immediate change in CTC count from day 0 - 7 was an early predictive biomarker of treatment response and PFS was significantly lower in those patients with stable or increasing CTC counts ($p=0.01$) or $\geq$1 CTC/ml at day 7 ($p=0.03$). In a similar study, Inhestern *et al.* assessed CTC count in 40 HNSCC patients as part of a trial of induction chemotherapy, using negative depletion and flow cytometry.[124] CTC count was assessed before induction chemotherapy and at up to 4 other time points before each round of treatment, in addition to end of treatment. They adopted a slightly different analysis strategy to others, by

calculating the median CTC count across patients at baseline and during treatment to separate patients into high and low risk groups. RFS was significantly lower (p=0.025) when the CTC count at baseline was greater than the median value, as was OS if the maximal CTC count during therapy was greater than the median value during treatment (p=0.049). Such early work demonstrates the translational potential to use dynamic changes in CTC burden when measured serially, as a prognostic and/or predictive biomarker.

1.5.3 CTC derived predictive and therapeutic biomarkers in HNSCC

Biomarker-led personalised anti-cancer therapy has been shown to produce better outcomes and lower toxicity than non-personalised treatment,[125] especially in the R/M setting.[126] Unfortunately, biomarkers predicting response to treatment, e.g. targeted-immunotherapy, are underdeveloped, both generally and specifically in HNSCC. Immune checkpoint inhibitors (ICIs), such as the PD-1 inhibitors nivolumab[10] or pembrolizumab[127], have shown promise in treating R/M HNSCC. However, response rates of 20-30% in these patients has shifted the focus of research to finding biomarkers that can identify those who will benefit most from potentially morbid and costly treatment. Expression of PD-L1 on tumour cells (TPS – 'tumour positive score' $\geq$1%) has been shown to be a poor predictive biomarker for response to anti-PD-1 ICI treatment.[128] Therefore, a test that could use CTC derived biomarkers for immune/targeted therapy patient stratification has genuine clinical translational potential, as has been highlighted in other cancer types.[129]

Strati *et al.* were one of the first groups to investigate CTC expression of PD-L1 in HNSCC – using the CellSearch platform to capture EpCAM+ve CTCs and assess PD-L1 expression with RT-qPCR.[84] Samples were taken before, during and after treatment (when patients were available). PD-L1 expressing EpCAM+ve CTCs were observed in 25.5% (24/94) and 22.2% (12/54) of patients at baseline and end of treatment respectively. Baseline CTC PD-L1 expression did not predict treatment response or PFS/OS, however an absence of PD-L1 CTC expression at end of treatment was strongly associated with complete response (p=0.002). Multivariate analysis revealed that patients with PD-L1 expressing CTCs at the end of treatment had significantly lower PFS and OS (p=0.001 and <0.001). Chikamatsu *et al.* expanded this research by investigating PD-L1, PD-L2 and CD47 immune checkpoint markers using RT-qPCR in CTCs enriched by negative depletion in a R/M HNSCC cohort.[112] Twenty four of 30 patients were positive for CTCs, and among these 20, 17 and 19 patients demonstrated CTCs with PD-L1, PD-L2 and CD47 expression respectively. Expression of all three markers correlated strongly with each other, indicating a potential 'immune evasion switch' occurring in this CTC subset. However, for the first time this study investigated the heterogeneity between tumour and CTC immune marker expression. In 23 patients with available tumour immunohistochemistry, 13 expressed PD-L1, compared to 20 patients with PD-L1 expressing CTCs. Only in 10 patients did the PD-L1 expression in CTCs match the tumour and in 10 patients the CTCs expressed PD-L1 with no expression in the tumour tissue. Such data poses an interesting question – *are the CTCs detecting intra-tumoural immune marker heterogeneity that could be a more accurate biomarker for immune therapy, or are the CTCs acquiring these changes in the circulation and do not accurately reflect the tumour?*

Groups of markers or 'signatures' that encompass multiple functional domains provide a more comprehensive picture of tumour activity. Over the past few years, research into the molecular profiling of CTCs in HNSCC has slowly moved away from single marker expression to the goal of examining multiple functionally/therapeutically relevant markers. However, unless additional samples are taken from each patient, using IF microscopy is limited in this regard, as previously discussed. Using RT-qPCR on enriched CTCs, several groups have been able to investigate expression of multiple genes/markers. Tada *et al*. enriched CTCs using the CellSieve® microfilter from 44 patients with both early and advanced HNSCC.[130] Enriched samples were first analysed for RNA expression of 4 epithelial markers (EPCAM, MET, KRT19, and EGFR) and those positive for at least one of these (28/44) then underwent analysis for 10 further markers from four functional domains: cell growth (PIK3CA and CCND1), EMT (SNAI1 and VIM), stemness (CD44, NANOG, and ALDH1A1), and immune regulatory (CD47, PD-L1, and PD-L2). CTC positive patients demonstrated higher rates of non-response to treatment and loco-regional recurrence, with reduced PFS (p=0.0107). However, they were unable to correlate increased expression of the above 10 markers to clinical outcome on univariate analysis. The exception to this was CTC PD-L1 expression, which appeared to relate to increased PFS and OS (p=0.0346 and 0.0378 respectively). This finding is contrary to previous reports[84], [112] although this study assessed CTC PD-L1 expression pre-treatment and those discussed above found significance of CTC PD-L1 expression only at the post-treatment time point. In tumour histopathology, pre-treatment PD-L1 expression has been shown to be a marker of improved PFS in HNSCC. A notable outcome from the work by Tada *et al*. was the ability to separate CTC positive patients into 2 groups based on clustering analysis of the expression of 8 of the 10 markers (excluding PD-L1 and PD-L2). While these groups did not correlate to prognosis, the 'high' expression group was strongly correlated to advanced stage

disease (p=0.0228). For the first time this depth of CTC analysis glimpsed the potential benefit of deriving prognostic signatures from multi-parameter characterisation.

The disadvantage of the majority of CTC sequencing techniques used to-date in HNSCC, is that they examine gene expression on a bulk scale, using pooled CTCs from one enrichment procedure. If, as suspected, evidence demonstrates phenotypic heterogeneity within the CTC population then single-cell characterisation is required to further illuminate the underpinning genomic and molecular profile of these groups. The challenge in this endeavour is the low starting point of CTC numbers and thus low DNA/RNA input volume for sequencing experiments. Culture of CTCs ex-vivo has shown promise as it negates the low starting volume of CTCs and provides a self-replenishing source of cells to perform numerous molecular and sequencing assays.[131] Early success of CTC 2-dimensional culture was quickly followed by 3-dimensional 'spheroid' and 'organoid' CTC-derived culture models.[131] Zhang *et al.* were able to culture CTC spheroids from 36 hepatocellular carcinoma patients and use this model to demonstrate proof-of-concept to assess CTC sensitivity to 2 chemotherapeutic agents.[132] Such pre-clinical ex-vivo models provide a means of personalised medicine and drug sensitivity testing - particularly beneficial in R/M patients who have relapsed and undergone multiple rounds of chemotherapy treatment and in whom a biopsy may not be possible to provide tumour tissue for conventional patient-derived xenograft (PDX) animal models.

CTC-derived xenografts (CDX) in mice have provided a means to test chemotherapeutic agents and understand CTC tumorigenic potential and biology in more depth.[133] Such models can be used *after* first-line treatment in cases of recurrence to elucidate and test for

mechanisms of treatment resistance. Drapkin *et al*. created serial CDX and CTC culture models from a R/M SCLC patient prior to combination chemotherapy treatment and at the time of relapse.[134] The CTC model demonstrated initial drug sensitivity, but then resistance at the time of progression – thus mirroring the clinical course. The authors argued that CTC-derived culture models could be an invaluable tool for drug development, treatment stratification and early identification of resistance, especially in multi-drug R/M therapeutic regimens.[134] In HNSCC, Liao *et al*. were able to culture CTCs from 20 patients in 3-dimensional spheroids and perform gene expression analysis of 17 markers, including the multi-drug resistant proteins (MRP) 1,2,4,5 and 7.[135] While this study provided proof-of-concept in a small patient cohort, it was able to demonstrate that patients with CTCs that expressed high levels of the MRP class proteins had significantly decreased PFS ($p=0.011$) and the overall potential of using CTC culture models in HNSCC drug sensitivity testing.

1.6 Summary – unanswered questions and future directions

In summary, we have highlighted that HNSCC represents a high burden of disease for patients, with continued poor survival and high rates of recurrence. HNSCC demonstrates considerable molecular ITH, which in turn is a marker for poor clinical outcomes. CTCs are one part of a 'liquid biopsy' blood test and are demonstrated to have prognostic utility in HNSCC with expression of potentially druggable targets. However, the following unanswered questions remain:

- *Which biological pathways underpin the adaptation and survival of tumour cells once entered into the circulation?*
- *What technique of cell capture and downstream analysis offers the optimum method of CTC multi-parameter characterisation for clinical translation?*
- *With considerable ITH contributing to poor clinical outcomes and specifically treatment resistance, are CTCs an accurate representation of tumour molecular ITH?*
- *Finally, can CTCs provide biomarkers to direct treatment, in particular those with recurrence/metastasis?*

In this thesis we will discuss the utility of a novel microfluidic platform to enrich CTCs in HNSCC and investigate techniques to achieve multi-parameter characterisation, beyond that which is already accomplished. In turn, we hope to provide a framework with which to answer and further research the above unanswered questions.

CHAPTER 2: CTC EXPRESSION OF IMMUNE-MARKERS IN HNSCC: A SYSTEMATIC REVIEW AND META-ANALYSIS

2.1 Introduction

The primary aim of this thesis is to optimise and validate a method to enable multiparameter characterisation of CTCs in HNSCC. Aside from the EMT status of CTCs, one area of particular interest is the expression of immune-related markers. Firstly, to elucidate possible interactions between CTCs and PBMCs, and secondly to determine if druggable targets related to immune checkpoint inhibitors can be detected on CTCs. Furthermore, it is unclear if expression of immune-related markers on CTCs are mirrored in the primary tumour. If identified, this would allow expression on CTCs to be used as therapeutic or predictive biomarkers in HNSCC. Thus, the ability to simultaneously profile multiple markers in the immune landscape is paramount to this goal and potential clinical translation, and will be discussed further in chapter 7. To assess current understanding with regards CTC expression of immune-markers a systematic review and meta-analysis was conducted.

The text and figures below were published as the following paper: *Circulating tumour cell expression of immune markers as prognostic and therapeutic biomarkers in head and neck squamous cell carcinoma: a systematic review and meta-analysis. International Journal of Molecular Sciences. 2020; 21(21): 8229. PMID: 33153130*

The development of immune checkpoint inhibitors (ICIs) has heralded a wave of optimism in treating R/M HNSCC patients. These drugs act to block the PD-1 receptor on T-cells which,

when activated by PD-L1 from cells in the tumour, causes immune suppression.[136] Trials of ICIs in HNSCC - for example the PD-1 inhibitors pembrolizumab[127] and nivolumab[10] - have demonstrated response rates in the region of 20% and improved overall survival in R/M patient groups. PD-1 inhibitors are therefore now recommended as monotherapy for patients who have platinum-refractory R/M disease (progressing within 6 months of chemotherapy).[137] Despite the success of these ICIs, 80% of R/M patients will still not respond to what is often highly morbid treatment. Biomarkers to direct treatment and predict response are therefore urgently required.[125] Research has sought to identify novel predictive and therapeutic biomarkers for ICIs, in addition to tumour PD-L1 expression.[53] These include genomic markers such as tumour mutational burden or microsatellite instability scores,[53] or gene expression profiles to identify tumours likely to respond to treatment, such as the T-cell inflamed tumour subtype.[138] One potential source of novel immunotherapy-related biomarkers is a 'liquid biopsy' – using a non-invasive blood test to detect circulating tumour DNA (ctDNA) or tumour cells (CTCs), which allows serial monitoring of treatment response.[83], [139] CTCs are cells derived from the primary or metastatic tumour mass that are circulating within the blood and are responsible for the formation of secondary metastatic deposits at distant sites.[72] CTCs have been demonstrated as a source of potential diagnostic and prognostic biomarkers in HNSCC, with the ability to detect recurrence and monitor tumour dynamics.[65] Thus, utilising immune-marker expression on CTCs as prognostic and/or immunotherapy-related biomarkers is particularly promising. Our aim was to systematically review immune-markers expressed on CTCs as potential biomarkers in HNSCC - to assess their efficacy and the potential future avenues of research.

2.2 Methods

A review protocol was designed a priori and approved by all authors. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was followed when performing this systematic review and meta-analysis.[140] A study selection flowchart is detailed in figure 2.1.

Search Strategy and inclusion criteria

Embase and Medline databases were searched using the Ovid interface. The following keyword search terms were used: ("head and neck cancer" or "head and neck squamous cell carcinoma" or "head and neck" or "HNSCC" or "SCCHN") and ("circulating tumour cell" or "circulating tumor cell" or "CTC" or "neoplastic circulating cell"). Records were searched with no time limits up to June 2020. The results of this search were filtered to remove duplicates. Reference lists from identified papers were also examined to search for other relevant studies that may not have been identified by the primary search strategy.

Papers were included if they reported a recognised immune marker expressed on CTCs as a biomarker in HNSCC - this included correlation to treatment or clinical outcome variables as predictive or prognostic biomarkers and also comparison to tumour marker expression as a therapeutic biomarker. No minimum follow-up duration criteria were set. Only English language written papers were included. Case reports and conference abstracts were excluded, but were screened to identify potentially eligible studies.

Study selection and data extraction

Titles and abstracts of all identified studies were independently evaluated by two authors (KP and MP) after the removal of duplicate results. Studies selected were then judged according to their suitability for inclusion in the systematic review based on the eligibility criteria outlined. Any disagreement regarding inclusion was referred to a third author (PN). Data was independently extracted from each paper by the two reviewers and entered into a pre-designed excel proforma sheet. For each study we collected the following data points: author and year, stage of disease, size of patient cohort, method of CTC enrichment/detection, method of marker/gene identification, marker assessed, treatment timepoint at which CTCs were analysed, clinical outcome variable assessed with statistical data and follow-up duration.

Data analysis

Due to considerable heterogeneity between included studies (including the different immune-markers investigated), a narrative synthesis was undertaken in the first instance. Data was presented graphically as forest plots and meta-analyses were performed using Cochrane RevMan.v5 software. Log-rank hazard ratios (HRs) derived from Kaplan-Meier curves were presented, where it was possible to calculate. If HRs and confidence intervals (CI) were not presented in the paper, the corresponding author was contacted to request further data. If log-rank analysis was not available, univariate cox regression modelling HRs and CIs were utilised and analysed separately to log-rank data. As heterogeneity between studies was high (I^2>50%), a random effects model was used for meta-analysis of individual markers investigated in multiple studies (PD-L1). In addition, 2 studies which presented post-treatment data from homogenous cohorts were also examined in separate meta-analyses.

Quality assessment of included studies was undertaken using the Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK) checklist.[141][142]

2.3 Results

Included studies and patient clinical demographic data

After excluding reports not meeting the selection criteria (figure 2.1), we identified 5 studies for qualitative analysis and quantitative meta-analysis (table 2.1 and 2.2). Of note, one paper which was not included did discuss CTC immune marker expression (TLR4, TLR7 and TLR9) in HNSCC.[143] However, while the paper stated that these markers were not associated with PFS or OS, it did not present the actual data and so was excluded.

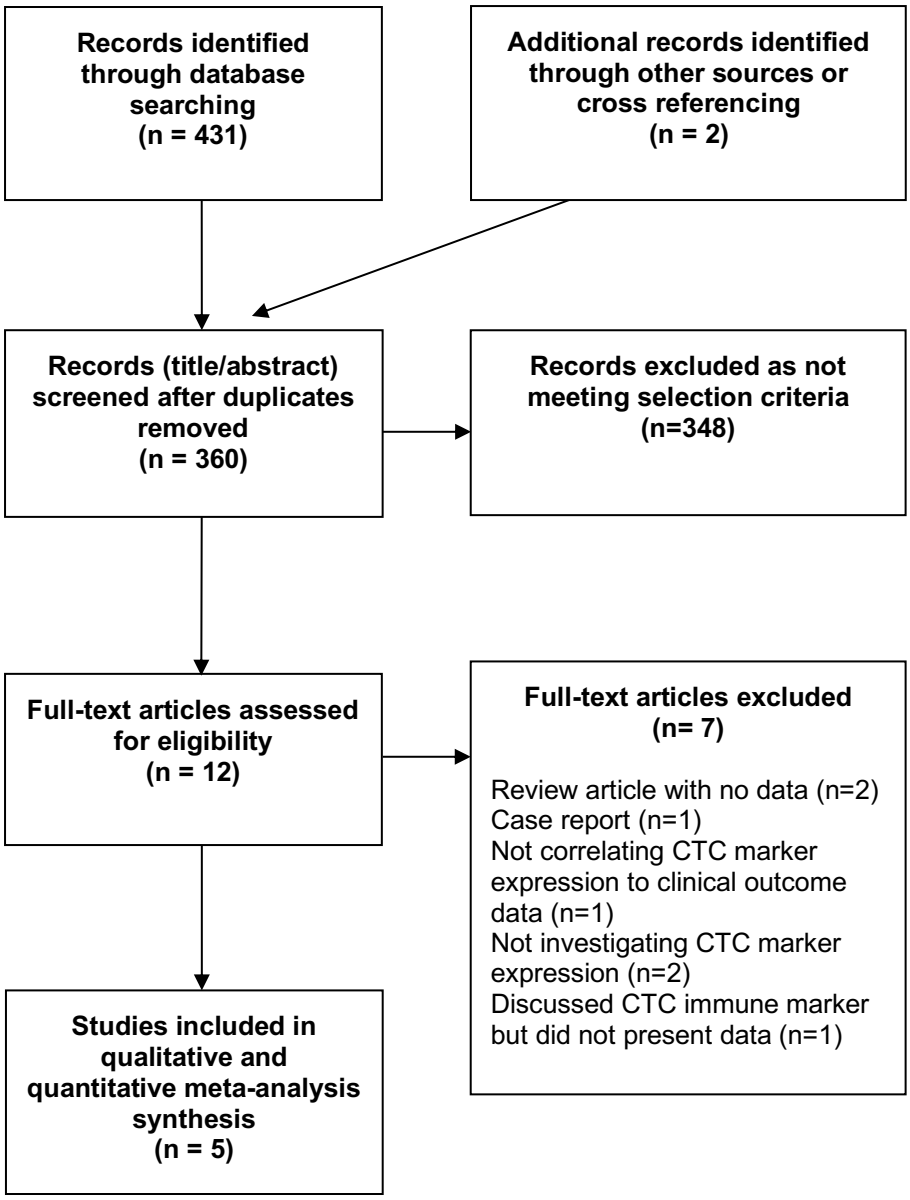


Figure 2.1: Systematic review study selection flowchart. Of 431 papers initially identified, 348 did not meet the eligibility criteria. A further 7 were excluded after reviewing the full text. A final cohort of 5 papers remained.

Considerable variability existed in the stage of disease between included studies – with 2 studies including stage I-IV HNSCC, 2 studies including locally advanced (stage III/IV) disease and 1 study only including R/M patients. Documentation of treatment regimens within studies on the whole was acceptable. Two studies that assessed pre-treatment CTC status

only stated that patients were 'treatment naïve'. In the R/M cohort study, previous primary treatment was detailed but secondary/palliative treatment was not discussed. In the two papers with post-treatment data, both cohorts received cisplatin-based chemo-radiotherapy.

CTC detection and disease stage

Assessing the correlation of CTC detection rate and disease stage between studies was not possible. In the two papers including locally advanced HNSCC patients, actual figures of 'epithelial CTC positive' patients were not reported.[84], [144] In these papers an EpCAM positive enrichment strategy was assumed to capture all epithelial expressing CTCs and gene expression values of this pooled cell group were normalised against a control value to assess overexpression of the target gene. In the R/M patient study, 80% (24/30) of patients were CTC positive[112] and in the stage I-IV patient studies 47.8% (11/23) and 63.6% (28/44) demonstrated detectable CTCs.[130], [145]

Immune marker expression and disease stage

Two studies compared CTC immune marker expression and stage of disease. Strati *et al*. reported no statistically significant relationship between CTC PD-L1 expression and stage of disease at pre-treatment or post-treatment timepoints (p=0.340 and 0.328 respectively). Similarly, Economopoulou *et al*. reported no relationship between stage of disease and CTC IDO-1 expression at pre-treatment or post-treatment timepoints (p=0.63 and 0.34 respectively). The remaining three studies did not compare CTC immune marker expression and stage of disease. Comparing CTC expression of all immune markers across 4 studies that reported actual numbers of positive CTCs we observed a statistically significant difference

between the R/M patient group and stage I-IV (including locally advanced disease) patient groups (p=0.00327), being increased in the former.

HPV status

Study stratification based upon HPV status was not possible. Only one study clearly presented HPV status,[145] with 47.8% (11/23) of patients being HPV positive. However, this study did not correlate HPV status to CTC PD-L1 expression. In another study, Strati *et al*. stated that of the CTC positive patients with known PD-L1 expression status, 4 were HPV positive and 3/4 of these demonstrated PD-L1 expressing CTCs.[84] However, they did not state the HPV status of the entire cohort, or correlate this to clinical outcome variables. The three remaining studies did not discuss HPV status.[112], [130], [144]

CTC immune-marker expression as tumour therapeutic and treatment response biomarkers

One study correlated tumour and CTC expression of the immune-marker PD-L1.[112] Twenty four of 30 R/M HNSCC patients (80%) were positive for CTCs, and among these 20 (83%) demonstrated PD-L1 expressing CTCs. In 23 patients with available tumour immunohistochemistry, 13 tumours expressed PD-L1, compared to 20 samples with PD-L1 expressing CTCs. In 10 patients (43.5%) the tumour and CTC PD-L1 expression were concordant (all PD-L1 positive cases). However, 10 patients demonstrated CTC PD-L1 expression where the primary tumour was PD-L1 negative. This equated to a positive and negative predictive value of 50% and 0% respectively. Importantly, this paper did not correlate CTC PD-L1 expression and clinical outcome.

Table 2.1: Characteristics of included studies. Data of clinicopathological variables, CTC enrichment/detection methods and investigated marker/clinical outcome as extracted from included studies (PFS – progression free survival, OS – overall survival, MACS – magnetic assisted cell sorting).

Author	Year	Tumour site/stage	No. of patients in cohort	Cohort gender	Age (years)	Tumour site	CTC enrichment technique	Marker detection methodology	Immune marker(s)	Timepoint of CTC analysis	Outcome measure [median] and (range) F/U
Strati *et al*. [18]	2017	Locally advanced HNSCC	113 (pre-treatment=94, post-treatment = 54)	Pre-treatment: M = 75, F = 19. Post-treatment: M=37, F=17.	Pre-treatment: ≥65=40, <65=54. Post-treatment: ≥65=25, <65=29.	Pre-treatment: Oropharynx = 21, Other = 73. Post-treatment: Oropharynx = 13, Other = 41.	Negative depletion (red cell lysis and CD45 MACS depletion) and EpCAM MACS enrichment	Gene expression using RT-qPCR	PD-L1	Pre-treatment and post-treatment	PFS and OS [18.9 months (0.2-54.9)]
Kulasinghe *et al*. [21]	2018	HNSCC stage I-IV	23	M=17 F=6	<60=10 >60=23	Oral cavity = 9 Oropharynx = 14	Microfluidic (ClearCell FX CTChip marker-independent) enrichment	Surface marker assessment with IF antibody	PD-L1	Pre-treatment	PFS [not stated]
Chikamatsu *et al*. [20]	2019	R/M HNSCC	30	M=27 F=3	Median 70.5 (range 53-86)	Oral cavity = 3 Nasopharynx = 1 Oropharynx = 3 Hypopharynx = 12 Larynx = 4 Paranasal sinuses = 6 Parotid gland = 1	Negative depletion (density centrifugation and CD45 MACS depletion)	Gene expression using RT-qPCR	PD-L1, PD-L2, CD47	R/M post-treatment	Biomarker of therapeutic PD-L1 target on tumour
Tada *et al*. [22]	2020	HNSCC stage I-IV	44	M=41 F=3	<66=21 ≥66=23	Nasal cavity = 3 Oral cavity = 4 Nasopharynx = 2 Oropharynx = 17 Hypopharynx = 14 Larynx = 4	CellSieve microfilter	Gene expression using RT-qPCR	PD-L1, PD-L2, CD47	Pre-treatment	PFS and OS [not stated]
Economopoulou *et al*. [19]	2020	Locally advanced HNSCC	113 (pre-treatment=94, post-treatment = 54)	Pre-treatment: M = 75, F = 19. Post-treatment: M=37, F=17.	Pre-treatment: ≥65=40, <65=54. Post-treatment: ≥65=25, <65=29.	Pre-treatment: Oropharynx = 21, Other = 73. Post-treatment: Oropharynx = 13, Other = 41.	Negative depletion (red cell lysis and CD45 MACS depletion) and EpCAM MACS enrichment	Gene expression using RT-qPCR	IDO1	Pre-treatment and post-treatment	PFS and OS [27.16 months (2.3-69.3)]

Table 2.2: Extracted data from included studies. Hazard ratio (HR), 95% confidence interval (CI) and p-value log-rank data extracted from included studies for progression free survival (PFS) and overall survival (OS) (* = univariate cox regression, ** = author derived from presented data).

		PFS			OS		
Marker	***Pre-treatment***	**HR**	**95% CI**	**p value**	**HR**	**95% CI**	**p value**
	Strati *et al*. 2017	1.43	0.63-3.25	0.39	0.55	0.23-1.30	0.17
	Kulasinghe *et al*. 2018	5.16	1.01-26.33	0.049			
PD-L1	Tada *et al*. 2020	0.29	0.10-0.86	0.035	not stated	not stated	0.038
PD-L2	Tada *et al*. 2020	0.81	0.24-2.77	0.74	1.05	0.11-9.70	0.97
CD47	Tada *et al*. 2020	1.56	0.51-4.69	0.45	2.21	0.33-12.97	0.45
IDO1*	Economopoulou *et al*. 2020	0.23	0.02-0.50	0.018	0.57	0.19-1.36	0.21
	Post-treatment						
IDO1*	Economopoulou *et al*. 2020	1.20	0.25-3.23	0.75	2.92	1.11-7.70	0.011
PD-L1*	Strati *et al*. 2017	4.07	1.67-9.91	0.002**	7.96	2.65-23.89	0.0002**

One study analysed pre-treatment CTC PD-L1 expression as a predictive marker of treatment response, with data from treatment outcome data from 73 treatment naïve patients all receiving chemo-radiotherapy.[84] Eighteen of the 59 CTC positive patients demonstrated CTC PD-L1 expression. Seven of these 18 (38.9%) achieved a complete response (CR) to treatment. Of those who were CTC PD-L1 negative at baseline (41/59), 27 (65.9%) achieved a complete response. Pre-treatment CTC PD-L1 expression was not correlated to treatment response (p=0.085). However, post-treatment CTC PD-L1 expression was strongly correlated to treatment response (p=0.001). Of 10 patients with overexpression of PD-L1 at the end of treatment only 2 (20%) achieved a CR, whereas among 35 patients who were PD-L1 negative at the end of treatment 28 (80%) demonstrated a CR. Thus, PD-L1 negative CTCs at the end of treatment was strongly associated with CR (OR:16.00, 95% CI:2.76–92.72, p=0.002).

Pre-treatment (baseline) CTC immune marker expression as a prognostic biomarker

Four studies reported pre-treatment CTC immune-marker expression as prognostic biomarkers of PFS and/or OS (figure 2.2 and 2.3).[84], [130], [144], [145] Of these, 3 studies reported PD-L1 expression,[84], [130], [145] 1 study PD-L2 expression,[130] 1 study CD47 expression[130] and 1 study reported IDO-1 expression.[144] The relationship between pre-treatment CTC PD-L1 expression and PFS and OS appeared mixed. In a cohort of 94 patients assessed at baseline, Strati *et al*. reported 24 patients with PD-L1 expressing CTCs. In this cohort, CTC PD-L1- expression was not associated with PFS (HR: 1.43, 95%CI 0.63-3.25, p=0.395) or OS (HR: 0.55, 95%CI 0.23-1.30, p=0.172). Kulasinghe *et al*. reported on a smaller cohort of 23 patients, of whom 11 were CTC positive and 6 patients exhibited PD-L1 expressing CTCs.[145] They correlated CTC PD-L1 expression to significantly decreased PFS (log-rank, HR: 5.159; 95% CI: 1.011-26.33, p=0.0485). In contrast, Tada *et al*. demonstrated *improved* PFS and OS in patients with PD-L1 expressing CTCs (HR:0.289, 95% CI:0.0982-0.8643, p=0.0346 and HR undefined, p=0.0378 respectively). In the same paper CTC PD-L2 and CD47 expression was assessed, however neither of these markers significantly correlated to PFS or OS. In this cohort of 44 patients, 64% (28) were CTC positive of which 11, 8 and 16 were PD-L1, PD-L2 and CD47 positive respectively. Economopoulou *et al*. assessed CTC IDO-1 expression as a biomarker of PFS and OS in a cohort of 60 HNSCC patients. Twenty-four patients were IDO-1 positive pre-treatment which correlated with improved PFS (log-rank, p=0.01) but not OS (log-rank, p=0.19). The association of IDO-1 pre-treatment expression and improved PFS was retained with multivariate analysis (cox regression, HR: 0.19, 95% CI:0.03-0.46, p=0.017). Meta-analysis of pre-treatment CTC PD-L1 expression among 3 studies demonstrated no association with PFS (HR: 1.17, 95% CI: 0.28-4.97, p=0.83) (figure 2.2) and a high heterogeneity between studies of 79%.

Post-treatment CTC immune marker expression

Two papers evaluated CTC immune-marker expression at the post-treatment timepoint, of which both cohorts received cisplatin based chemo-radiotherapy regimens.[84], [144] In contrast to their findings pre-treatment, Strati *et al*. reported post-treatment CTC PD-L1 expression as a strong prognostic marker of deceased PFS and OS (log-rank, p=0.001 and <0.001 respectively).[84] Univariate Cox regression analysis demonstrated HRs of 4.07 (95% CI: 1.67-9.91) and 7.96 (95% CI:2.65-23.89) for PFS and OS respectively and in both cases these far outweighed HR significance from other conventional prognostic clinicopathological variables (figures 4.4 and 4.5). In the aforementioned study by Economopoulou *et al*. IDO-1 expression assessed post-treatment was found to be a strong predictor of decreased OS but not PFS (contra to pre-treatment findings from this study) (figures 2.4 and 2.5). Multivariate analysis corroborated this association (cox regression, HR: 3.27, 95% CI:1.03-2.05, p=0.008).

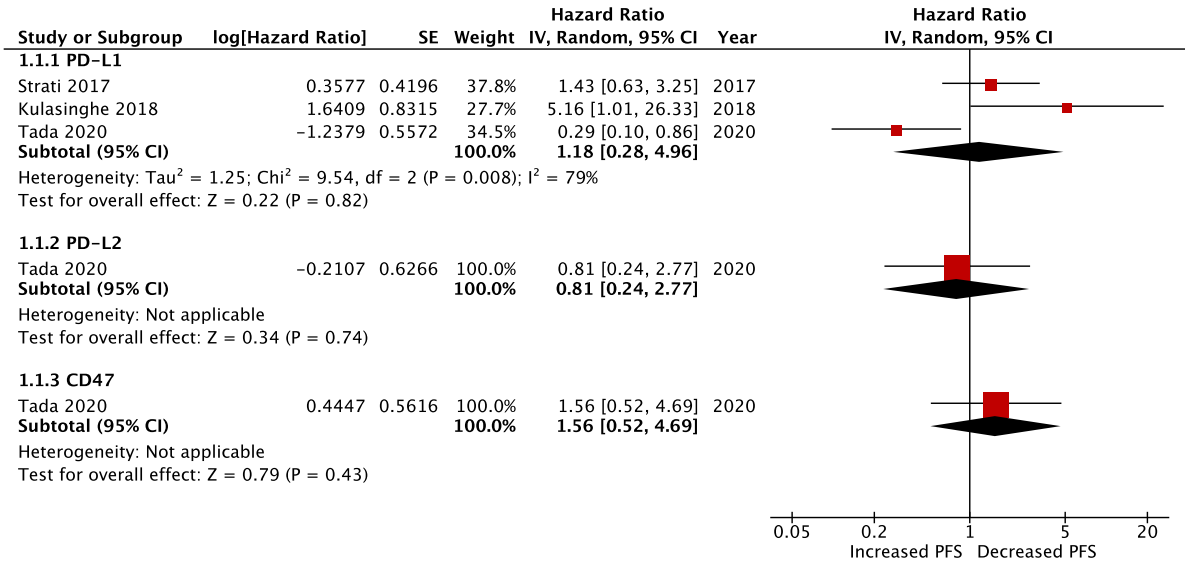


Figure 2.2: No association between pre-treatment CTC immune-marker expression and progression-free survival (PFS). Forest plot of pre-treatment CTC immune-marker expression as a biomarker of PFS, with no significant association in any marker (presented data may vary from table 4.2 due to conversion formulae).

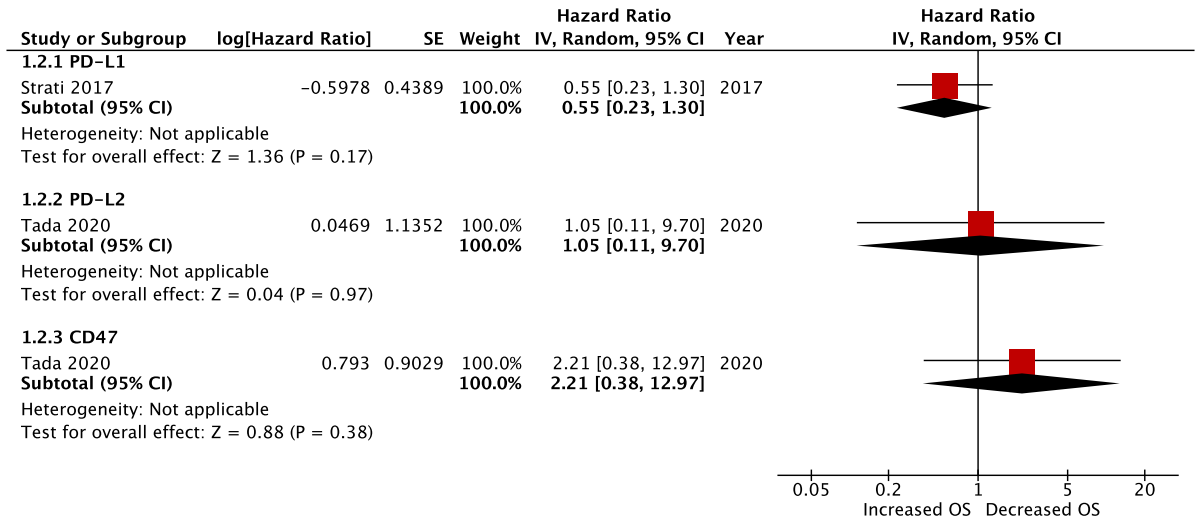


Figure 2.3: No association between pre-treatment CTC immune-marker expression and overall survival (OS). Forest plot of pre-treatment CTC immune-marker expression as a biomarker of OS, with no significant association in any marker (presented data may vary from table 4.2 due to conversion formulae).

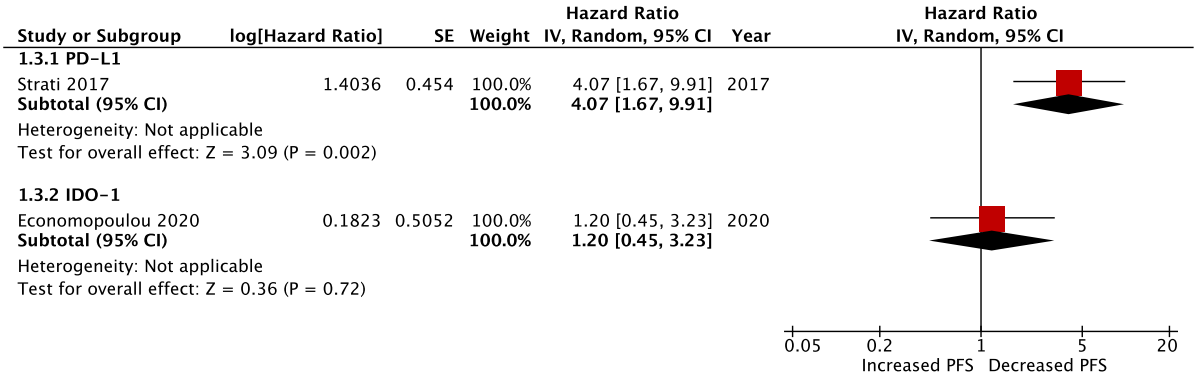


Figure 2.4: A mixed pattern of significance when post-treatment CTC expression of PD-L1 and IDO-1 is correlated to progression-free survival (PFS). Forest plot of post-treatment CTC immune-marker expression as a biomarker of PFS (presented data may vary from table 4.2 due to conversion formulae). Post-treatment CTC PD-L1 expression is significantly associated with decreased PFS, however no association is seen with IDO-1 expression.

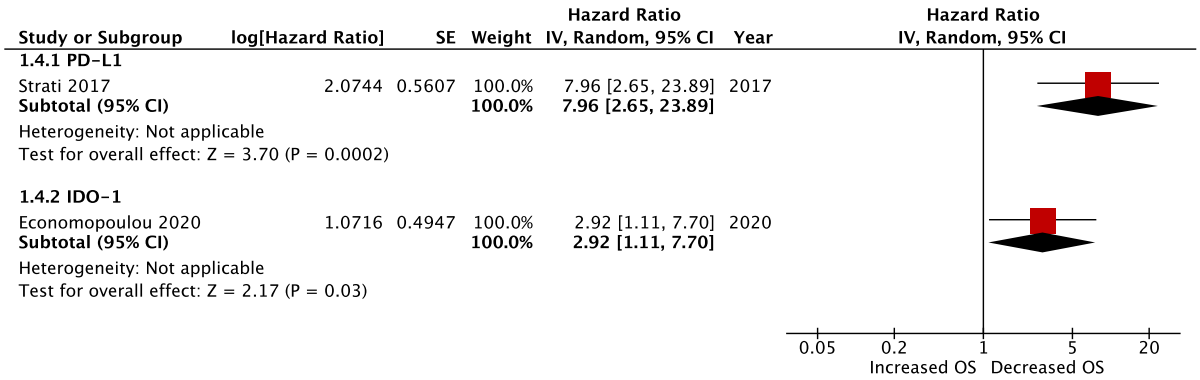


Figure 2.5: CTC expression of PD-L1 and IDO-1 post-treatment is associated with decreased overall survival (OS). Forest plot of post-treatment CTC immune-marker expression as a biomarker of OS (presented data may vary from table 4.2 due to conversion formulae). Both PD-L1 and IDO-1 expression are significantly correlated to decreased OS.

Study quality

The REMARK checklist was used to assess the quality of included studies. Notable shortcomings across all studies was the description of patient recruitment including exclusion criteria, randomisation and patient flow through the study. CTC enrichment strategies were generally well described, but blood collection strategies and specifically sample storage strategies were often vague. On the whole univariate and multivariate statistical analysis was presented in a clear concise fashion, however missing data or patient dropouts were poorly reported across studies.

2.4 Discussion

Since early research demonstrated proof-of-principle to identify CTC PD-L1 RNA transcript expression[146] and PD-L1 surface protein expression in HNSCC,[85] several studies have

sought to investigate multiple immune-markers expressed by CTCs as potential biomarkers. In this systematic review we identified 5 studies, all published within the last 3 years, which analysed small to large cohorts (range 24-113 patients) presenting correlation of CTC immune marker expression to clinical outcome data (PFS and OS). In meta-analysis, pre-treatment (baseline) CTC immune-markers did not demonstrate evidence of being prognostic biomarkers in HNSCC. In sub-group meta-analysis, data from 3 studies evaluating pre-treatment CTC PD-L1 expression demonstrated no significant correlation to PFS. These findings are similar to a recent meta-analysis of tumour PD-L1 expression in HNSCC tumour tissue, which reported no significant difference in survival between the presence of PD-L1 positive and negative tumour cells in IHC HNSCC.[147] To-date, few studies have been able to assess CTC target immune-marker expression in cohorts of ICI treated patients. In a cohort of 40 melanoma patients treated with pembrolizumab, pre-treatment CTC PD-L1 expression was identified as an independent predictive biomarker of PFS.[148] More recently, Zhang *et al*. demonstrated novel data that monitoring of PD-L1 expressing CTCs and circulating tumour endothelial cells (CTECs) *during* anti-PD-L1 therapy predicted treatment response. Albeit in a small cohort of 16 patients, the presence of PD-L1 expressing CTECs was associated with significantly worse PFS (p=0.012) and notably this trend was not observed in CTCs.[149] The investigation of immune-marker expression on CTC sub-populations, for example with endothelial, stemness or mesenchymal marker expression is particularly relevant to HNSCC, given evidence that these subtypes correlate with worse clinical outcomes.[65] Unfortunately, none of the studies identified in this review evaluated immune-markers in CTC sub-groups (instead investigating epithelial expressing CTCs only) or as predictive biomarkers for ICI therapy stratification. Clinical outcomes were studied in largely heterogenous treatment groups.

However, evidence from 2 studies identified in this review indicates that post-treatment CTC immune-marker expression appears to have prognostic significance. Strati *et al*. and Economopoulou *et al*. demonstrated that CTC expression of the immune-markers PD-L1 and IDO-1 at the post-treatment timepoint was associated with poor prognosis and reduced OS. IDO-1 is an interferon-induced enzyme associated with tumour immunosuppression.[150] Such prognostic significance of immune-markers may represent an 'immune-evasion' switch occurring in CTCs post-treatment leading to a poorer outcome. These findings corroborate those from Chikamatsu *et al*. who observed that CTC expression of the immune markers PD-L1, PD-L2 and CD47 were significantly correlated to each other (p=0.0285, 0.0005 and 0.003 respectively) in R/M patients who had failed primary treatment. Thus, reinforcing the hypothesis of a CTC sub-population with an immune-evasion phenotype.[151] In patients treated with ICI's, the persistence of target immune-marker expression on CTCs of patients failing treatment has been highlighted as a means to identify resistance through 'therapy escape'.[129] While at first glance the prognostic utility of pre-treatment CTC immune-marker (PD-L1) expression may appear poor, the benefit of liquid biopsies to enable serial testing of dynamic CTC immune-marker expression may provide a valuable tool in monitoring response to treatment, including ICI therapy. Among the studies included here the epithelial positive CTC detection rate was 48-80% (similar to other CTC studies in HNSCC [65]), thus raising questions as to the clinical applicability of this approach as a standardised biomarker assay. Further large cohort studies are required to investigate this. We found no evidence of serial CTC immune-marker monitoring in the post-treatment setting in HNSCC, however evidence is growing in other cancer types.[152] Disease progression in patients treated with ICI therapy has been correlated to increased numbers of post-treatment PD-L1 expressing CTCs in lung cancer.[153], [154] Strati *et al*. also clearly observed a significant trend for HNSCC

non-responders (treated with chemo-radiotherapy) to demonstrate PD-L1 expressing CTCs. In this cohort, CTC PD-L1 negative patients were up to 16 times more likely to have a CR to treatment than PD-L1 positive patients, although this was not related to ICI therapy per se, it demonstrates a potential differing CTC immune profile in non-responders.

Patient stratification for ICI therapy remains an ongoing dilemma in oncology.[83] Currently, tumour expression of PD-L1 is the only accepted means of patient stratification in HNSCC.[137] Expression of PD-L1 on tumour cells (TPS - tumour proportion score) has been shown to be a poor predictive marker of treatment response. [155], [156] Instead the composite proportion score/combined positive score (CPS), which assesses the ratio of PD-L1 expressing tumour *and* stromal cells is a more accurate biomarker and CPS $\geq$1 is currently recommended for patient selection.[137] However, CPS immunohistochemistry assays suffer from both inter-rater variability and inter-test variability from differing antibody epitopes.[157] In addition, tumour PD-L1 expression demonstrates both inter-patient (around 40% of HNSCC tumours express PD-L1[147]) and intra-tumoural heterogeneity.[158] Thus, if CTCs can be shown as a reliable source of druggable immune-markers they have potential to be a valuable clinical tool.[53] We identified only one study in HNSCC that correlated tumour and CTC PD-L1 expression.[112] CTC PD-L1 expression was seen to be a poor predictor of tumour PD-L1 expression, only being concordant in 50% of cases. Such findings of poor tumour PD-L1 expression predictive value in CTCs have been observed in other cancers.[153][154] There are several posited reasons for this discrepancy. Firstly, tumour PD-L1 expression is known to be both spatially and temporally heterogenous with considerable dynamic expression.[158] Of note, as seen by Chikamatsu *et al.*, studies in other cancers such as non-small cell lung carcinoma, have observed CTCs to be more often PD-L1

positive than tissue, up to two-fold.[153] This may represent PD-L1 intra-tumoural heterogeneity not detected in spatially biased single site tissue biopsies. Secondly, the complex relationship between CTCs and immune cells is still largely unknown. Such data poses an interesting question – *are the CTCs detecting intra-tumoural immune marker heterogeneity that could be a more accurate biomarker for immune therapy, or are the CTCs acquiring these changes in the circulation and do not accurately reflect the tumour?* Evidence continues to grow that CTC expression profiles are fluid and dependent upon the micro-environment, with adaptation in the circulation in response to immune cell interactions.[86] PD-L1 expression may be one facet of this adaptation,[151] however as yet this has not been directly investigated.

Assessment of gene expression in pooled CTCs from enriched blood samples allows interrogation of increased numbers of markers than is possible with conventional immunofluorescent microscopy. However, it does not provide quantifiable numbers of single-CTCs expressing the marker in question. Only one study in this review assessed expression of protein surface markers on CTCs.[145] Aside from correlating PD-L1 expression to PFS, a key message from this paper was the considerable heterogeneity of PD-L1 expression among CTCs from the same patient, even if over a small cohort. Of those 6 patients with PD-L1 expressing CTCs (range 1-20 CTCs per 3.75ml blood), only in 2 patients were all detected CTCs PD-L1 positive. Regardless of using gene expression or surface marker PD-L1 positivity, all the papers identified in this review adopted a binary approach of patients being CTC PD-L1 negative *or* positive. However, as validated in tumour PD-L1 assessment, CTC immune-marker heterogeneity warrants a scale or percentage to quantify marker expression. Such an approach of low, medium or high PD-L1 expression on single CTCs has been demonstrated to

more accurately stratify patient response to ICI therapy,[152], [159] and may start to tease out associations with tumour expression and survival outcomes not currently seen in reported studies of HNSCC. The increasing utilisation of single-cell sequencing technologies to allow single-CTC gene-expression profiling is sure to answer several questions regarding CTC immune-marker expression in relation to intra-tumoural heterogeneity. It remains to be seen if predictive ICI gene-expression profiles from HNSCC tumours are also expressed in CTCs and if such multi-omic characterisation will enable more accurate patient treatment stratification.

2.5 Conclusion

In summary, we identify only a small number of studies examining immune-marker expression by CTCs in HNSCC with sufficient clinical outcome follow-up data. As expected, the majority of studies assessed the expression of the druggable target PD-L1. No clear trend was observed between PFS/OS and pre-treatment CTC immune-marker expression – although multiple markers investigated by different studies make meta-analysis challenging. Despite these equivocal findings, post-treatment CTC immune-marker expression appears to hold prognostic significance and warrants further investigation. A key message is that to-date no studies have investigated CTC immune-marker expression in HNSCC ICI therapy patient cohorts. If the translational line of site is utilising CTC liquid biopsies to risk stratify and monitor treatment response in immunotherapy patients, then future research should focus on these patient groups in HNSCC.

CHAPTER 3: AIMS AND OBJECTIVES

3.1 Research question

Can a CTC based liquid biopsy in HNSCC inform patients of the molecular heterogeneity of their tumour and thus potentially match patients with appropriate treatment?

3.2 Aims

Primary aim: Optimise a marker-independent CTC enrichment technique in HNSCC to facilitate downstream single-CTC analysis

Secondary aim: Investigate if CTCs can identify HNSCC tumour-specific proteomic and/or transcriptomic ITH, to define potential therapeutic biomarkers

3.3 Objectives

1. Design a cell line model to determine optimal blood collection and storage conditions for microfluidic CTC enrichment and proteomic/transcriptomic characterisation
2. Demonstrate proteomic and transcriptomic profiling of CTC sub-populations in HNSCC and correlate this to patient-matched tumour heterogeneity and clinicopathological variables

3. Provide proof-of-principle to perform multi-parameter proteomic characterisation of single-CTCs using mass cytometry and identify upregulated oncogenic and druggable pathways that could direct treatment

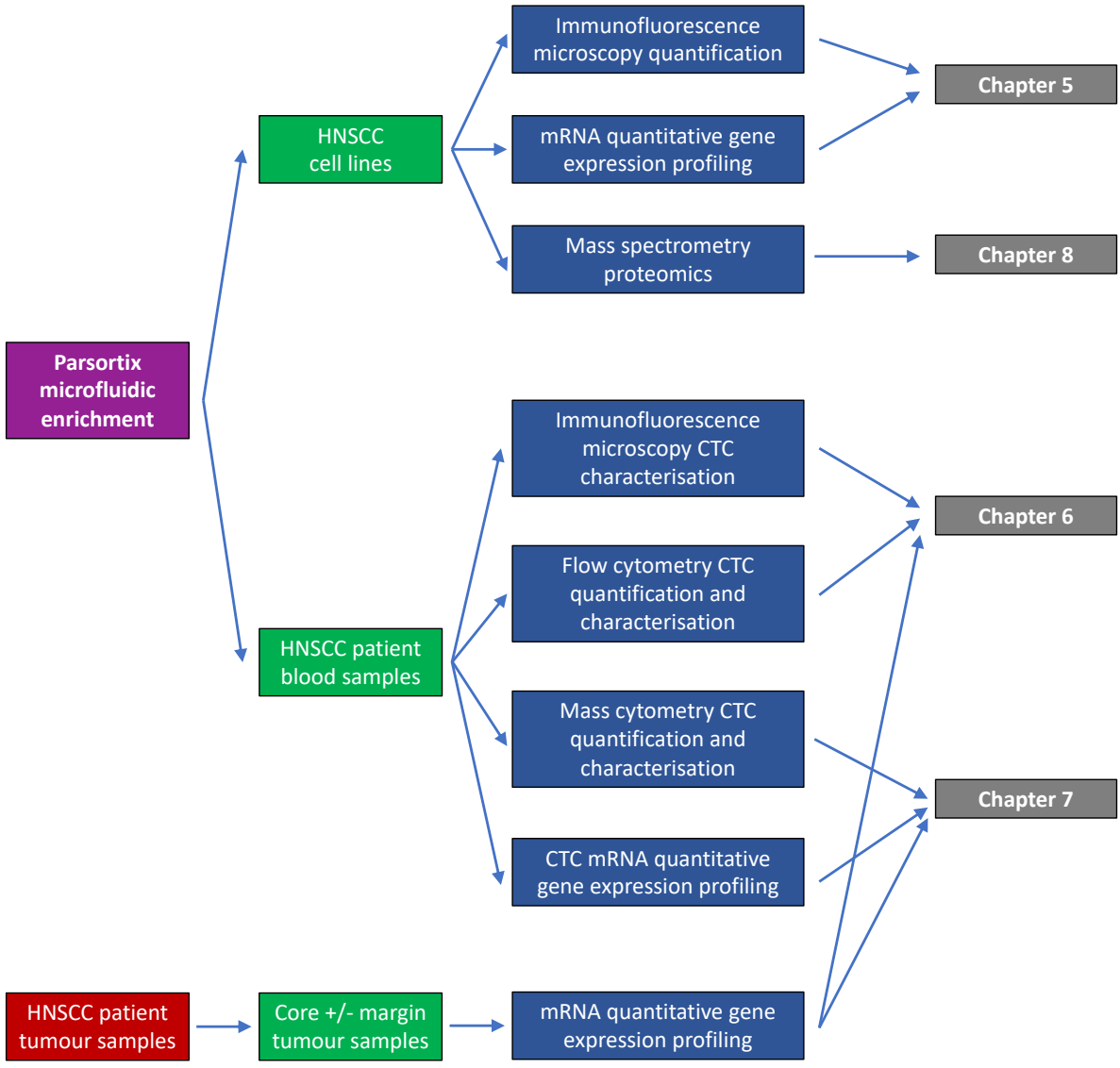


Figure 3.1: Thesis structure. A Schematic diagram of the biological samples used and the method of analysis and data output that contributed to each chapter.

CHAPTER 4: MATERIALS AND METHODS

4.1 Reagents

- Cell culture media
  - Dulbecco's modified Eagles medium (DMEM) – high glucose, Hepes modification (Sigma Aldrich D6171) - 500ml.
  - 10% foetal calf serum (Sigma Aldrich F9665) 500ml. 50ml added to 500ml bottle DMEM.
  - 1% non- essential amino acids NEAA x100 (Sigma Aldrich M7145) 100ml. 5ml added to 500ml bottle.
  - 1% GlutaMAX x100 (Life technologies 35050-038) 100ml. 5ml added to 500ml bottle.
  - 1% Penicillin/Streptomycin 5,000units/ml penicillin 5,00 ug/ml streptomycin (15070-063) 100ml. 5ml added to 500ml bottle.
- Lymphoprep – white cell isolation media (Axis shield 1114544).
- Phosphate buffered saline (PBS)
- Foetal calf serum (FCS)
- Freezing media – 90% FCS and 10% dimethyl sulfoxide (DMSO)
- Cell staining media (CSM) – PBS with 0.5% bovine serum albumin
- Paraformaldehyde (PFA)
- Iridium/saponin intercalator solution – 9.9ml CSM, 100μl 2% saponin and 0.5μl iridium

- Tissue dissociation media – 4ml cell culture media, 500µl 10X collagenase/hyaluronidase enzyme (Stemcell) (final concentration 1X) and 1µl Benzonase enzymes (Sigma aldrich) (concentration of 50U/ml)

4.2 HNSCC cell line culture and harvest

4.2.1 Cell line culture

The human papillomavirus (HPV) negative cell line FaDu (hypopharyngeal SCC: ATCC HTB-43) and SCC040 (tongue SCC: UPCI-SCC-040; ACC-660; DSMZ) were used for cell line spiking experiments. Cell lines were cultured as per a published protocol,[160] in cell culture media (see reagents section 4.1) at 37^{o}C in T75 flasks. Once >80% confluent, cells were split (1:10) and passaged – approximately every 3-4 days. All cell lines were tested for mycoplasma every six months and authentication via Short Tandem Repeat (STR) profiling was carried out on cell lines once a year.

4.2.2 Cell line harvest

Culture media was removed from the flask and cells washed with 10ml warmed PBS. 2ml of Trypsin/ETDA was then added and the cells incubated for further 2-3 minutes at 37°C. Cell dissociation was confirmed using light microscopy. 8ml of DMEM was then added to the flask to neutralise the Trypsin. The 10ml cell suspension was placed into a 15ml falcon tube and spun at 1500 rpm for 5 minutes. The supernatant was decanted and cell pellet resuspended in 5ml of PBS or DMEM. Cell concentration was evaluated by adding 10µl of the cell suspension to 10µl trypan blue and counted using the automated Invitrogen Countess cell counter.

4.3 Cell line spiking of healthy donor blood

To simulate cancer patient blood samples containing CTCs and facilitate enrichment protocol optimisation, a model of cell lines spiked into donor blood was used. Healthy donor blood was collected using the chosen BCT and either 4ml or 9ml used for each spiking experiment. Prior to spiking into blood, cell lines were labelled with the immunofluorescent dye CellTracker™ (CTr) Red (Invitrogen™) to facilitate identification and enumeration of captured cells after enrichment. 1mg vials of CTr were dissolved in 360μl of DMSO and 20μl volumes aliquoted into Eppendorf tubes for each experiment. One 20μl CTr sample was added to 8ml of PBS. Cell lines were cultured as described in section 4.2. For CTr labelling, the culture media was first removed and the CTr/PBS solution was added to the T75 flask and incubated for 5 minutes at 37°C. Following this the CTr/PBS solution was removed and cells harvested as per section 4.2.2. The cell suspension was serially diluted in PBS and X-Yμl added to the blood aiming for a final concentration of 10, 50 or 100 cells/ml blood. In each experiment, the same volume of the cell suspension spiked into the blood was added to three control wells of a 96 well plate. Cells in these wells were counted manually and a mean count taken as the final spiked concentration in the blood sample.

4.4 Parsortix cell line and CTC enrichment, antibody staining and immunofluorescence microscopy

4.4.1 Cell line/CTC enrichment

Cell line spiked blood was loaded into the Parsortix blood draw chamber in a 50ml falcon tube. Patient blood samples were loaded within the BCT. The manufacturer approved workflow was followed for enrichment protocol *PX2*_S99F.[161] A 6.5μm separation cassette was used.

4.4.2 Antibody staining

For flow cytometry experiments, captured cells were stained within the separation cassette after the enrichment protocol had been completed. The antibodies were loaded as a cocktail (table 4.1) in a single 15ml falcon tube together with DAPI, made up to 500μl with Inside Perm antibody buffer (Miltenyi Biotec) and placed into line 1 on the Parsortix device, using protocol *PX2_stain1.* Antibody optimisation is discussed in section 6.2.

Table 4.1. Antibody panel used for CTC detection. Table lists the phenotype marker and antibody clone, together with the conjugated fluorochrome and source

CELL TYPE MARKER	ANTIBODY	HOST SPECIES	TYPE	CLONE	FLOUROCHROME	SOURCE
EPITHELIAL	EpCAM (CD326)	mouse	Monoclonal, conjugated	9C4	Alexa Fluor 488	Biolegend
MESENCHYMAL	N-cadherin (CD325)	mouse	Monoclonal, conjugated	8C11	APC	Biolegend
PAN-LEUKOCYTE	CD45	mouse	Monoclonal, conjugated	2D1	Phycoerythrin (PE)	Biolegend
NUCLEAR DYE	DAPI					Biolegend

4.4.3 Elution of the Parsortix enriched cell population

Once the enrichment and antibody staining protocols had been completed, the enriched cell population was harvested from the cassette. The Parsortix device elutes the enriched cells in a 200µl volume of PBS buffer. This eluted fraction was used as per the different experiment designs.

4.5 Negative depletion protocol

A previously reported negative depletion technique,[106] comprising a density gradient centrifugation to isolate PBMCs followed by CD45 depletion using magnetic activated cell sorting (MACS), was compared against Parsortix enrichment in spiked cell line experiments.

Lymphoprep protocol - Four ml of lymphoprep media was pipetted into a 15ml falcon tube. The 4ml volume of cell line spiked blood was diluted with 4ml of DMEM in a 15ml falcon tube. The blood dilution was carefully layered onto the lymphoprep media and centrifuged at 591g for 30 minutes with no brake. The cells were removed from the interface of the lymphoprep media with a sterile pastette into a fresh 15ml tube, and washed twice in 10mls of DMEM (931g for 10 minutes).

MACS protocol - The cell suspension was then subjected to MACS using CD45 microbeads (Miltenyi Biotec) at a concentration of 20µl per 10^7 cells, following the manufacturers protocol[162] using a mini MACS magnet and MS separation columns. The enriched fraction

was eluted into a 96 well plate together with the Parsortix enriched fraction (and 3 x spiked cell volume control wells). CTr positive cells in each well were counted using the EVOS immunofluorescent microscope.

4.6 Flow cytometry of Parsortix enriched cells

Prior to use, the flow cytometer was washed and primed according to the manufacturers protocol. Compensation set-up was performed using the prepared antibody compensation controls. Laser voltage values from the most previous experiment were used as a starting point and adjusted accordingly. Blood samples from HNSCC patients taken before surgery (baseline) were collected into 9ml Transfix BCTs and stored for 24hrs at 4°C. Samples were run on the Parsortix device and antibody stained as per section 4.4.1 and 4.4.2. The Parsortix enriched cell fraction was eluted into a FACS tube and analysed on the flow cytometer immediately (<1hr). VersaComp antibody capture beads (Beckman Coulter) were used for compensation set-up of the three antibodies at time of processing.

Antibody compensation bead preparation – In a dark sterile hood one drop of the positive and negative antibody beads was added to 3 x FACS tubes, together with 100μl of PBS. 5μl of antibody was added in turn to respective FACS tubes (EpCAM, N-cadherin and CD45). These were gently agitated by hand, then incubated at 4°C for 30 minutes. Following this, 295μl of PBS was added to each tube to make up the final volume to 400μl. In addition, a non-stained control tube was made up, whereby one drop of positive and negative beads was added to 400μl PBS in a FACS tube.

4.7 Immunofluorescence microscopy of Parsortix enriched and antibody stained cells

The Parsortix enriched and eluted cell suspension was dropped onto a glass microscope slide (10µ per slide) and allowed to air dry for ≈10 minutes. Following this a coverslip was placed using ProLong Gold antifade slide mountant and allowed to cure for 24hrs. Slides were stored in darkness at 4°C. Staining was initially confirmed using the EVOS microscope, at 20X magnification. If cells were identified, then the Leica DM6000B microscope was used to take higher quality images at 40X and 100X magnification.

4.8 PBMC stimulation for phospho-antibody optimisation experiments

To test the binding of phospho-antibodies we used stimulated healthy donor PBMCs as a control. Cell activation cocktail (CAC) containing PMA and Inomycin (Biolegend) was used to stimulate the cells. The time required for cell stimulation varies between cells, with the manufacturer protocol stating a stimulation time of up to an hour depending on which pathways are required to be activated. However, overstimulation can cause cell death. We tested times of 5, 10 and 15 minutes in a healthy donor control PBMC population, using flow cytometry and fluorophore conjugated phospho-antibodies for pERK and pSTAT3. Cells were stimulated in cell culture media at RT, at a concentration 2µl CAC per ml of $2x10^6$ cells/ml suspension. We observed that 15 minutes was the optimum time for cell stimulation (figure 4.1). All phospho-antibody tests and control cells for mass cytometry samples were subsequently stimulated for 15 minutes with CAC.

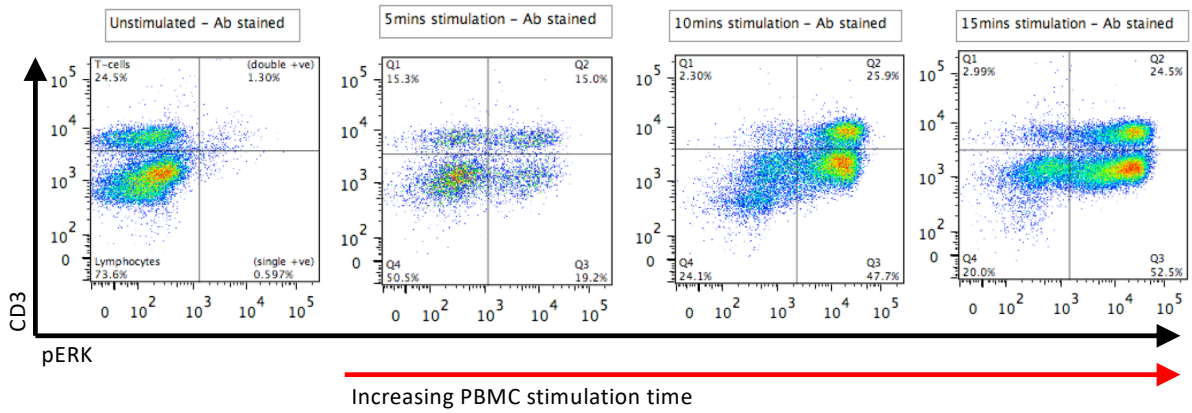


Figure 4.1 Stimulation of PBMCs for 15 minutes produces optimum signalling pathway expression. Results from flow cytometry analysis of PBMC stimulation experiments at 5, 10 and 15 minutes stimulation time. Bi-axial plots show the expression of pERK in CD3 positive (T-cells) and negative populations, with good stimulation at 15 minutes incubation time (results from pERK only are displayed as an example marker).

4.9 Sample processing protocol for mass cytometry

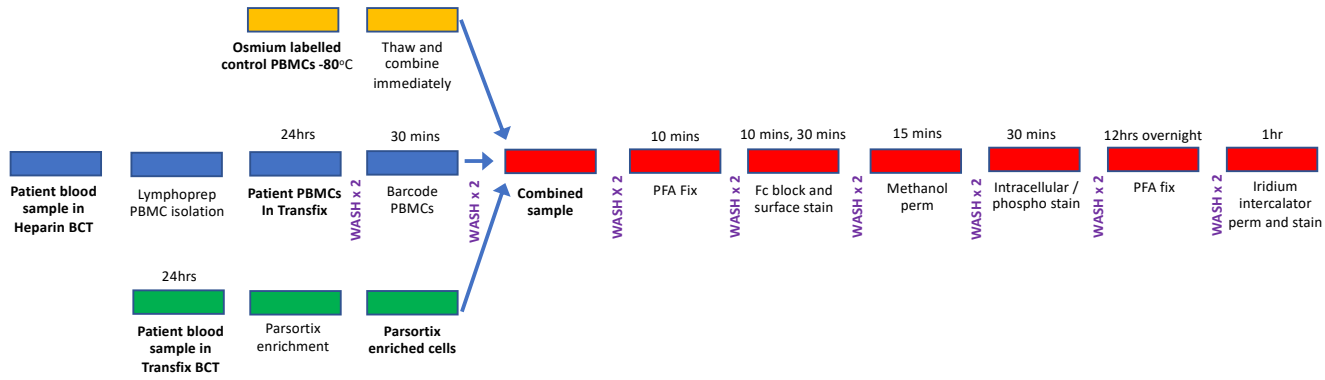


Figure 4.2: Workflow for patient sample processing and mass cytometry antibody staining. The above workflow describes the processing of blood samples to obtain patient carrier PBMCs and Parsortix enriched cells which are then combined with a stimulated PBMC control population prior to antibody staining and mass cytometry analysis.

PBMC stimulated positive control

As a positive control population for phospho-antibody staining we created a stimulated PBMC population from a healthy donor and osmium labelled the cells for gating out purposes. PBMCs from a single healthy donor at a single timepoint were isolated as described in section 4.5 and stimulated as described in section 4.8. After stimulation cells were kept on ice and washed 3 times in cold PBS and pelleted at 300g for 5mins. Stimulated cells were labelled with the heavy metal Osmium and aliquots of 100,000 cells were frozen in 1ml of freezing media in a 1ml cryovile at -80°C. For each experiment, one aliquot was thawed slowly in a water bath and immediately added to the Parsortix enriched cells and PBMC control carrier cells for PFA fixation.

PBMC carrier cells isolation and barcoding

Patient matched PBMCs were used as carrier cells during the antibody staining protocol to reduce the impact upon CTC numbers of cell loss during multiple washing steps. Whole blood collected in a heparin BCT underwent PBMC isolation via lymphoprep (section 4.5). One million cells in 1ml of culture media were incubated with 130μl of Transfix solution for 24hrs in a FACS tube. After incubation the cells were washed twice in CSM and pelleted at 600g for 5 mins. The pellet was resuspended in the residual volume (≈50μl) and stained with beta-2-microglobulin/ATPase antibody combination labelled with cadmium 106 as a barcode for gating purposes. After 30 minutes incubation at RT, the cells were washed twice in CSM, pelleted at 600g for 5 mins and combined with the other samples (figure 4.2).

Mass cytometry antibody staining of patient samples

Patient blood samples were collected in a Transfix BCT (9ml) and stored at 4°C for 24hrs prior to Parsortix enrichment as detailed in section 4.4.1. Once eluted, the enriched cell population was added directly to the PBMC carrier cells and PBMC control, in a FACS tube, washed twice in CSM and pelleted at 300g for 5 minutes. Mass cytometry antibody staining was performed as per a previously optimised protocol.[163] The pellet was resuspended and 250µl of CSM added, followed by 250µl 3.2% PFA to make a final concentration of 1.6% PFA. The cells were incubated at RT for 10 minutes. Following PFA incubation the cells were washed twice in CSM and pelleted at 800g for 5 minutes. The pellet was resuspended in the residual volume (≈50µl) and 1µl of the Fc blocker added for 10 minutes incubation at RT. Following this the extracellular antibody cocktail was added directly and incubated for a further 30 minutes at RT. After extracellular antibody staining the cells were washed twice in CSM and pelleted at 800g for 5 minutes. The pellet was resuspended in residual volume and kept on ice for 10 minutes prior to permeabilization. 1ml of pre-chilled methanol was added to the cells and incubated for 15 minutes on ice. Following permeabilization the cells were washed twice with CSM and pelleted at 800g for 5 minutes. The pellet was resuspended in residual volume and the intracellular/phospho antibody cocktail added and incubated for 30 minutes at RT. After intracellular staining the cells were washed twice in CSM and pelleted at 800g for 5 minutes. The pellet was resuspended and 250µl of CSM added, followed by 250µl 3.2% PFA to make a final concentration of 1.6% PFA. The cells were incubated at 4°C for 12 hrs (overnight). After the final fixation the cells were washed twice in CSM and pelleted at 800g for 5 minutes. After pellet resuspension, 1ml of iridium intercalator solution was added and the cells incubated for 1 hr at 4°C. After 1hr of intercalator staining the sample was either run immediately on

the mass cytometer (University of Birmingham MDS Core Facility), or frozen at -80°C for subsequent analysis at a later date.

Mass cytometry datasets were analysed using OMIQ and all analysis performed by Dr Karl Payne, with supervision from Dr Graham Taylor and confirmation of results by Dr Naeem Khan.

4.10 Establishing a confidence threshold for cytometry based (flow and mass cytometry) positive identification of CTCs

Healthy donor (HD) blood samples (section 4.12) were processed alongside patient samples in cytometry experiments (flow cytometry n=5 section 6.5.2 and mass cytometry n=5 section 7.4.1). In each experiment the HD sample was collected, processed and data analysed in the exact same method as per the patient samples. Using cytometry to characterise CTCs is highly novel, therefore the purpose of these HD samples was to provide a negative control for subsequent data analysis. From these results we set a confidence threshold with which we would define events meeting our CTC criteria to be 'CTC positive'. In both experiments the maximum number of events in any single HD was set as our confidence threshold. The number of replicates was chosen as we observed a low false discovery rate across 5 samples and the logistical difficulty of obtaining further HD samples. The primary purpose of this strategy was to negate against false discovery in 'CTC negative patients'. Cytometry is a method of cell characterisation primarily optimised for larger populations of cells, thus we were required to adapt and modify existing data analysis strategies for our specific needs.

4.11 RNA extraction, sequencing and analysis

4.11.1 RNA extraction from tissue, plasma and CTCs

RNA was extracted from tissue and unfixed cells using the RNeasy micro kit (Qiagen) as per the manufacturers protocol. RNA from Transfix preserved cells and/or Parsortix enriched samples was extracted using the Pico Pure RNA extraction kit (Thermofisher) as per the manufacturers protocol.

4.11.2 RNA sequencing

Library preparation was performed using the Lexogen Quantseq 3‘mRNA-Seq FWD Library Prep method and sequencing performed on an Illumina NextSeq 500 platform – both undertaken by the Genomics Birmingham (Institute of Cancer & Genomic Sciences, University of Birmingham).

4.11.3 Bioinformatics pipelines

Data from Quantseq sequencing underwent analysis using R. Sequencing reads were trimmed to remove adapter contamination, polyA read through, and low-quality tails using bbduk utility according to Lexogen guidelines. The trimmed reads were mapped to the GRCh38 (hg38) human genome with STAR aligner (v2.5.2b).[164] Reads mapping to genes were counted by the same software. Normalisation of read counts and differential expression analysis was performed with the DESeq2 R Bioconductor package [165], taking into account

the paired sample design of the experiment. Gene set enrichment analysis was performed using the GAGE (v.2.36.0) R Bioconductor package[166] with the Gene Ontology (GO) database. Hierarchical clustering of samples was performed using pheatmap in R, with a euclidean distance. RNAseq data analysis was undertaken by Dr Boris Noyvert, Dr Yi Pan and Dr Mohammed Elasrag from the Birmingham Centre for Computational Biology.

4.12 Mass spectrometry protocol and data analysis

Cells (fixed and unfixed) were processed for proteomics using the in-cell digest, as described by Kelly *et al*.[167] Cells were resuspended in digestion buffer (0.1 M triethylammonium bicarbonate + 1 mM MgCl2) and digested with benzonase (>99%, Millipore) for 30 min at 37 °C, followed by trypsin (ThermoFisher Scientific, 1:50 w/w protein) overnight at 37 °C. A second aliquot of trypsin (1:50) was subsequently added and incubated at 37 °C for 4 hours. A minimum of 25 ng of trypsin was added. Digests were acidified and desalted using Empore-C18 columns (Agilent). Non-fixed cell pellets were processed using S-TRAP (Protifi) using the manufacturer's protocol. Samples were then analysed by LC-MS/MS. Peptides were chromatographed on an Ultimate3000 RSLCNano over a 2-hour gradient before introduction to an EasySpray-Q-Exactive Orbitrap (ThermoFisher) and analysed using data-dependent acquisition. Raw files were processed using MaxQuant with the human reference proteome from UniProt.[168] Data were then analysed using R. Mass spectrometry data mapping was performed by Dr Tony Ly and subsequent statistical analysis performed by Dr Karl Payne.

4.13 Patient sample recruitment and data capture

Patient samples and healthy donor samples were recruited prospectively through the *Accelerated2 Sample Collection* Platform (REC ref:16/NW/0265). All patients and healthy donors consented to allow blood and tissue collection. Final TNM stage was taken from operative histology reports. In all patients except one, the pathologist reported using TNM edition 7 criteria. In the single case that used edition 8, the stage was converted to edition 7 notation. Follow-up time was measured from the date of consent and entry into the study, which for the majority of patients was the day of surgery. In a small number of cases when the patient was consented on the same day but surgery was delayed for a few weeks, the follow-up time was still measured from time of consent, so as to stay consistent.

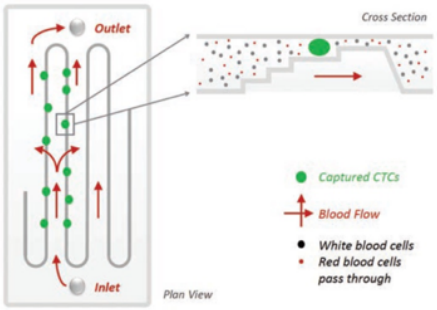
CHAPTER 5: OPTIMISATION OF THE PARSORTIX DEVICE FOR CTC CAPTURE AND DOWNSTREAM ANALYSIS IN HNSCC

5.1 The Parsortix CTC enrichment device

Platforms utilising microfluidic principles for CTC enrichment have gained traction due to their 'marker-independent' approach which negates the detection bias of previously utilised platforms that target single markers on the surface of CTCs. In addition, CTCs are 'captured' in a cassette, reducing the potential for cell loss as seen with alternative marker independent approaches using negative depletion techniques. The Parsortix device is a CTC enrichment platform that uses size-based sorting principles to enrich CTCs from whole blood samples.[169] Blood is loaded into the machine either in a blood collection tube (BCT) or 50ml falcon tube (figure 5.1). The sample is then pumped at 99bar pressure through a specially designed enrichment cassette. Seen on a cross section the cassette has a gradient tiered structure that traps CTC-like cells due to the microfluidic forces. The Parsortix enrichment process works on the assumption that CTCs are generally larger in diameter than PBMCs and less compressible due to their squamous cell origin. A 10ml whole blood sample takes approximately 2hrs to run through the machine, with PBS used as a buffer during the enrichment process.

A 

B 

C 

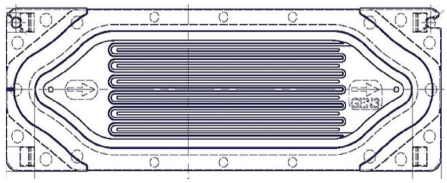
D 

Figure 5.1: The Parsortix CTC enrichment device. A) The Parsortix enrichment device: buffers loaded on left side, blood tube in the middle, antibody reagents in tubes on the right and the enrichment cassette sits within the metal housing on the top right. B) A close-up picture of the plastic CTC enrichment cassette. C) Pictorial representation of the tiered structure of the enrichment cassette and how the flow of blood captures larger CTCs. D) Schematic drawing of the enrichment cassette to show the multiple lanes and tiered design. (images reproduced with permission from Angle Plc).

The ideal CTC enrichment platform to enable clinical translation should provide an easily reproducible protocol for marker-independent CTC isolation. We chose to evaluate the Parsortix device because it allows a 'minimum user input' standardised approach to CTC enrichment with inbuilt protocols for numerous applications, for example in-cassette fixation and/or antibody staining. Once CTCs are enriched and captured within the Parsortix cassette they can then be antibody-labelled using a pre-installed protocol to confirm CTC status. Captured and antibody-stained cells can be viewed within the cassette using IF microscopy or harvested from the cassette in 200μl of PBS buffer for downstream applications.

Early research into the Parsortix device sought to validate its performance using spiked cell line models. Hvichia *et al*. evaluated capture rates of 5 cancer cell lines (pancreatic, melanoma, prostate, lung and bladder) spiked into healthy donor whole blood at concentrations of 10, 50 and 100 cells/ml.[170] Capture rates across the cell lines at differing spiked concentrations ranged from 42-70%. While the authors observed that at the lowest spiked cell concentration the capture rate decreased slightly, they did not perform statistical analysis to compare different spiked concentrations. A notable finding highlighted by the authors was that cell size was related to enrichment value. At the extremes of their range, the smaller T24 bladder cancer cell line (mean cell size 18μm) had the lowest capture rate of 42% compared to the larger PANC-1 pancreatic cancer cell line (mean cell size 23μm) capture rate of 70%. Using three lung cancer cell lines, Chudziak *et al*. compared Parsortix capture rates of spiked cells (50 cells/ml) at different blood volumes.[171] They observed no difference in capture rates between 0.5ml and 4ml of blood (70% and 69% respectively), demonstrating that an accurate CTC/ml reference can be obtained even from very small blood samples. In addition, the authors compared Parsortix enrichment to the Cell Search device (EpCAM marker-dependent CTC enrichment). As expected, for cell lines expressing EpCAM the capture rates were similar between platforms. However, with a cell line with low EpCAM expression the Parsortix device was far superior ($p=0.02$).

These findings, derived from spiked cell line models, were corroborated in a cohort of 12 small-cell lung cancer patients, whereby the Parsortix device successfully enriched cytokeratin positive epithelial CTCs that were not identified by the CellSearch device.[171] The benefit of Parsortix to enrich EpCAM low/negative CTCs, compared to CellSearch has been demonstrated in several studies as a key advantage of the device.[172], [173] In

metastatic breast cancer patients, Lampignano *et al.* compared Parsortix and CellSearch enrichment to assess CTCs with high or low EpCAM expression (cells with low EpCAM expression were confirmed as CTCs using cytokeratin marker expression).[173] Across 52 blood samples from 47 patients, 54% (n=28) demonstrated *both* EpCAM high and EpCAM low CTCs. CTCs were sorted from a group of 13 patients to examine *PIK3CA* mutational status (an independent negative prognostic factor in breast cancer). Of note, the presence and type of *PIK3CA* mutation was heterogeneous between EpCAM high and low CTCs. Moreover, in 1 patient the *PIK3CA* mutation was detected in EpCAM low CTCs only i.e. those CTCs that were detected with Parsortix and *not* CellSearch. The Parsortix device has now been optimised to detect CTCs within multiple cancer types, including lung[174], breast[86], prostate[175] and melanoma[176], among others. These early studies have provided proof-of-concept that molecular and genomic characterisation of CTC populations is possible following microfluidic enrichment using Parsortix.

Since commencing this project, one paper has reported the use of Parsortix in HNSCC. Zavridou *et al*. compared Parsortix and CellSearch based CTC enrichment in 50 HNSCC patients, evaluating gene expression (*CK19, PD-L1, EGFR, TWIST, CDH2)* and methylation (*RASSF1A, MLL3*) of key markers.[177] The Parsortix device was observed to be superior in detecting *CK19* expressing CTCs and in turn an increased number of CTCs with target marker expression, when compared to CellSearch. Although not to statistical significance, increased methylation was observed in Parsortix enriched CTC samples compared to CellSearch for the 2 CTC gene markers.

The results and figures from this chapter were published as the following paper: *Immediate Sample Fixation Increases Circulating Tumour Cell (CTC) Capture and Preserves Phenotype in Head and Neck Squamous Cell Carcinoma: Towards a Standardised Approach to Microfluidic CTC biomarker Discovery. Cancers. 2021; 13(21): 5519. PMID: 34771681*

5.2 Aims and objectives

The aim of this chapter was to validate Parsortix to isolate CTCs in HNSCC, using a cell line model, and compare the CTC capture rate to alternative negative depletion CTC enrichment techniques. In addition, we sought to determine the optimum collection and storage strategy for patient blood samples to facilitate downstream analysis, given the possible requirement to store samples on a short-term basis within this project, or transport samples between sites in a future clinical trial.

Objectives

1) Determine Parsortix capture rate of HNSCC spiked cells in healthy donor blood
2) Assess the impact of blood collection tube type on cell capture and cell viability/RNA integrity
3) Assess the impact of blood storage time on cell capture and cell viability/RNA integrity
4) Refine a protocol for RNA extraction from Parsortix enriched cells
5) Investigate if the Parsortix enrichment and antibody staining protocols impact upon the RNA expression profile of spiked cells

5.3 Capture rate of HNSCC spiked cells by the Parsortix device

5.3.1 The impact of spiked cell count on Parsortix capture rate

Initial cell line spiking experiments sought to determine the efficacy of Parsortix to enrich CTC-like HNSCC cells and compare this to a previously reported negative depletion technique. SCC040 and FaDu HNSCC cell lines were spiked into 4ml healthy donor (HD) blood (collected using EDTA BCTs) as a CTC model - aiming for concentrations of 10 (n=3), 50 (n=5) and 100 (n=5) cells/ml (methods section 4.3.1) - and processed immediately. These spiked cell concentrations were chosen for two reasons - firstly to know if the Parsortix device was able to enrich HNSCC cells at the low concentrations of CTCs reported in HNSCC patient samples [106], [117], [120], [121], [124] and secondly these concentrations were comparable to those in other cancer type cell line spiking experiments using the Parsortix device[170], [171] thereby allowing direct comparison.

Parsortix demonstrated a mean cell capture rate of 53.5% (95% CI: 50.394 – 56.696), across a range of 9-129 cells/ml (n=13) (figure 5.2a). Spiked cell concentration did not significantly correlate with capture rate (Spearman's r=0.382, p=0.276) and no difference was observed between the two cell lines. Both cell lines were similar in size - the SCC040 cell line had a mean cell diameter of 11.2 μm (n=4, range 10.5-12.2) and the FaDu cell line a mean cell diameter of 12.8μm (n=9, range 10.7-13.5). Grouping data into target spiked populations of 10, 50 and 100 cells/ml, there was a slight decrease in mean capture rates for 10 cells/ml compared to 50 and 100 cell/ml (50% vs 54% and 55% respectively), although this was not to a significant level (figure 5.2b, ANOVA, p=0.415).

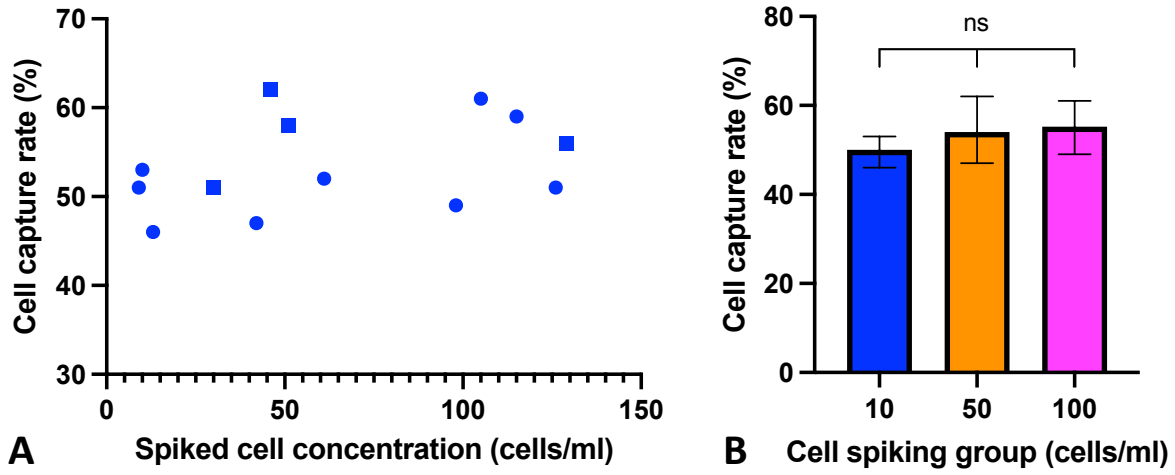


Figure 5.2: Spiked cell concentration does not impact capture rate. A) FaDu and SCC040 cells spiked into 4ml of healthy donor blood at stated concentrations. Each symbol represents an individual experiment (n=13). B) Collating results from experiments into target spiked concentration groups of 10 (n=3), 50 (n=5) and 100 (n=5) cells/ml, there was no significant difference in capture rate.

Following this, Parsortix enrichment of FaDu cell line spiked blood samples (collected in ETDA BCTs) was compared to a previously published protocol of CTC enrichment using negative depletion - comprising a lymphoprep and CD45 MACS depletion (see methods). Across triplicate matched samples with a mean spiked cell concentration of 115 cells/ml (range 105 - 126 cells/ml) the mean Parsortix capture rate was significantly higher than the negative depletion enrichment protocol, 57.0% vs 28.0% respectively (p=0.02) (figure 5.3). Thus, demonstrating the advantage of microfluidic enrichment over an alternative depletion protocol.

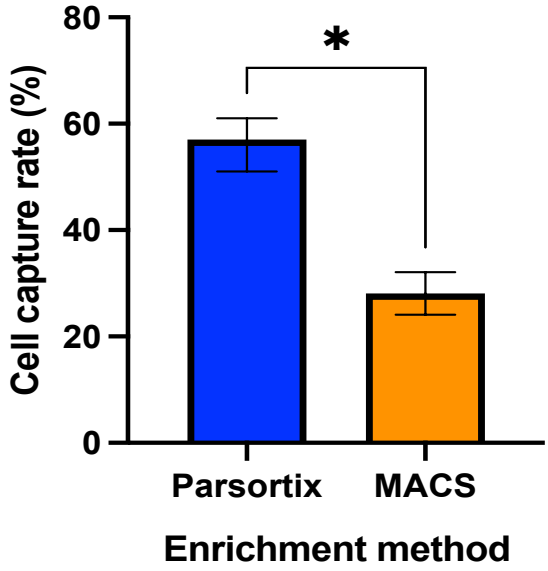


Figure 5.3: Parsortix is superior to CD45 MACS depletion to enrich spiked cells. FaDu cells were spiked into 4ml of HD blood at target concentrations of 100 cells/ml and enriched using Parsortix or lymphoprep and CD45 MACS depletion to test cell capture rate. Bars represent mean and whiskers range of values for n=3 per enrichment protocol. * = $p<0.05$.

5.3.2 Investigating cell preserving blood collection tubes and sample storage time

Having established a baseline capture rate of HNSCC cell lines spiked into ETDA BCTs, we sought to investigate if (a) cell preserving BCTs and (b) sample storage time had an impact upon CTC capture rate. The rationale behind this investigation was that in a clinical trial setting samples are taken at irregular times of the day and may not be able to be processed immediately. In addition, samples from multicentre trials are posted and can take up to 3 days to arrive at the host centre. Therefore, it is important that the reproducibility of CTC capture and subsequent downstream analysis between samples can be ensured if a sample is kept overnight or processing is delayed. Two methods of cell preservation with EDTA blood

collection were compared: the Transfix BCT, which contains a fixative agent and other proprietary cell stabilising compounds, and the Cytodelics whole blood stabilising kit, which allows whole blood to be frozen for up to 1 year with reported minimal cell loss upon thawing. The ability to freeze multiple samples would enable batch processing and downstream analysis. Eight millilitres of HD blood taken in EDTA and Transfix BCTs with the FaDu cell line at a concentration of 100 cells/ml. Samples were processed at 0hrs (either immediately/ <4hrs), 24hrs and 72hrs. The Cytodelics kit allows up to 1ml whole blood to be processed, therefore 100 cells in 1ml healthy donor blood was used for each assay and samples were frozen at -80°C for 24hrs.

Triplicate experiments for each spiked BCT at each timepoint were performed. At baseline 0hrs (sample processed immediately but requiring ~4hrs processing time) samples collected in Transfix BCTs demonstrated a significantly higher capture rate of 65.3% compared to 51.0% for EDTA samples (figure 5.4, p=0.0329). At subsequent timepoints (24hrs and 72hrs) the capture rate of Transfix preserved samples increased slightly to 73% and 73.1% respectively – potentially due to the prolonged fixation time resulting in increased cell rigidity. Across these timepoints Transfix demonstrated a significantly higher capture rate than EDTA samples (figure 5.4, p= 0.0145 and 0.0018 for 24hrs and 72hr respectively). EDTA samples stored for any period of time displayed significantly decreased cell capture rates compared to Transfix samples at the same timepoint - at 24hrs the capture rate decreased to 32.3% (p=0.0236) and then 6.0% at 72hrs (p=0.0016). The Cytodelics kit demonstrated a mean capture rate of 30.7%, which was significantly less than both EDTA and Transfix baseline 0hrs (ANOVA, p<0.0001) values, but comparable to the EDTA capture rate at 24hrs.

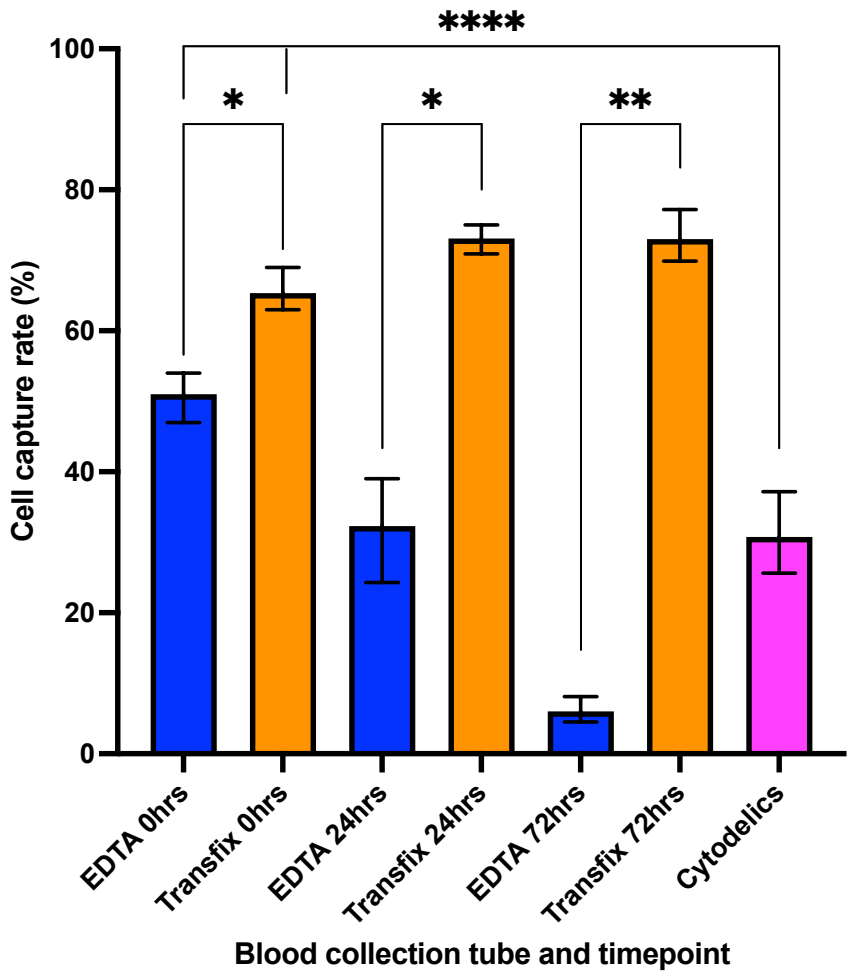


Figure 5.4: Transfix BCTs demonstrate superior cell capture rates compared to EDTA and Cytodelics blood preservation. HD blood samples collected into EDTA or Transfix BCTs were spiked with FaDu cells (100/ml target concentration), then enriched using Parsortix at time points, 0, 24 and 72hrs post-collection For Cytodelics, 1ml of HD blood was spiked with 100cells, then preserved at -80^{O}C for 12-24hrs before thawing and processing via Parsortix. Bar denotes mean cell capture rate and whiskers range of values for n=3 EDTA/Transfix, or n=5 Cytodelics, at each time point* = p<0.05, ** = P<0.001, **** = P<0.0001.

5.4 Impact of Parsortix microfluidic enrichment on gene expression

Cell fixation with formaldehyde/PFA halts cellular metabolic activity by forming covalent bonds between proteins and nucleic acids within the cell, while aiming to cause minimal protein structural change.[178] The stabilisation and 'freezing' of CTC metabolic activity at the time of sample collection using a fixation BCT, such as Transfix, has the benefit of protecting CTCs from changes in gene and/or protein expression during potentially traumatic enrichment protocols, thus accurately preserving the phenotype at point of blood collection, unbiased by subsequent processing. However, the majority of CTC research to date has used unfixed blood samples, partly due to the difficulty in nucleic acid extraction from fixed cells (as will be discussed).

To assess transcriptional changes occurring as a result of microfluidic enrichment, the gene expression profile of unfixed FaDu cells (n=3 paired samples) was assessed before and after Parsortix processing - via quantitative mRNA expression analysis (Lexogen QuantSeq). Analysis of transcript counts revealed that 303 genes were differentially expressed in Parsortix enriched cells, the majority of which (75.6%, 229/303) were downregulated (figure 5.5a). To interpret the functional impact of this altered gene expression profile we performed gene ontology (GO) functional enrichment analysis. GO analysis was able to group differentially expressed genes into biologically significant gene sets. A clear pattern emerged that cell function related to transcription - mRNA/RNA processing/binding, ribosome and protein processing/binding - was downregulated in enriched cells (table 5.1).

There were no GO sets upregulated, however when upregulated individual genes were examined there were some notable findings. Differentially expressed genes related to cell stress/injury (GDF15 [179] and ZNF557 [180]) and apoptosis (MALAT1 [181], TXNIP [182] and TRIML2 [183]) were upregulated in the enriched cell group. Of note, numerous genes relating to oxidative stress were upregulated after microfluidic enrichment, including: MMP1 [184], ZNF721 [185], HILPDA [186], SOD2 [187], MAP2K3 [188] and WSB1 [189].

To assess the potential implications of cell phenotype changes observed after microfluidic enrichment on CTC identification and clinical biomarker discovery, we compiled a list of 34 genes associated with HNSCC and CTC identification, including druggable targets, key pathway proteins and markers of EMT, stemness and proliferation/metastasis (table 7.1). Differential analysis of these genes revealed that keratins 17 and 19 and ribosomal protein S6 were significantly downregulated ($p=0.0207$, 0.0341 and 0.0112 respectively). In addition, keratin 5 trended towards downregulation ($p=0.0511$). Markers of EMT, stemness and druggable targets such as EGFR, CTLA4 and PD-L1 were not significantly altered by the enrichment process (figure 5.5b).

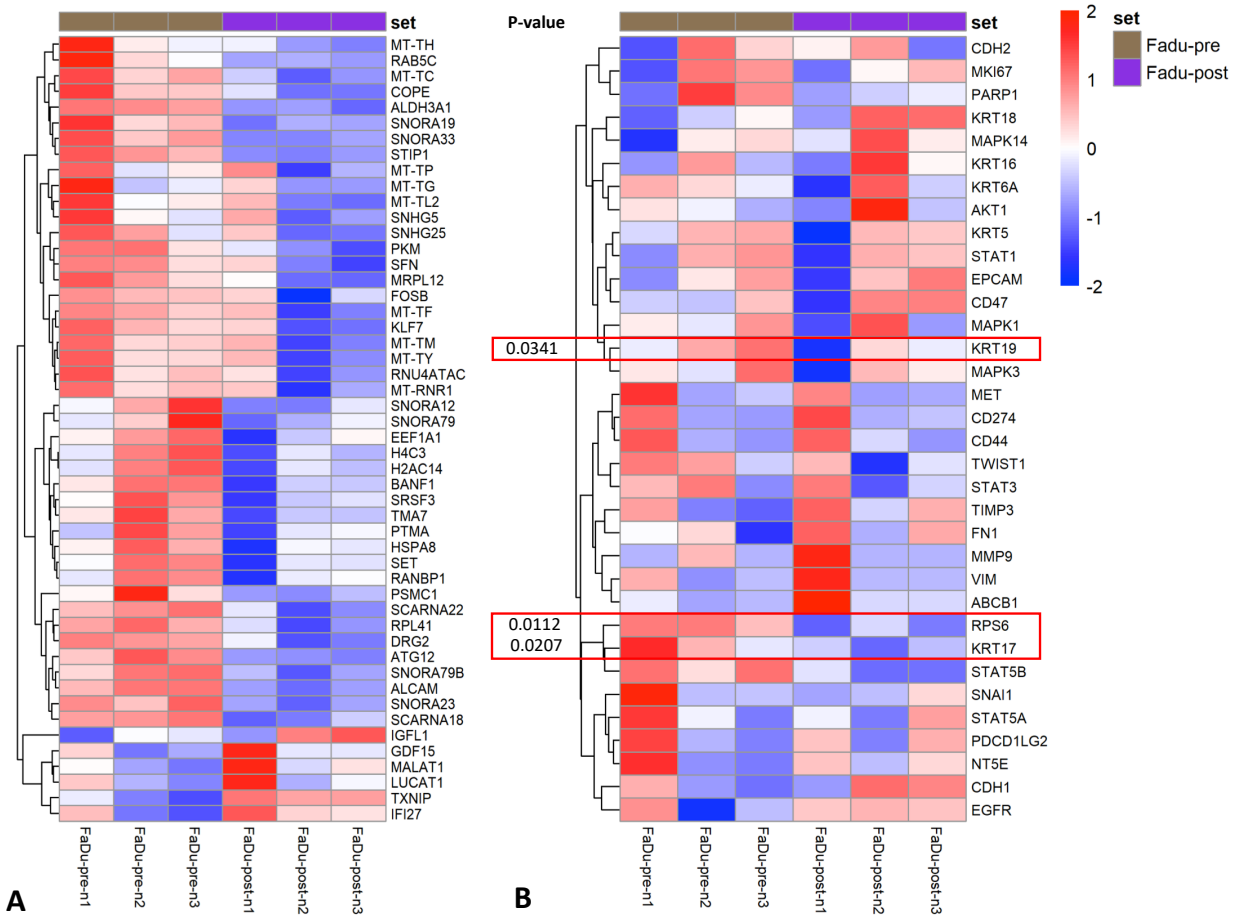


Figure 5.5: Differential gene expression analysis of FaDu cell line before and after Parsortix microfluidic enrichment demonstrates a cell stress response and downregulation of keratin proteins. Heatmap of differential gene expression analysis of paired unfixed FaDu cells before and after Parsortix enrichment. A) The top 50 significantly differentially expressed genes are displayed, the majority of which are downregulated and involved in mRNA/RNA and ribosomal protein processing. Upregulated genes indicate apoptosis and an oxidative stress response. B) Heatmap of gene expression for specific HNSCC/CTC target markers (rationale for marker inclusion outlined in table 7.1). Three genes (keratins 17 and 19 and ribosomal protein S6, highlighted within red boxes and p-values displayed) demonstrate significant downregulation after Parsortix enrichment, the remaining genes are not significantly altered (p-values not shown).

Table 5.1: Downregulated differentially expressed genes in Parsortix enriched FaDu cells demonstrate enrichment of mRNA/RNA and ribosomal processing gene sets. Functional enrichment analysis of gene ontology (GO) gene sets (biological process, cell component and molecular function) of downregulated genes in the Parsortix processed FaDu cells. Top five for each domain are displayed. There is a clear pattern of downregulation in mRNA/RNA processing and ribosomal proteins – as seen in cells undergoing a stress response and/or apoptosis.

Gene Ontology (GO) category and gene set	Gene set size	Direction	p value	q value
GO Biological process				
GO:0016071 mRNA metabolic process	750	Down	1.30E-07	0.000694
GO:0006402 mRNA catabolic process	330	Down	4.70E-07	0.000975
GO:0006396 RNA processing	884	Down	6.20E-07	0.000975
GO:0006412 translation	560	Down	7.33E-07	0.000975
GO:0022613 ribonucleoprotein complex biogenesis	423	Down	1.09E-06	0.000999
GO Cell component				
GO:1990904 ribonucleoprotein complex	637	Down	2.79E-10	0.000000194
GO:0005730 nucleolus	843	Down	0.000000324	0.000113
GO:0005840 ribosome	210	Down	0.000000945	0.000219
GO:0044391 ribosomal subunit	174	Down	0.00000153	0.000266
GO:0022626 cytosolic ribosome	99	Down	0.00000884	0.00123
GO Molecular function				
GO:0003723 RNA binding	1452	Down	2.20E-14	1.98E-11
GO:0003735 structural constituent of ribosome	151	Down	1.91E-06	0.000779
GO:0044877 protein-containing complex binding	836	Down	2.60E-06	0.000779
GO:0005198 structural molecule activity	374	Down	7.24E-06	0.00162
GO:0045296 cadherin binding	291	Down	9.66E-05	0.0174

5.5 Optimising RNA extraction from the Parsortix enriched cell population

Having established that Transfix BCTs produced a superior yield of CTCs in a HNSCC cell line model, and with the intention of performing gene expression profiling on enriched CTCs from patient samples, it was necessary to investigate if RNA extraction of the Parsortix enriched cell population would be compromised by the fixative agent used in the Transfix tubes. While the exact fixative agent used in Transfix tubes is not made public by the manufacturer, this is presumed to be a combination of low concentration paraformaldehyde (PFA) and other protein stabilising agents, similar to alternative cell preserving BCTs. As previously highlighted, aldehyde fixation results in cross-linkages between RNA (and DNA) molecules that can inhibit successful purification, thus preventing sequencing.[97]

The widely utilised Qiagen 'RNeasy micro kit' is designed to purify up to 45μg of RNA from < 500,000 cells or 5mg of tissue[190] but is *not* optimised for fixed samples. Conversely, the Qiagen 'RNeasy FFPE kit' is not optimised for low input samples. An alternative RNA extraction kit is the Thermofisher 'PicoPure'.[191] The Picopure kit is optimised for extraction of low volumes of RNA and anecdotally (from discussions with fellow researchers) is observed to perform better on fixed samples. Therefore, we compared the performance of the Qiagen and Thermofisher kits to extract RNA from Transfix preserved cells. The FaDu HNSCC cell line was used, at concentrations of 100, 1,000, 10,000 and 100,000 cells (in a fixed volume of 1ml buffer). A non-fixed control sample was compared to cells preserved in Transfix solution for <4hrs or 24hrs, so as to mimic real life clinical trial sample collection. Extracted RNA was quality assessed on an Agilent 2200 RNA tape station, to produce an RNA integrity (RIN) score and RNA concentration value. Results are summarised in Table 5.2. Both the Qiagen and

Thermofisher kits performed equally on control samples. RNA concentration for 100 cells input was poor for both kits. Across 1,000-100,000 cell input numbers, mean RNA integrity was 9.7 (9.1-9.6) and 9.1 (8.9-9.3) for the Qiagen and Thermofisher kits respectively; RNA concentration increased accordingly and was comparable between kits. In Transfix preserved samples the PicoPure kit outperformed the Qiagen kit, with higher RIN scores and RNA concentrations across cell input numbers. Cells that were kept in Transfix solution for 24hrs demonstrated decreased RIN scores and extracted RNA concentrations compared to near immediate processing (<4hrs). However, the benefit of the Quantseq library prep is that only a relatively small sequence from the 3' end is required for transcript counting, therefore it is an excellent sequencing technique to use on highly fragmented RNA (i.e. fixed/paraffin embedded samples).[192] Anecdotally our group has performed Quantseq on paraffin embedded tissue derived RNA with RIN values of ≈1, therefore we considered our RIN values of ≈3 from Transfix samples to be acceptable for subsequent sequencing experiments.

Table 5.2: The ThermoFisher PicoPure kit out performs the Qiagen RNeasy kit for RNA extraction from Transfix preserved samples. RIN values and RNA concentrations are displayed for extraction experiments from 100 – 100,000 cell input numbers (FaDu HNSCC cell line), comparing unfixed and Transfix preserved cells.

			Qiagen RNeasy		Thermofisher PicoPure	
Sample fixation	Timepoint	Cell count	RIN score	Concentration ng/ml	RIN score	Concentration ng/ml
Control unfixed	0hrs	100	2.5	0.106	7.2	0.216
		1,000	9.9	0.264	9	1.62
		10,000	9.6	1.82	9.3	1.39
		100,000	9.7	18.9	8.9	7.92
Transfix preserved	<4hrs	100	nil	nil	4.1	0.513
		1,000	2	0.0873	9.6	0.167
		10,000	7.7	0.526	9.7	9.42
		100,000	9.5	1.39	7.4	41.6
	24hrs	100	nil	nil	nil	nil
		1,000	3.9	0.431	1.3	0.278
		10,000	3.5	0.679	3.7	1.91
		100,000	3	0.262	7.8	2.9

5.6 Gene expression profiling of Parsortix enriched cells

Finally, we sought to demonstrate proof-concept to perform gene expression profiling on Transfix preserved and Parsortix enriched patient blood samples – both from an RNA extraction/sequencing technical viewpoint and data analysis strategy to see rare cell (CTC) specific transcripts from 'bulk' sequencing of the heterogenous Parsortix enriched population.

FaDu cell line spiked healthy donor (HD) Transfix blood samples were enriched using Parsortix (time-to-processing 24hrs). Transfix preserved matched HD PBMCs and FaDu cells were used as negative and positive controls respectively. We performed differential gene expression (DGE) analysis and specifically looked at our HNSCC/CTC marker gene list (as per figure 5.5) to compare the Parsortix-enriched spiked cell population and the PBMC control. Numerous epithelial cell (FaDu) specific transcripts - including EpCAM, E-cadherin (CDH1), and cytokeratins (KRT) 5,17,18 and 19 (figure 5.6a) – were highly expressed within the enriched spiked cell samples. These markers – routinely used for CTC identification in HNSCC – were also highly expressed in the FaDu positive control cells, thus demonstrating that FaDu specific transcripts were identifiable within the enriched population. In addition, druggable targets such as EGFR and PD-L1 (CD274), known to be expressed in the FaDu cells, were also able to be demonstrated as highly expressed compared to the PBMC control.

To demonstrate that the pattern of differentially expressed genes was originating from our spiked FaDu cells, we compared the DGE between the Parsortix-enriched (spiked PBMC) samples and PBMC controls and between the PBMC and FaDu control samples. With the hypothesis that a high correlation would confirm that not only the epithelial cell transcripts (identified above) but also other differentially expressed genes in the spiked sample were originating from our spiked in 'CTCs'. Such analysis was important to provide proof-of-concept that variance in non-epithelial transcripts could be reliably identified as being of CTC origin when bulk sequencing was performed. A high correlation was observed between the two datasets (Pearson R = 0.933, Spearman R= 0.934, $p<0.001$, figure 5.6b) indicating that HNSCC-specific gene expression might be successfully identified from CTCs using Transfix BCTs and Parsortix enrichment.

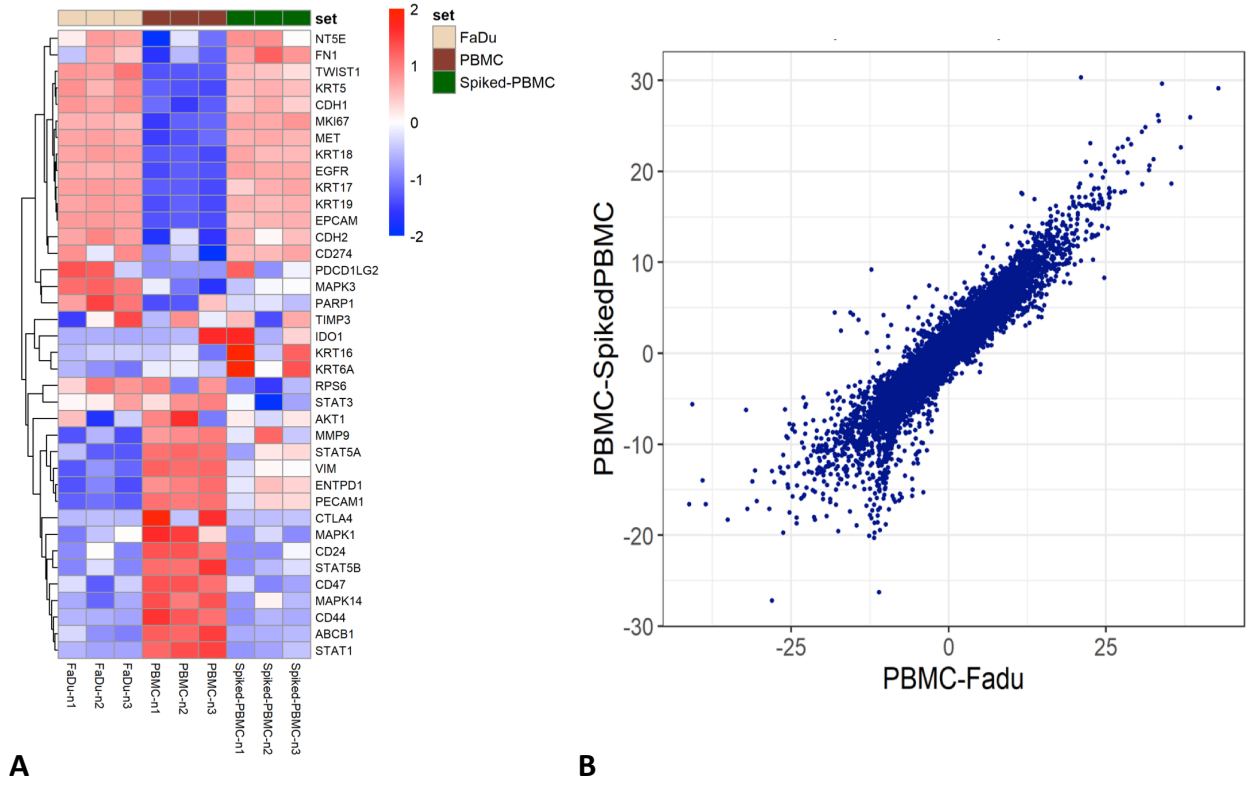


Figure 5.6: Proof-of-concept that a rare cell population (epithelial CTC) can be identified from bulk sequencing of Transfix-preserved, Parsortix enriched blood samples. Analysis of gene expression from 'bulk sequencing' of Transfix preserved and Parsortix enriched healthy donor blood samples spiked with 1000 FaDu cells (n=3). A) Heatmap to visualise expression of the 34 HNSCC/CTC target/marker genes in control samples (FaDu and PBMC) and enriched spiked cell sample. Note that epithelial and key HNSCC pathway genes are highly expressed in the 'spiked-PBMC' population with a similar expression pattern to the FaDu control cells, confirming the presence of enriched epithelial (FaDu) cells. B) Scatterplot of differentially expressed genes between the Parsortix enriched spiked PBMC samples and PBMC controls vs the FaDu and PBMC control samples. Axes denote 'stat' score as a measure of number of standard deviations of the significant difference in gene expression between samples i.e. negative values are downregulated genes and positive values upregulated genes. High correlation (Pearson R = 0.933, Spearman R= 0.934, $p<0.001$) confirms the presence of FaDu specific i.e. rare epithelial cell, transcripts detectable from the enriched rare cell population.

5.7 Summary of key findings

- *Collecting blood in Transfix BCTs increases Parsortix cell capture rate compared to EDTA BCTs and Cytodelics cell preservation*
- *Transfix preserved blood samples can be reliably stored for up to 72hrs prior to processing*
- *Parsortix microfluidic enrichment of unfixed cells induces a quiescent phenotype and cell stress response related to oxidative stress, with downregulation of certain cytokeratin proteins*
- *Extraction of RNA is possible from Transfix preserved cells and bulk RNA sequencing of Parsortix enriched cells allows identification of epithelial cell-specific transcripts*

CHAPTER 6: CHARACTERISING CTC SUB-POPULATIONS IN HNSCC USING IMMUNOFLOURESCENCE AND FLOW CYTOMETRY

6.1 Moving from *in-vitro* to *ex-vivo* CTC characterisation

Having established proof-of-concept for the utility of Parsortix to enrich CTCs in HNSCC, we progressed onto processing patient samples. The following chapter will describe the validation of Parsortix for enriching CTCs in HNSCC patient blood samples and subsequent downstream characterisation via two complementary approaches. First, using immunofluorescence microscopy. Second, using flow cytometry to accurately quantify CTCs and characterise sub-groups undergoing an epithelial-mesenchymal transition (EMT).

6.2 Aims and objectives

Our initial aim in this chapter was to visualise CTCs in HNSCC patient samples. Following initial validation, we then sought to optimise a method to quantify enriched CTCs and explore a means to characterise phenotypic sub-groups.

Objectives

1. *Optimise an antibody panel for analysis of Parsortix enriched Transfix preserved blood samples*
2. *Visualise CTCs in HNSCC and accurately quantify CTC burden*
3. *Investigate if CTC count and phenotypic groups correlate to clinicopathological and survival outcome variables*

6.3 Antibody optimisation

As discussed, the definition of a CTC is widely accepted to be a nucleated epithelial marker positive cell, that is negative for the pan-leukocyte marker CD45. With this in mind our antibody panel comprised of the epithelial marker EpCAM and mesenchymal marker N-cadherin, together with the pan-leukocyte marker CD45 (table 4.1). While several markers have previously been utilised to define epithelial and mesenchymal CTC phenotypes across a range of malignancies, we selected EpCAM and N-cadherin for this purpose because, as highlighted in the introduction, they had shown positive results in previous reports in HNSCC. Collectively these three markers would allow us to define 'epithelial CTCs' (EpCAM high, N-cadherin low, CD45 low) and 'EMT CTCs' (EpCAM high, N-cadherin high and CD45 low).

Antibody compatibility with Transfix preserved cells was assessed using flow cytometry of FaDu and CAL27 HNSCC cell lines (for EpCAM and N-cadherin binding) and healthy donor PBMCs for CD45 binding. We chose antibody clones with published evidence of efficacy on fixed cells, based on the logic that this would allow antibody binding to epitopes on Transfix

preserved cells. The EpCAM-specific antibody (clone 9C4) demonstrated excellent binding to both FaDu and CAL27 cell lines, with no evidence of decreased signal intensity between un-fixed and Transfix preserved cells (figure 6.1a). The CAL27 cell line is known to have a more mesenchymal phenotype than the FaDu cell line, with moderate expression of N-cadherin.[193] The N-cadherin-specific antibody (clone 8C11) demonstrated acceptable binding efficacy to Transfix preserved CAL27 cells, with a small decrease of 1.3% N-cadherin positive cell population size when comparing unfixed and Transfix preserved cells (figure 6.1b). Using PBMCs as a positive control, the CD45 clone 2D1 was also unaffected by Transfix fixation (figure 6.1c).

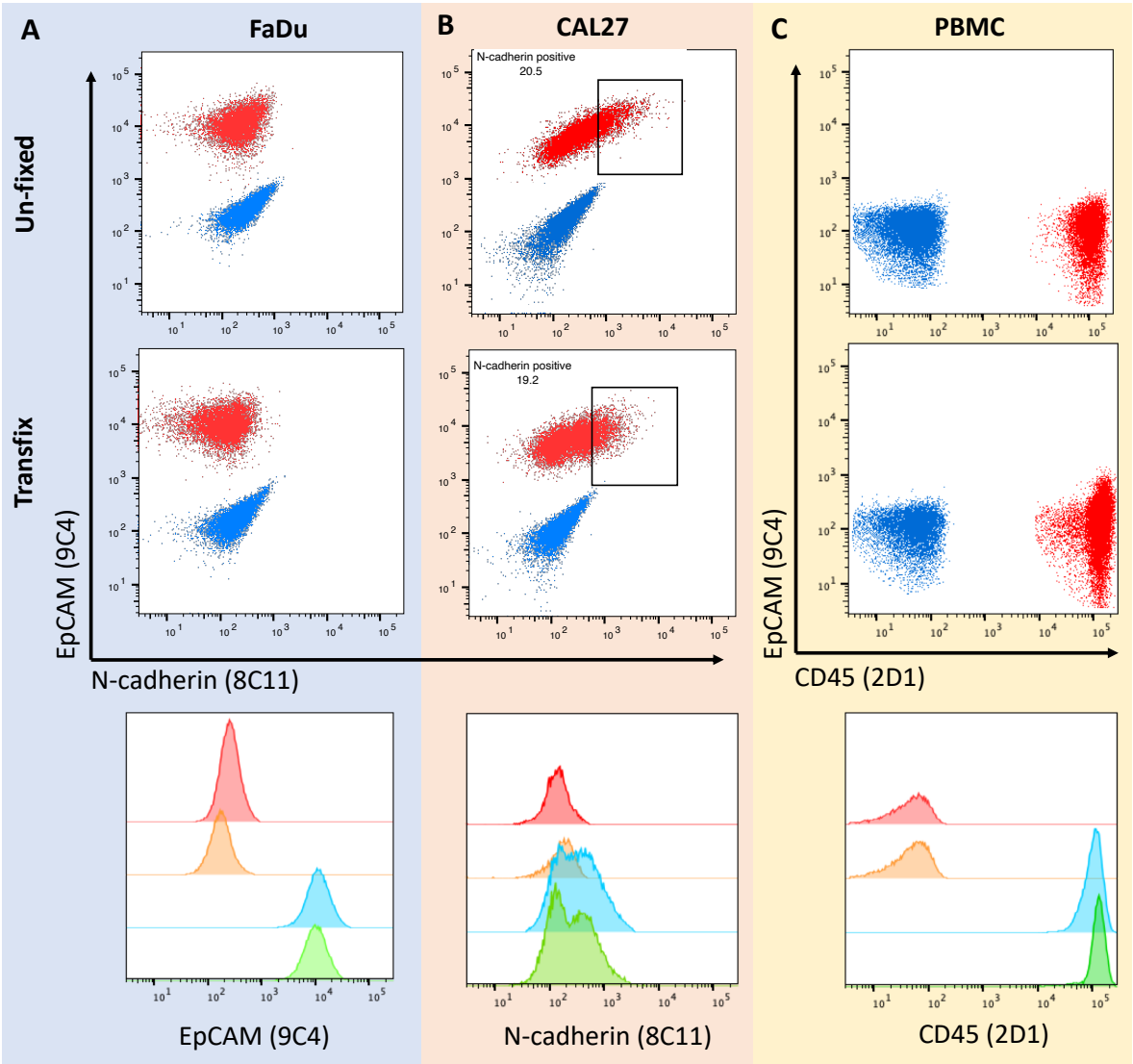


Figure 6.1: Testing antibody compatibility with Transfix fixed HNSCC cell lines. EpCAM, N-cadherin and CD45-specific antibody clones were tested against three different cell types: left column, FaDu; middle column, CAL27; right column, PBMCs. The bi-axial plots in rows one and two show the results for negative-control unstained cells (blue) and antibody-stained cells (red) with each plot showing unfixed or Transfix fixed cells as indicated on the figure. The bottom row of histogram plots displays intensity values for each antibody to four populations of cells as follows: red – unfixed unstained, orange – Transfix unstained, blue – unfixed stained, green – Transfix stained). Overall the three antibody clones were unaffected by Transfix fixation.

6.4 Visualisation of Parsortix enriched CTCs stained with fluorophore labelled antibodies

Having identified antibody clones able to effectively stain Transfix preserved cells, we sought to optimise a method of visualising fluorophore labelled Parsortix enriched cells. The Parsortix system allows enriched cells to be stained within the cassette which can then be temporarily removed from the machine to be inspected using a separate immunofluorescent microscope. The advantage of this 'closed circuit' analysis is that it does not disturb the cells. The technique was first trialled using whole blood from a healthy donor that was spiked with 100 cells/ml of the FaDu HNSCC cell line then added to a Transfix BCT, thus mimicking a CTC-containing patient blood sample. While we were able to detect the spiked epithelial cells, the quality of images was suboptimal (figure 6.2). The double walled structure and thickness of the cassette created adverse light refraction and a poorly focused image. In addition, due to the tiered design of the cassette, cells are captured at multiple levels and focusing on several cells within the same image frame was challenging. Different microscopes, including the EVOS digital colour fluorescent microscope and the Leica DMI6000B inverted confocal microscope

were trialled to enhance the image quality. The latter was an improvement, as seen in the images from a negative control healthy donor blood sample (figure 6.3), however the focusing of cells at higher magnification remained difficult. Due to the cassette thickness, mounting the cassette in a standard upright confocal microscope was not possible.

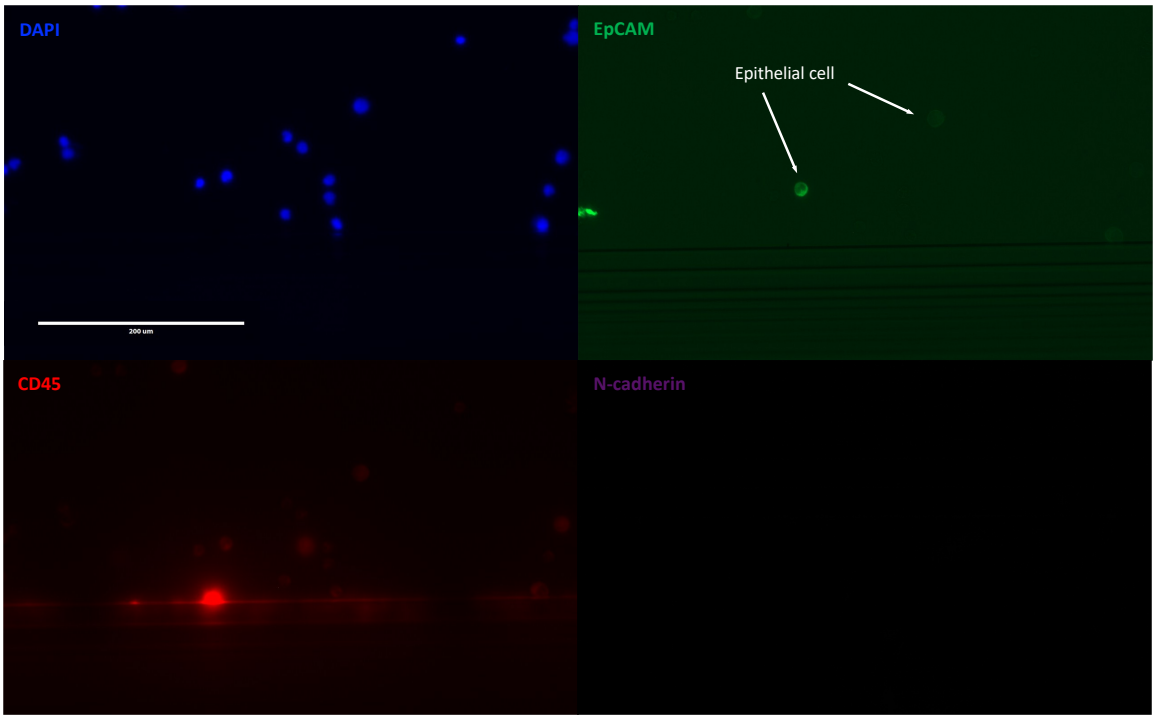


Figure 6.2: Enriched HNSCC spiked cells are detectable but poorly visible within the enrichment cassette. 20x magnification images taken using an EVOS fluorescence microscope (Invitrogen) of Parsortix enriched spiked FaDu cells in healthy donor blood (in Transfix). Epithelial cells are visible within the cassette as EpCAM high, DAPI high, CD45 low. Of note, the quality of images was poor and focusing on multiple cells (epithelial cells and leukocytes together) was difficult.

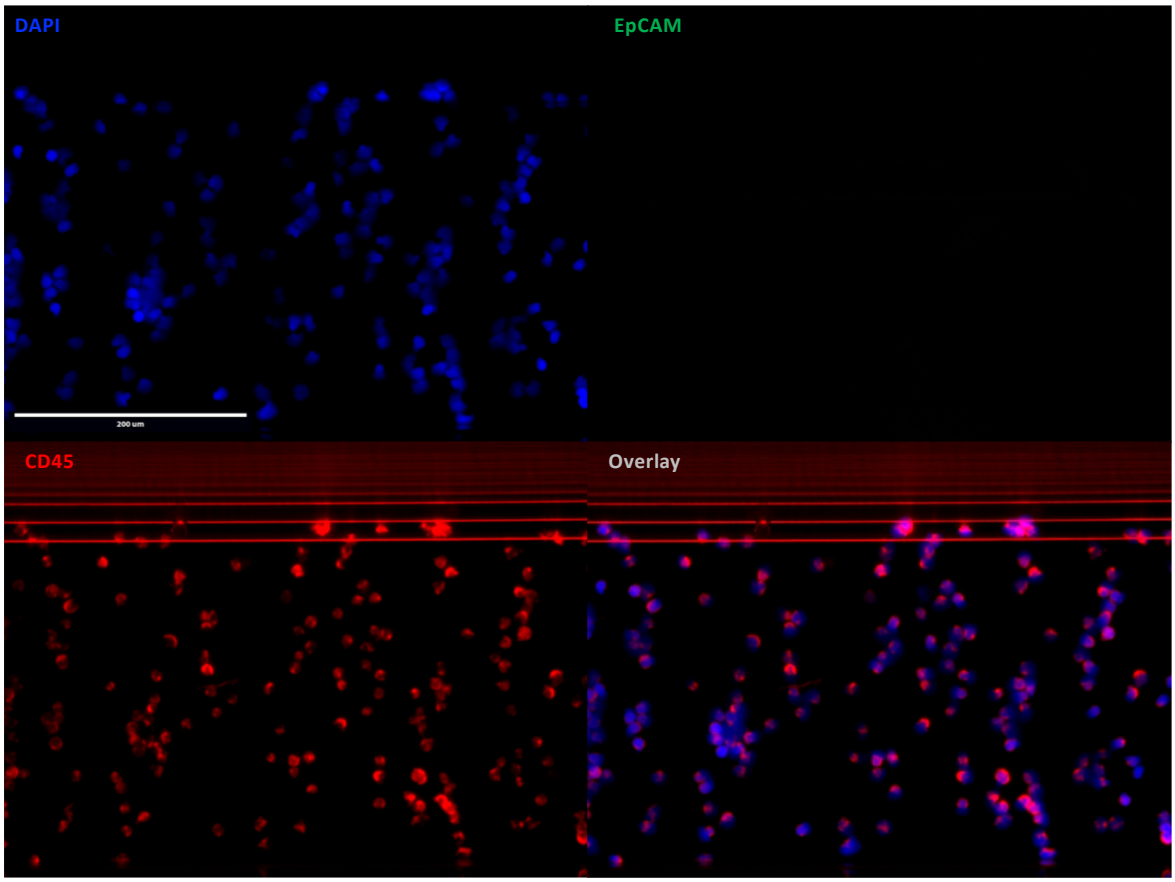


Figure 6.3: Parsortix enriched healthy donor sample does not demonstrate EpCAM high cells. 20x magnification of healthy donor blood in Transfix blood collection tube, enriched using Parsortix. Changing the microscope to an inverted confocal microscope (Leica DMI6000B) allowed an improved quality of image, however we were still unable to focus within the cassette at higher magnifications.

To combat the issue of 'in-cassette' microscope cell visualisation we decided to trial transferring the enriched cells from the cassette onto glass slides, thus enabling the use of a conventional upright confocal microscope (section 4.7). The results of slide mounting cells were significantly improved compared to in-cassette viewing, enabling images to be obtained at a higher magnification using a Leica DMI6000B microscope (figures 6.4 and 6.5).

Having optimised the antibody labelling and visualisation steps, a proof-of-principle experiment was undertaken to demonstrate that the Parsortix system could successfully

identify CTCs in blood samples from HNSCC patients. Pre-treatment blood samples from two HNSCC patients with high disease burden were used to provide the best chance of CTCs being present. The first patient had a T4 laryngeal SCC recurrence with neck metastases, the second had a T4aN2b primary oral SCC. Cells meeting the criteria for CTCs, including epithelial and EMT CTCs, were identified in both samples (table 6.1, figure 6.4 panel a and b). At higher magnification considerable heterogeneity of CTC morphology was observed – with some cells demonstrating a granular type appearance and a larger nucleus to cytoplasmic ratio. Also present were entities that appeared to be either cells having undergone cell lysis during the enrichment protocol, or microparticles of epithelial cells attached to nucleus material (figure 6.5a-d). Expression of N-cadherin on EMT CTCs was generally weak (figure 6.5c), which may be expected given that these are partially transitioned cells. N-cadherin positive, EpCAM negative, CD45 negative cells (i.e. fully transitioned mesenchymal CTCs) were not seen.

Table 6.1: Immunofluorescence microscopy CTC characterisation test patients. Clinical and demographic data and observed CTC phenotypes from two immunofluorescent test cases.

PT. ID	AGE/ SEX	SITE	STAGE	EPCAM+ CTC (DAPI+, CD45-)	EPCAM+/ N-CADHERIN+ CTC (DAPI+, CD45-)
TEST 1	77 M	Laryngeal (recurrence)	T4	+	+
TEST 2	81 M	Oral	T4a N2b M0	+	-

In addition, in the patient with recurrent disease we observed a single CTC cluster of 5 epithelial (EpCAM-positive, N-cadherin-negative) cells (figure 6.5e). In the same patient a strange arrangement of cells was seen in close proximity or clustered together, that were

CD45 positive and N-cadherin negative but with one cell also showing weak EpCAM expression (figure 6.5f). An unexpected finding was the presence of a small number of cells with positive DAPI staining but negative for all three markers (EPCAM, N-cadherin and CD45 negative; a single example is shown in figure 6.4 panel b, 'DAPI+'). The identity of these cells was not clear.

In summary, while immunofluorescence staining of cells was variable we were able to positively identify cells in patient samples that met the criteria for CTCs. The heterogeneity of CTC epithelial marker staining, for example moderate or high EpCAM expression, was a notable observation and has been observed by other groups in HNSCC who describe CTCs with variable epithelial maker expression.[130] In particular, positive control (spiked in epithelial cell line) CTCs demonstrated reliable staining, which built confidence in results from patient samples that such variability was biological and not technical. This issue of technical versus biological staining variation is further discussed in chapter 8.

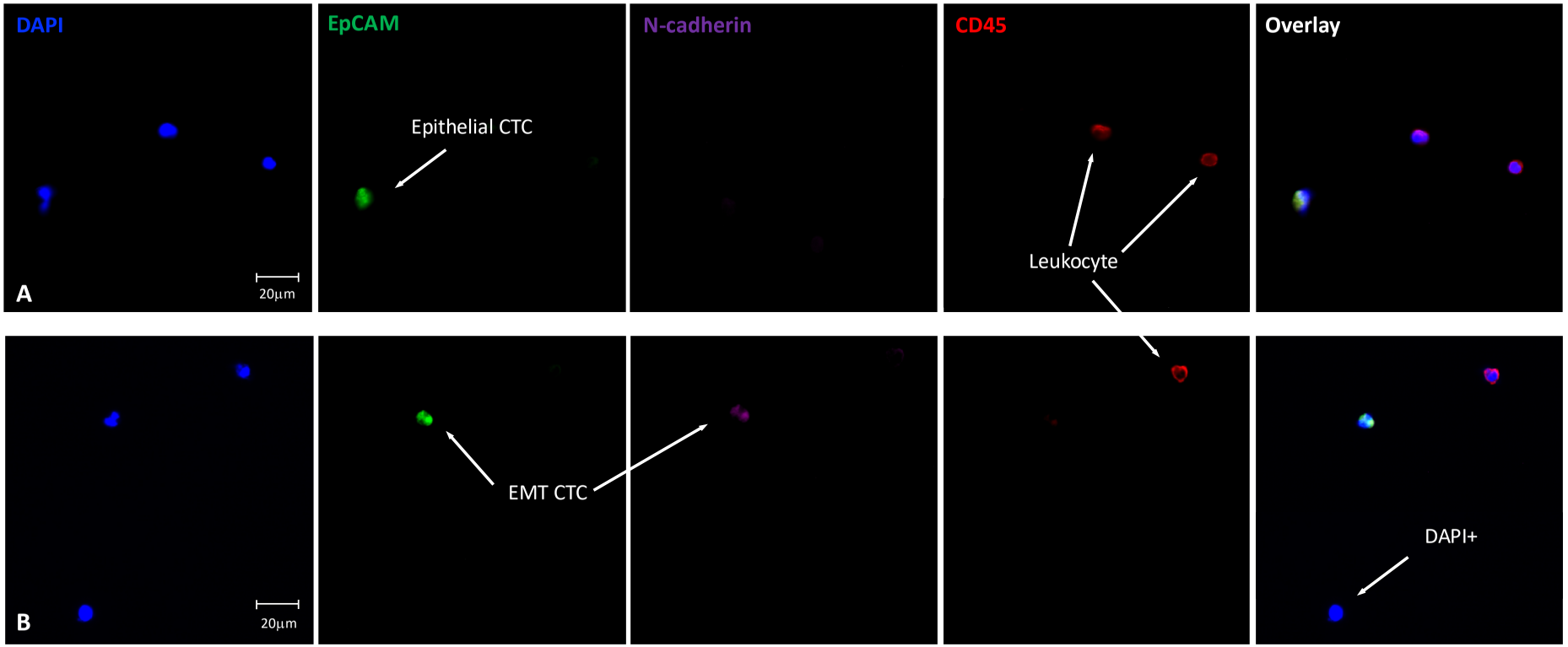


Figure 6.4: Parsortix enriched cells with characteristics of epithelial and EMT CTCs. Representative images of slide mounted cells enriched from a single patient sample (patient ID: test 1, 20x magnification). Panel A shows cells meeting the criteria for epithelial CTC (DAPI positive, EpCAM positive, CD45 negative). Panel B shows an example of an EMT CTC (DAPI positive, EpCAM positive, N-cadherin positive, CD45 negative) and a DAPI positive cell lacking EpCAM, N-Cadherin and CD45 (located bottom left of image, labelled 'DAPI+')

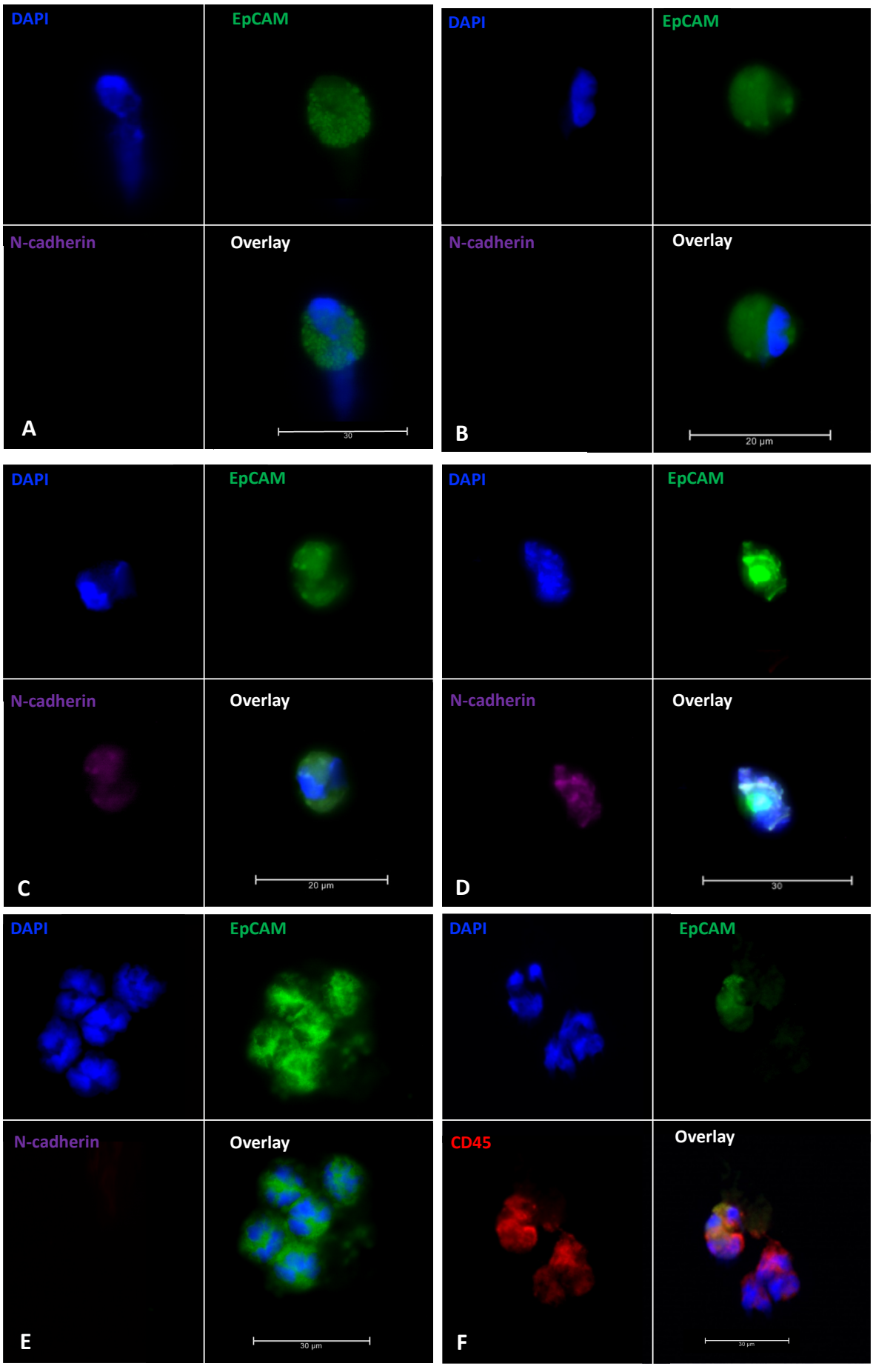

DAPI
EpCAM
N-cadherin
Overlay
A
30
DAPI
EpCAM
N-cadherin
Overlay
B
20 µm
DAPI
EpCAM
N-cadherin
Overlay
C
20 µm
DAPI
EpCAM
N-cadherin
Overlay
D
30
DAPI
EpCAM
N-cadherin
Overlay
E
30 µm
DAPI
EpCAM
CD45
Overlay
F
30 µm

Figure 6.5: High magnification images of CTCs reveals differing morphology and CTC clusters. Higher magnification immunofluorescence images from the same patient sample shown in Figure 6.4 (patient ID: test 1, 100x magnification). Note that cells in boxes a-e are all CD45 negative and cells in box f are n-cadherin negative (to aid clarity these negative stains are not shown). CTCs are present with epithelial (a,b) and EMT (c,d) expression. Of note is the granular appearance of EpCAM staining in CTCs. In addition, the presence of clusters (e,f) in a patient with recurrent disease.

6.5 Flow cytometry characterisation of CTCs

Once we had demonstrated proof-of-principle in a test cohort that Parsortix could enrich CTCs and that the antibody panel was optimised for Transfix fixed cells, we proceeded to gather data from a larger patient cohort. While direct visualisation of CTCs using immunofluorescent microscopy was able to confirm the presence of CTCs, this method was not feasible for analysing the total Parsortix enriched population from each patient (this would require 20 slides per patient to be manually screened). Analysis was therefore transferred to flow cytometry to enable reproducible and accurate quantification of CTCs per ml of whole blood collected from each patient.

6.5.1 Gating strategy

To optimise a flow cytometry gating strategy for Parsortix enriched blood samples a control sample was prepared. This comprised whole blood spiked with a small number of cells from the FaDu HNSCC cell line. This preparation was then added to a Transfix tube to mimic a

patient blood sample. Standard flow cytometry data clean-up involved gating single events on forward scatter (FSC) area vs FSC height plots, then gating on side scatter (SSC) and FSC to gate out low signal debris (figure 6.6a). Of note was the difficulty in setting a singlet gate in a heterogenous population of 2 cell types (FaDu and PBMCs). We therefore set a tight gate on FSC vs SSC plots to decrease the potential of including doublet events. EpCAM and N-cadherin gating is described in figure 6.6 and figure 6.1.

Gating on CD45 vs EpCAM clearly demonstrated the ability to separate out spiked epithelial cells and CD45 high PBMCs. Within the Parsortix enriched sample a population of EPCAM high and CD45 high events was also observed (figure 6.6a). This was examined this further using FaDu and donor PBMC positive controls (figure 6.6b). The 'double positive' population was not present in either the FaDu positive control or PBMC control – but it was evident when a mixture of these 2 controls was run. Therefore, the assumption was that these were false positive doublets, present because of the difficulty in defining a singlet gate with 2 differing cell populations (figure 6.6b). Nonetheless, this test demonstrated a successful protocol and gating strategy to be able to identify and quantify epithelial cells (i.e. CTCs) using Parsortix and flow cytometry. We were satisfied that when running patient samples, the PBMC population enriched by Parsortix would serve as an internal negative control with which to set a gate to define CD45 low and EpCAM high events. In addition to PBMCs, the Parsortix enriched fraction contained CD45 positive cells of similar size and granularity as the spiked FaDu cells, i.e. a CTC mimic. These were provisionally identified as granulocytes that were a minor population in the PBMCs but were being enriched alongside the epithelial cells. A finding noted for future marker panel design.

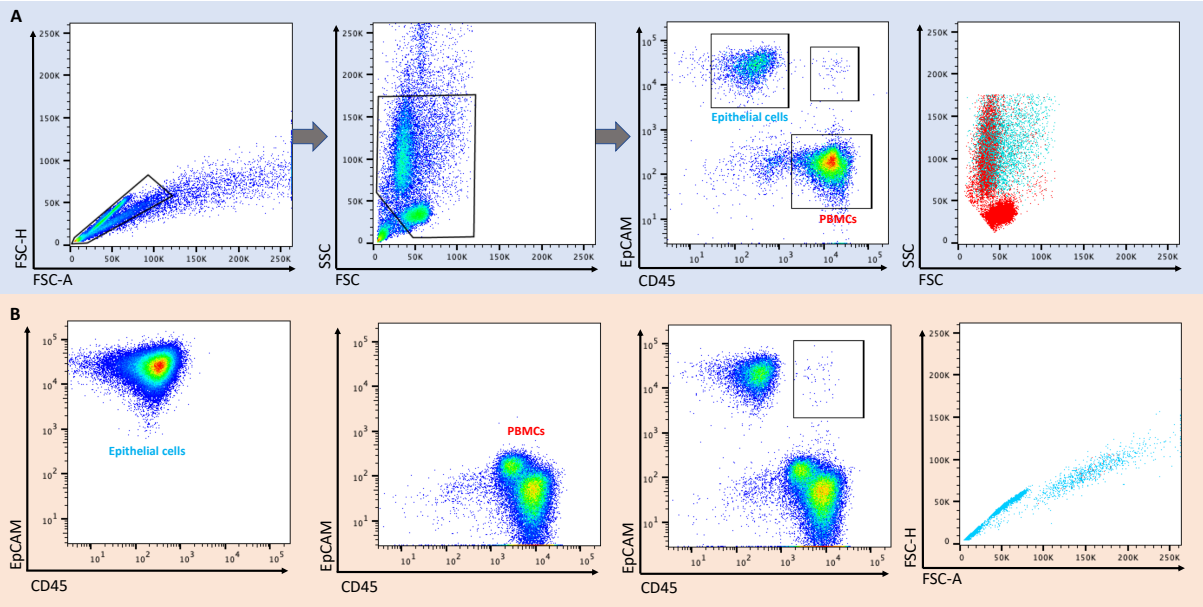


Figure 6.6: Determining gating strategy to identify enriched epithelial cells. A) FaDu spiked donor blood enriched with Parsortix. From left to right: singlet gating, gating on non-debris cells, identifying epithelial and PBMC cell populations and backgating to determine position on scatter plots. The unlabelled square in plot 3 is a doublet population, further explored in panel B. B) Data from FaDu and PBMC positive controls to determine the origin of 'double positive' cell population identified in spiking experiment. When the cells are combined together the double positive population appears and were therefore assumed to be doublet artefact as seen from final FSC plot (doublets in red).

6.5.2 Patient cohort

A cohort of 20 HNSCC patients were recruited prospectively (methods 4.10). Blood samples (9ml) were collected into Transfix BCTs, at the time of surgery, and processed as per methods section 4.6. Baseline clinicopathological data and unadjusted CTC count from the 9ml blood samples is shown in table 6.2. CTC count was derived from flow cytometry gating, as detailed in figure 6.7. Mean patient age was 65.6 years (range 24-84) with 65% (n=13) male and 35%

(n=7) female. The majority of patients (13/20) presented with advanced (stage III/IV) disease, with an 'oral' anatomical subsite predominance (11/20). An example of gating on two different patient samples to quantify CTC phenotypic sub-groups is shown in figure 6.7. The patient sample in figure 6.7 panel a, demonstrates an EpCAM high, CD45 low population with N-cadherin expression in a subset of these cells, demonstrating an EMT sub-group. In contrast, the patient sample from figure 6.7 panel b, is an EpCAM high, cD45 low population with no N-cadherin expression.

Table 6.2: Patient demographic and clinical data with CTC count. Total count of epithelial or EMT CTCs identified from 9ml Transfix BCT is shown (S = surgery, CRT = chemoradiotherapy, RT = radiotherapy, P = palliative treatment).

PT NO.	AGE	SEX	T STAGE	N STAGE	STAGE GROUP	TUMOUR SITE	TREATMENT (PRIMARY/ PALLIATIVE)	EPI-CTC	EMT CTC	TOTAL CTC	STATUS AT TIME OF CENSORING
1	68	F	4	2	IV	oral	S + CRT / P-RT	23	19	42	Death due to cancer
2	72	M	2	1	III	hypo-pharynx	P	35	16	51	Death due to cancer
3	62	M	x	2	III	unknown	CRT	12	11	23	Alive and well
4	62	M	2	0	II	oral	S	0	0	0	Alive and well
5	69	F	3	0	III	oral	S	15	7	22	Alive and well
6	67	M	2	0	II	oropharynx	S	36	4	40	Alive and well
7	57	F	2	0	II	oral	S / P-CRT	5	0	5	Alive with recurrence
8	50	M	4	2	IV	oropharynx	CRT	14	11	25	Alive and well
9	73	F	4	0	IV	oral	S + CRT	62	39	101	Alive and well
10	64	F	4	2	IV	oral	S + CRT / P-RT	26	10	36	Death due to cancer
11	68	M	2	0	II	larynx	S	22	0	22	Alive and well
12	25	F	1	0	I	oral	S	4	0	4	Alive and well
13	84	F	4	0	IV	oral	S	34	0	34	Death, cause unknown
14	61	M	2	0	II	oral	S	5	0	5	Alive and well

15	79	M	4	0	IV	larynx	S	3	2	5	Alive and well
16	54	M	4	3	IV	larynx	P	2	0	2	Death due to cancer
17	65	M	2	0	II	oral	S	1	0	1	Death, other cause
18	76	M	4	2	IV	larynx	S	16	9	25	Death, cause unknown
19	58	M	4	0	IV	oral recurrence	S	21	13	34	Alive and well
20	79	M	2	2	III	oral	S	18	0	18	Death due to cancer

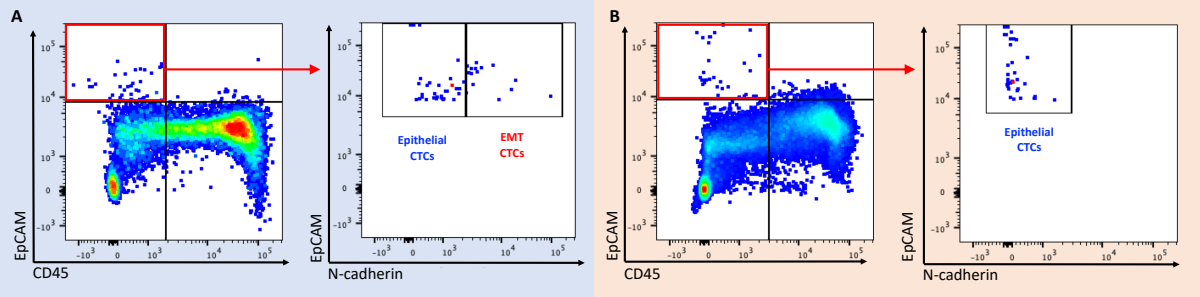


Figure 6.7 Gating strategy for identification of epithelial and mesenchymal CTCs. Flow cytometric data for samples, from 2 different patients post-Parsortix enrichment. Both samples have undergone clean-up gating as detailed in figure 6.6. A) PBMC population is used as a negative control to define boundaries for quadrant gate and identify 'EpCAM high, CD45 low' CTCs. CTC population in red box in plotted for EpCAM vs N-cadherin to identify epithelial and EMT CTC phenotypic subgroups (gate set using positive control compensation bead data). B) Same gating strategy, but CTCs are exclusively epithelial with low mesenchymal marker expression.

Defining a confidence threshold

In parallel, blood samples from five healthy donor (HD) controls underwent the same CTC enrichment protocol, staining, flow cytometry and data analysis procedures to define a confidence threshold for positive events. For three of the five HDs EpCAM high and CD45 low events were detected albeit at a low frequency – 6, 8 and 9 events per sample. Such a finding was expected, given the difficulty in clean-up gating experienced in our spiked test case and the potentially subjective nature of flow cytometry target cell manual gating. At this level of

events the accuracy of gating to define target populations is significantly reduced and the margin of error to include debris, doublets or false positive events is high. We set a minimum threshold value as the maximum of these 3 false positive results. Thus, >9 events per 9ml blood tube was used to define the presence of CTCs, which equated to a value of >1 CTC/ml of whole blood. Based on this threshold we redefined our dataset to exclude values from samples with ≤9 events and presented data as CTCs enriched per ml of whole blood, the standard metric (figure 6.8). Reassuringly, results from patient samples fell into two broad categories – those with clear evidence of CTCs (a total CTC count of at least 20 cells per 9ml blood sample) and those with very few cells (5 or less cells per 9ml blood sample). As such, none of the patients were close to the defined minimum threshold of 9 events. Among CTC positive patients, adjusted CTC count per ml of enriched blood was 4 CTCs/ml blood (range 2 - 11.2) (figure 6.8).

Despite previous reports describing N-cadherin positive, CD45 negative cells as 'mesenchymal CTCs'[106] we observed events meeting this criteria in 2 of our 5 healthy donor blood samples. In our data analysis, we used the combination of epithelial *and* mesenchymal marker positivity to confidently describe an EMT CTC (i.e. mesenchymal *transitioning*, but not fully *transitioned* cells). While the criteria governing the positive identification of a CTC may be evolving, we exercised caution when assigning the label of a 'CTC' to a cell with no epithelial expression and based solely on the expression of one mesenchymal marker that is also known to be expressed at low levels in certain immune cells (e.g granulocytes can express N-cadherin and are CD45 low[194]). Such a dilemma in CTC characterisation clearly demonstrates why multi-parameter characterisation is required – to be able to confidently identify multiple markers from a single phenotype builds confidence in CTC identification and characterisation

(as will be discussed using mass cytometry in chapter 7). As has been highlighted by other groups, defining a 'pure mesenchymal CTC' is very difficult.[195] Instead, we are able to confidently identify 'mesenchymal characteristics' of positively identified CTCs - both through the identification of mesenchymal markers and transcription factors, but also proliferative/stemness markers such CD44 expression which are seen in mesenchymal behaving cells.[195]

6.5.3 Correlating CTC count to tumour pathological variables

Comparing the presence of a CTC phenotypic sub-group to clinical data demonstrated some notable trends. Five of the seven patients with no CTCs present were early stage (I/II) disease and the presence of CTCs of any phenotype was significantly associated with advanced stage (III/IV) disease (figure 6.8, chi-square, p=0.0121). Furthermore, EMT CTCs were almost exclusively observed in advanced stage disease patients (9/10). The presence of CTCs (of either phenotype) was not associated with T (1/2 vs 3/4) or N (negative vs positive) status (figure 6.8, chi-square, p=0.109 and 0.0850 respectively). Multivariate analysis was performed using Poisson regression – modelling CTC phenotype count (epithelial and EMT) against T and N status. Neither epithelial nor EMT CTC count could be fitted into a prediction model for T stage (p=0.882 and 0.418 respectively) or N stage (p=0.280 and 0.224 respectively).

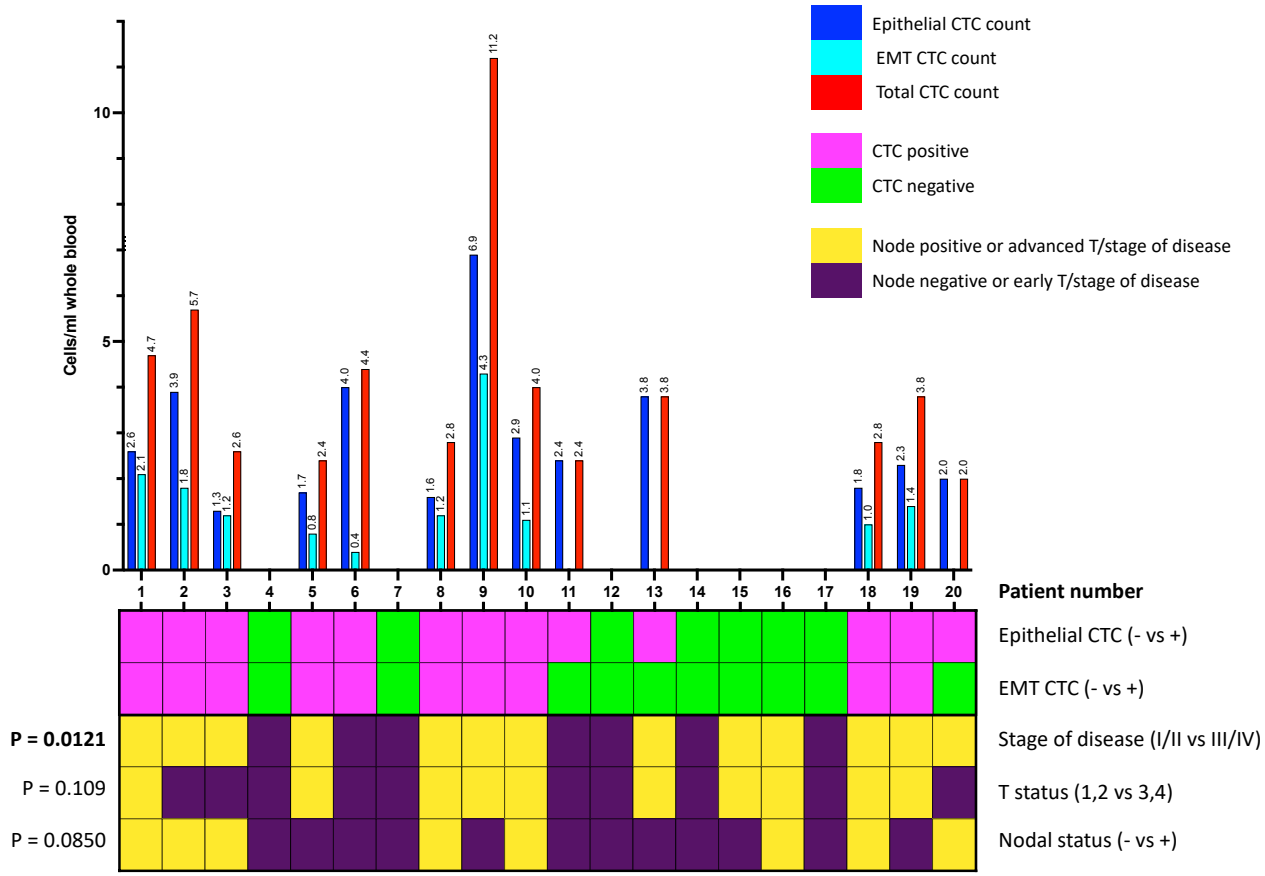


Figure 6.8: CTC count as cells per ml of whole blood and correlation to stage of disease. Bar chart (above) displays number of epithelial, EMT and total CTCs (per ml of enriched whole blood) for each patient. Heatmap (below) correlates the presence of epithelial and/or EMT CTCs to T, N and overall stage of disease, with significant association with late stage of disease (chi-square derived p-values).

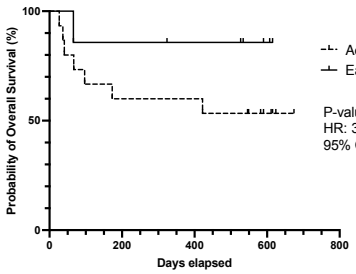
Next, the potential correlation between EMT CTCs and total CTC count was explored. *In other words*, is a higher CTC burden associated with increased transition of epithelial CTCs towards a mesenchymal phenotype. The ratio of EMT to epithelial CTCs was calculated and compared to total CTC count. These values were significantly correlated (Pearson's R= 0.564, P-value = 0.00958) thus, as the total CTC count increased so did the fraction of EMT CTCs within the entire CTC population. In all patients with EMT CTCs (n=10), epithelial CTCs were also present.

6.5.4 Baseline CTC presence as a prognostic biomarker of clinical outcome

Median patient follow-up time (section 4.13) was 586 days (range 324 – 674). Of the twenty patients within the cohort, at the time of censoring - eight patients had died (5 of which died due to disease progression and 3 due to unknown or other causes) and one patient was alive but with recurrence. In a small pilot cohort of 20 patients, firm conclusions with regard clinical outcomes cannot be made. However, analysis has been performed to demonstrate trends that would warrant further investigation.

Regarding tumour pathological variables, advanced stage of disease (III/IV) demonstrated a clear trend for decreased probability of overall survival (OS) and progression-free survival (PFS), albeit not significant (figure 6.9 a and b). The T status of the tumour (comparing 1,2 vs 3,4) did not correlate with OS or PFS (figure 6.9 c and d). However, the presence of nodal metastases was significantly associated with decreased OS and PFS (p= 0.0190 and 0.0219 respectively) (figure 6.9 e and f).

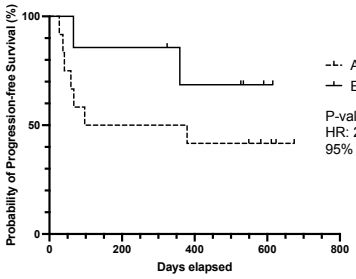
A



Number at risk

Days elapsed	Advanced Stage	Early Stage
0.000	15	7
27.000	15	
37.000	14	
41.000	13	
66.000		7
67.000	12	
97.000	11	
173.000	10	
324.000		6
422.000	9	
527.000		5
534.000		4
546.000	8	
549.000	7	
582.000	6	
590.000	5	3
607.000		2
611.000	4	
615.000	3	1
624.000	2	
674.000	1	

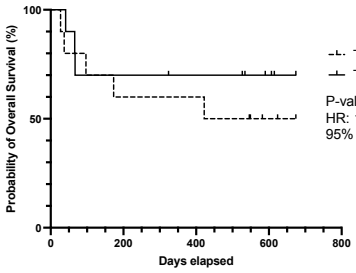
B



Number at risk

Days elapsed	Advanced stage	Early stage
0.000	12	7
27.000	12	
37.000	11	
41.000	10	
59.000	9	
66.000		7
67.000	8	
97.000	7	
324.000		6
359.000		5
379.000	6	
527.000		4
534.000		3
549.000	5	
582.000	4	
590.000		2
611.000	3	
615.000		1
624.000	2	
674.000	1	

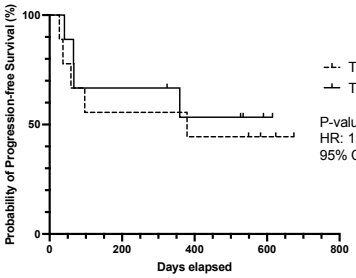
C



Number at risk

Days elapsed	T3/T4	T1/T2
0.000	10	10
27.000	10	
37.000	9	
41.000		10
66.000		9
67.000		8
97.000	8	
173.000	7	
324.000		7
422.000	6	
527.000		6
534.000		5
546.000	5	
549.000	4	
582.000	3	
590.000		4
607.000		3
615.000		2
624.000	2	
674.000	1	1

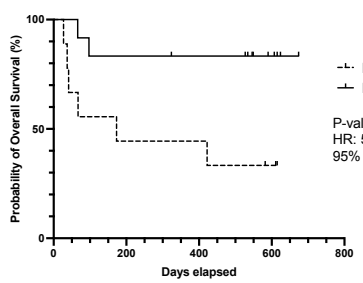
D



Number at risk

Days elapsed	T3/T4	T1/T2
0.000	9	9
27.000	9	
37.000	8	
41.000		9
59.000	7	
66.000		8
67.000		7
97.000	6	
324.000		6
359.000		5
379.000	5	
527.000		4
534.000		3
549.000	4	
582.000	3	
590.000		2
615.000		1
624.000	2	
674.000	1	

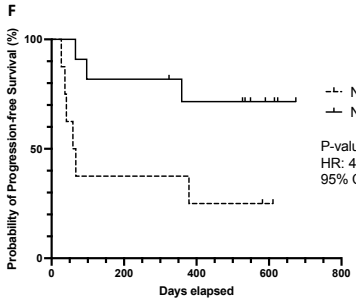
E



Number at risk

Days elapsed	Node positive	Node negative
0.000	9	12
27.000	9	
37.000	8	
41.000	7	
66.000		12
67.000	6	
97.000		11
173.000	5	
324.000		10
422.000	4	
527.000		9
534.000		8
546.000		7
549.000		6
582.000	3	
590.000		5
607.000		4
611.000	2	
615.000	1	3
624.000		2
674.000		1

F

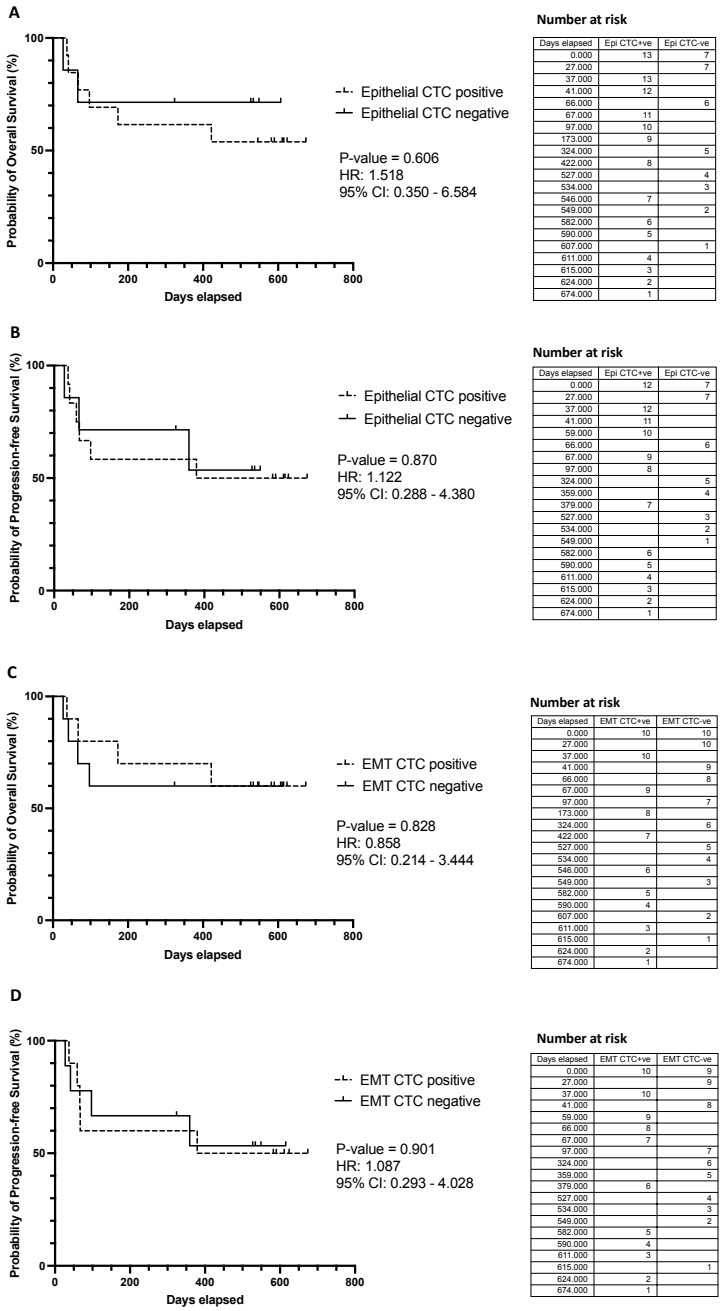


Number at risk

Days elapsed	Node positive	Node negative
0.000	8	11
27.000	8	
37.000	7	
41.000	6	
59.000	5	
66.000		11
67.000	4	
97.000		10
324.000		9
359.000		8
379.000	3	
527.000		7
534.000		6
549.000		5
582.000	2	
590.000		4
611.000	1	
615.000		3
624.000		2
674.000		1

Figure 6.9: Survival outcomes correlated to tumour pathological variables. Kaplan-Meier plots of overall (OS) and progression-free survival (PFS) for overall stage of disease (a,b), T status (c,d) and node positive disease (e,f). Log-rank (Mantel-Cox) derived p-value and hazard ratio reported. Number at risk data tables are also displayed. Advanced stage of disease trended towards poorer OS and PFS. Node positive patients demonstrated significantly decreased OS and PFS.

When examining the prognostic utility of epithelial and EMT CTCs being present at baseline, while the presence of epithelial CTCs demonstrated a decreased probability of OS (71% vs 54%) this was not to significant levels. Neither the presence of epithelial nor EMT CTCs correlated significantly with OS nor PFS (figure 6.10). We were unable to determine a cut-off for CTC count with prognostic significance for OS or PFS.



Days elapsed	Epi CTC+ve	Epi CTC-ve
0.000	13	7
27.000		7
37.000	13	
41.000	12	
66.000		6
67.000	11	
97.000	10	
173.000	9	
324.000		5
422.000	8	
527.000		4
534.000		3
546.000	7	
549.000		2
582.000	6	
590.000	5	
607.000		1
611.000	4	
615.000	3	
624.000	2	
674.000	1	

Days elapsed	Epi CTC+ve	Epi CTC-ve
0.000	12	7
27.000		7
37.000	12	
41.000	11	
59.000	10	
66.000		6
67.000	9	
97.000	8	
324.000		5
359.000		4
379.000	7	
527.000		3
534.000		2
549.000		1
582.000	6	
590.000	5	
611.000	4	
615.000	3	
624.000	2	
674.000	1	

Days elapsed	EMT CTC+ve	EMT CTC-ve
0.000	10	10
27.000		10
37.000	10	
41.000		9
66.000		8
67.000	9	
97.000		7
173.000	8	
324.000		6
422.000	7	
527.000		5
534.000		4
546.000	6	
549.000		3
582.000	5	
590.000	4	
607.000		2
611.000	3	
615.000		1
624.000	2	
674.000	1	

Days elapsed	EMT CTC+ve	EMT CTC-ve
0.000	10	9
27.000		9
37.000	10	
41.000		8
59.000	9	
66.000	8	
67.000	7	
97.000		7
324.000		6
359.000		5
379.000	6	
527.000		4
534.000		3
549.000		2
582.000	5	
590.000	4	
611.000	3	
615.000		1
624.000	2	
674.000	1	

Figure 6.10: Correlation between CTC positivity and survival outcomes. Kaplan-Meier plots of overall (OS) and progression-free survival (PFS) for epithelial (a,b) and EMT (c,d) CTC positive groups. Log-rank (Mantel-Cox) derived p-value and hazard ratio reported. Number at risk data tables are also displayed. Neither epithelial nor EMT CTC presence correlated with OS or PFS.

6.5.5 Longitudinal patient samples

To evaluate the dynamic change in CTCs in response to treatment we aimed to collect longitudinal blood samples at 6 weeks, 3 months and 12 months after completion of treatment. Unfortunately, due to the ongoing global pandemic closing clinical trials and halting sample collection the availability of samples was severely limited. Post-treatment samples at 12 months were collected from patient numbers 3, 6 and 7 and a 6 week sample was also collected from patient number 6 (figure 6.11).

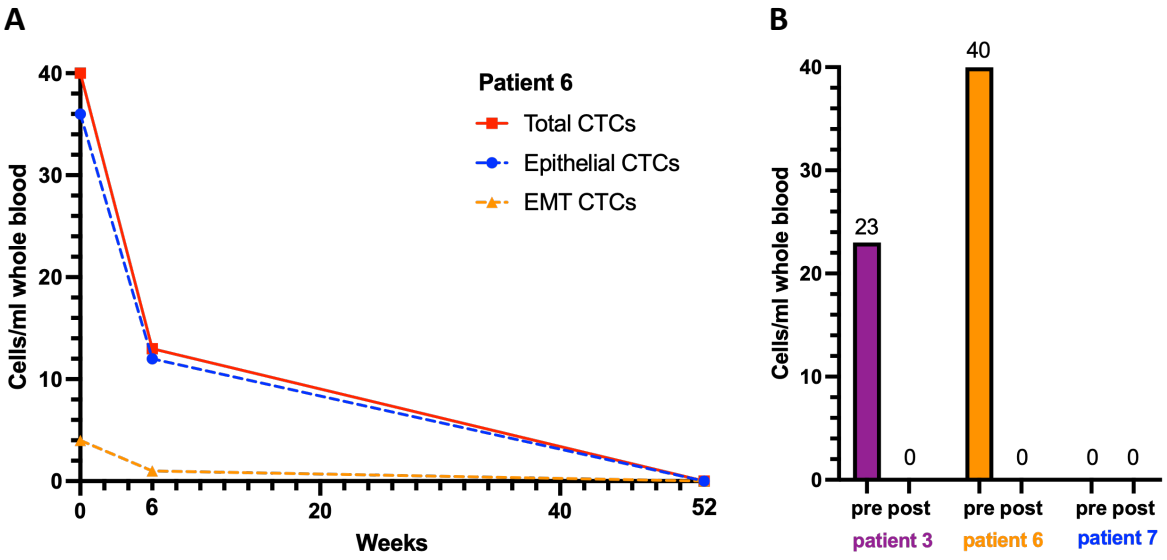


Figure 6.11: Post treatment CTC levels decrease in patients responding to treatment. A) Baseline and post-treatment CTC counts observed in patient 6. CTC burden decreases markedly at 6 weeks post-treatment and CTCs are not present at 12 months. B) Comparison of pre-treatment and post-treatment (12 months) CTC counts in three treatment responders.

In patient 6 there was a stepwise drop in CTCs post-treatment - CTCs were not cleared at 6 weeks post-treatment, but went on to record as zero at 12 months (figure 6.11a). At 6 weeks

there was a reduced total CTC count and almost no EMT CTCs detectable. In all three patients, CTCs were not detectable at 12 months post-treatment (figure 6.11b). Clinical data from all three patients reported a complete response to treatment and none of the three patients recurred during or immediately after the 12-month post-treatment timeframe.

6.5.6 Comparison of tumour and CTC EMT profiles

Having observed differing epithelial and EMT CTC phenotype profiles in patient samples, we were interested to investigate if the mesenchymal status of the CTCs could predict the tumour mesenchymal status. Given strong evidence that tumour mesenchymal status is a prognostic biomarker of PFS in HNSCC,[75] such a finding would be clinically significant. As stated in section 6.5.5, in a small pilot cohort with limited access to tumour tissue due to the COVID pandemic, firm conclusions cannot be made. Analysis was performed to demonstrate trends that would warrant future investigation.

Previous reports analysing the gene expression profile of tumour samples, initially for lung cancer and then pan-cancer, had developed and validated a method of deriving a tumour-based EMT gene signature.[196] This involved correlating gene co-expression with four primary EMT 'seed' genes: E-cadherin (*CDH1*), N-cadherin (*CDH2*), Vimentin (*VIM*) and Fibronectin (*FN1*). Using the same method, Jung *et al.* have recently utilised TCGA data from 513 HNSCC patients to derive an 82 gene EMT signature (38 epithelial and 44 mesenchymal), which was validated in a secondary cohort of 270 HNSCC patients.[75] Using this EMT signature we profiled the tumour mesenchymal status of available core tissue samples from

14 of the 20 HNSCC patients within our cohort. Comparative analysis of gene expression of the above gene signature was performed between patients and results were hierarchically clustered. Of the 14 tumours, three tumours clearly had increased mesenchymal related gene expression with the other eleven having decreased mesenchymal gene expression and/or increased epithelial related gene expression (figure 6.12). The presence of EMT CTCs in the blood of these patients was not significantly associated with tumour mesenchymal status (chi-square, p= 0.347). In addition, the tumour mesenchymal status was not associated with stage of disease or nodal metastasis (chi-square, p= 0.515 and 0.837 respectively).

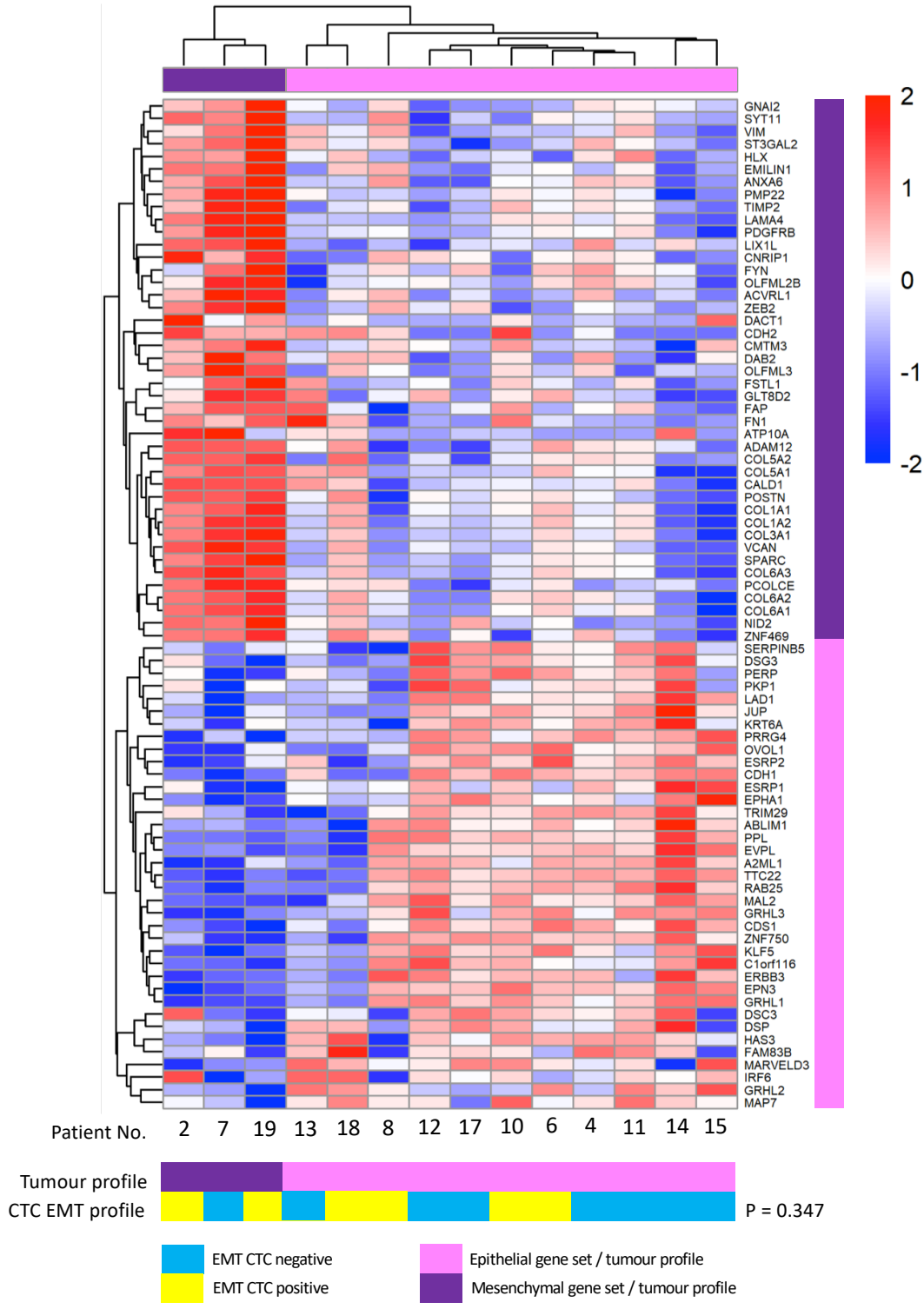
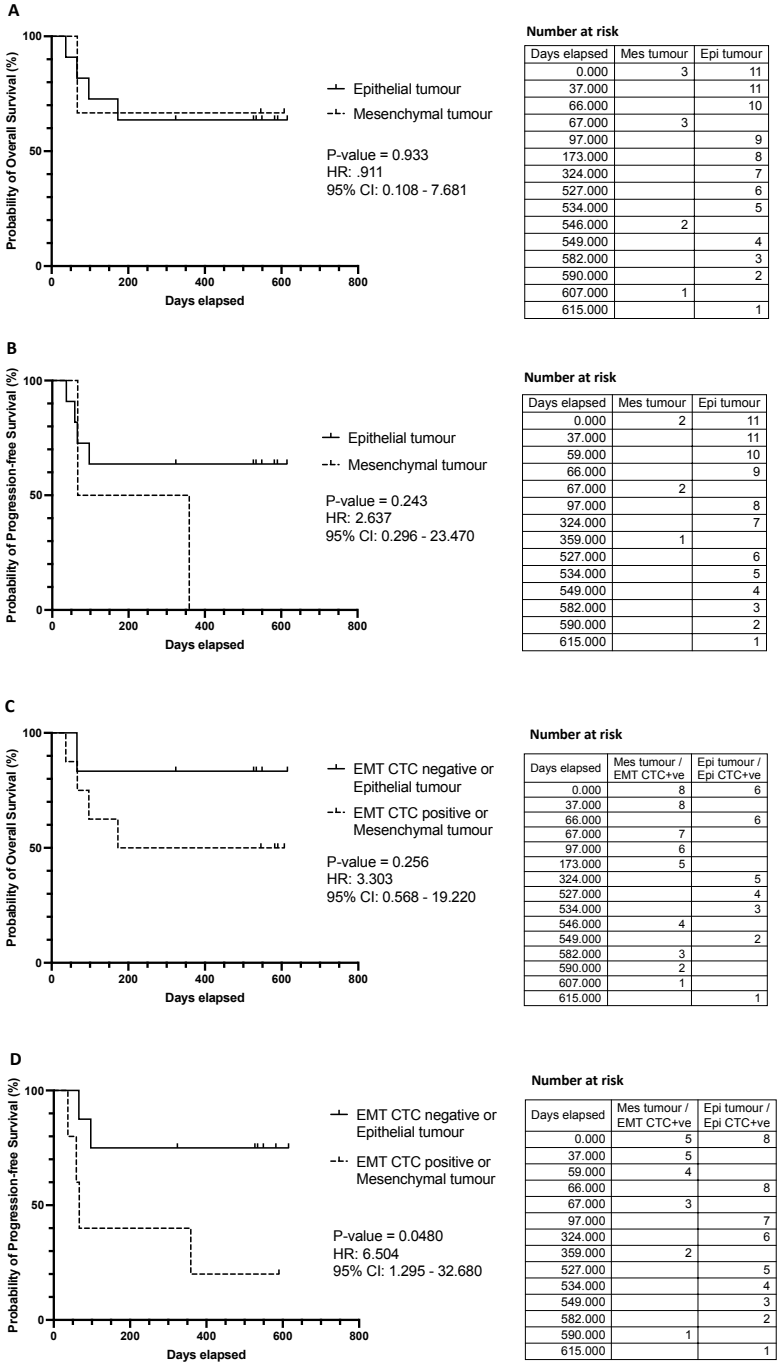


Figure 6.12: **Tumour mesenchymal status does not correlate with blood CTC EMT profile.** Heat map of gene expression in tumour samples from 14 HNSCC patients in our cohort. The genes that are shown are based on the 82 gene EMT signature. Patients clustered into mesenchymal (n=3) and epithelial (n=11) groups. Tumour profile was compared to CTC EMT expression (below) with no significant association (chi-square).

While tumour mesenchymal status did not significantly associate with OS or PFS, there was a trend of decreased PFS within mesenchymal profile tumours (figure 6.13a and b). All patients with a mesenchymal tumour profile suffered recurrence (in addition to one patent who was recruited as a recurrence and was omitted from survival analysis). Following this, combining information on tumour and CTC mesenchymal status was investigated to test whether it could provide improved prognostic stratification in our cohort. Patients with a mesenchymal profile in *either* tumour *or* CTCs demonstrated significantly worse PFS, with a HR of 6.5 and a difference in probability of survival of 75% vs 20% at 600 days follow-up (figure 6.13d). There was also a trend towards decreased OS in the mesenchymal group, with a HR of 3.3, but this was not to statistical significance (figure 6.13c).



Number at risk

Days elapsed	Mes tumour	Epi tumour
0.000	3	11
37.000		11
66.000		10
67.000	3	
97.000		9
173.000		8
324.000		7
527.000		6
534.000		5
546.000	2	
549.000		4
582.000		3
590.000		2
607.000	1	
615.000		1

Number at risk

Days elapsed	Mes tumour	Epi tumour
0.000	2	11
37.000		11
59.000		10
66.000		9
67.000	2	
97.000		8
324.000		7
359.000	1	
527.000		6
534.000		5
549.000		4
582.000		3
590.000		2
615.000		1

Number at risk

Days elapsed	Mes tumour / EMT CTC+ve	Epi tumour / Epi CTC+ve
0.000	8	6
37.000	8	
66.000		6
67.000	7	
97.000	6	
173.000	5	
324.000		5
527.000		4
534.000		3
546.000	4	
549.000		2
582.000	3	
590.000	2	
607.000	1	
615.000		1

Number at risk

Days elapsed	Mes tumour / EMT CTC+ve	Epi tumour / Epi CTC+ve
0.000	5	8
37.000	5	
59.000	4	
66.000		8
67.000	3	
97.000		7
324.000		6
359.000	2	
527.000		5
534.000		4
549.000		3
582.000		2
590.000	1	
615.000		1

Figure 6.13: Mesenchymal expression in tumour and CTC results in significantly worse progression-free survival. Kaplan-Meier plots of overall (OS) and progression-free survival (PFS) for epithelial and EMT CTC positive groups. Log-rank (Mantel-Cox) derived p-value and hazard ratio reported. Number at risk data tables are also displayed. Neither epithelial nor EMT CTC presence correlated with OS or PFS.

6.6 Summary of key findings

- *CTCs can be successfully enriched and antibody stained using Parsortix and visualised using immunofluorescence microscopy - being present with expression of epithelial and mesenchymal markers*
- *Flow cytometry analysis provides a means of quantifying CTC epithelial and EMT phenotypic subgroups*
- *The presence of CTCs was significantly associated with advance stage of disease but did not correlate with tumour (T) or nodal (N) stage, however EMT CTC count was correlated to total CTC burden*
- *Epithelial or EMT CTC presence did not significantly associate with OS or PFS*
- *In this small pilot cohort with limited access to tissue samples, we were unable to correlate tumour mesenchymal status with CTC EMT status*

CHAPTER 7: MASS CYTOMETRY SINGLE-CELL CHARACTERISATION OF CTCs IN HNSCC

7.1 Introduction to mass cytometry

Originally described in 2009, mass cytometry is a technique to analyse cell protein expression using metal labelled antibodies. In brief, each antibody stained cell is nebulised into droplets containing single cells that are then vaporised. The signal from ionised metals bound to the cell is detected by a 'time-of-flight' spectrometer. The use of isotopically pure metals conjugated to primary antibodies removes the problem of fluorophore spectral overlap experienced in flow cytometry, thus allowing a far greater number of available marker channels. While some spill over between channels is observed (due to metal impurities), in contrast to flow cytometry, compensation of data is not routinely required. With rapidly progressing advances in mass cytometry technology, there are currently 45 channels available (with potential to increase to 100) in contrast to ≈20 used in high-plex flow cytometry. Such advancements have allowed a depth of single-cell protein characterisation that was previously unobtainable.[197] As research into the application of mass cytometry in oncology continues to grow, researchers are demonstrating how it can be used to characterise tumour cells[198] and immune cells in solid tumours.[199]

Arguably, to-date the focus of single-CTC characterisation has been dominated by transcriptomics. There is no doubt that the ability to interrogate the expression of thousands of genes from a single cell is a powerful tool. As discussed in chapter 1, proof-of-principle

studies have demonstrated how this approach can be successfully applied to CTC analysis. However, there are several (often overlooked) limitations to this approach. Perhaps surprisingly, the correlation between transcriptome and proteome datasets in tumour cells is shown to be modest, in the range of 0.4 to 0.6.[200], [201] Furthermore, protein expression is more stable than transcripts in human tissue,[202] therefore being a more accurate and reliable method of cell characterisation. Undeniably, the one feature that transcriptomics cannot assess is that of post-translational modification - the most common being amino acid phosphorylation, which alters the binding capacity of a protein.[203] The addition of phosphoprotein-specific antibodies to mass cytometry panels has proved a powerful tool to illuminate cellular signalling pathways. [204] For example, a recent study correlated STAT3, CREB and NF-κB signalling in monocytes with improved post-operative recovery in hip-replacement patients.[205]

With this in mind we sought to go beyond the current standard of phenotypic marker expression defining CTC sub-populations – using mass cytometry to investigate the intracellular signalling pathways underpinning these CTC populations and thus identify potential therapeutic targets.

7.2 Aims and objectives

The aim of this chapter was to demonstrate proof-of-principle that mass cytometry could be utilised to provide single-cell characterisation of CTCs.

Objectives:

1. Design and optimise an antibody panel to perform mass cytometry on Transfix fixed cells
2. Optimise a protocol to perform mass cytometry on a Parsortix enriched cell population to enable the identification of low frequency CTCs
3. Process a pilot cohort of blood samples from HNSCC patients to identify CTCs and potential heterogeneity within sub-groups
4. Compare tumour and CTC expression profiles

7.3 Marker choice, antibody optimisation and panel design

7.3.1 Rationalising marker inclusion

At the time of writing, no studies have reported using mass cytometry to characterise CTCs in any cancer type. Therefore, without a baseline reference we were required to design our own panel. We found one report that had successfully used mass cytometry to characterise dissociated HNSCC tumour tissue in a small cohort. Although the samples in this study were unfixed cells, so all antibodies needed further optimisation for Transfix fixed cells, it was a useful starting point.

Panel design was especially challenging. It may seem surprising, given we were going from a flow cytometry panel of 4 markers up to 42 markers. However, we sought to build a future-proof panel that would allow both CTC and tumour tissue analysis. Furthermore, where

possible, we sought to include multiple markers for each phenotype/cell function. Firstly, this would create confidence in our data analysis, and secondly, we were interested to see if these markers were always co-expressed in pre-defined CTC sub-groups (as presumed in current research), or if variance in expression could define further novel CTC populations.

We defined the following four categories as an overarching framework for marker inclusion:

1. Cell lineage markers - including CTCs, tumour cells, immune cells and stromal cells
2. CTC / tumour cell phenotype markers - including EMT, stemness and proliferation
3. Immune evasion/suppression markers and druggable targets
4. Phosphorylated signalling pathway proteins

Table 7.1: Reason and references for CTC marker inclusion in cytometry panel. The table below lists markers included in the mass cytometry panel together with references to demonstrate their applicability specifically to CTC research and/or HNSCC tumour molecular biology.

MARKER	PHENOTYPE/FUNCTION	REFERENCES
CTC IDENTIFICATION		
EPCAM	Epithelial cell/CTC marker	[121] [104] [124] [118] [117] [112] [123] [206]
PAN-CYTOKERATIN	Epithelial cell/CTC marker	[117] [112] [115] [106] [123] [120] [206]
E-CADHERIN	Epithelial cell marker and 'switch' to EMT with loss of expression	[206] [207]
EGFR	Epithelial/proliferative cell marker and druggable target	[117] [112] [115] [123] [130]
EPITHELIAL-MESENCHYMAL TRANSITION		
VIMENTIN	Mesenchymal cell marker	[117] [206] [207] [208] [130] [209]
N-CADHERIN	Mesenchymal cell marker	[117] [106] [207] [209]
SNAIL1	EMT transcription factor	[130] [210] [209]
TWIST	EMT transcription factor	[130] [209]
STEMNESS/PROLIFERATION		
CD133	Stemness marker	[106] [206]
CD24	Stemness marker	[109] [211]

CD44	Stemness/proliferation marker	[117] [212] [213] [109]
Ki67	Proliferation marker	[214][215][216]
IMMUNE EVASION/SUPPRESSION		
CD47	Immune evasion	[112] [130]
PD-1	Immune response	[112]
PD-L1	Immune suppression	[112] [85] [217] [130]
PD-L2	Immune suppression	[112] [130]
CTLA4	Immune suppression	[53][218]
CD39	Immune suppression	[219]
CD73	Immune suppression	[220]
PHOSPHORYLATED SIGNALLING PROTEINS		
pSTAT1	Immune surveillance	[221] [222] [223]
pSTAT3	Proliferation and druggable target	[221] [222] [224]
pSTAT5	Cell survival	[221] [222] [225]
pPARP	DNA damage repair and druggable target	[226] [227]
pAKT	Proliferation	[228] [229]
pERK	Proliferation	[230] [231] [232]
pCREB	Proliferation/survival and druggable target	[233] [234]
p38	Proliferation/survival	[235]
pS6	Cell survival	[236] [237]

In choosing markers we were balancing the requirement to include sufficient cell lineage markers to be able to successfully identify and gate-out PBMC sub-populations, against the primary goal of multi-parameter CTC characterisation. Table 7.1 lists markers included for CTC identification and phenotype characterisation (with references of use in other CTC studies or HNSCC research). For CTC identification we included the epithelial markers EpCAM, pan-cytokeratin, EGFR and E-cadherin, the latter was also used to demonstrate EMT progression. Markers of EMT and stemness are well described in CTCs. We included Vimentin and N-cadherin as mesenchymal markers, with Twist and Snail1 included as EMT markers. Stemness markers included CD133, CD44 (also a marker for proliferation), and CD24 (a marker of

stemness in nasopharyngeal carcinoma). Ki-67 was also included as a key marker of proliferation. Regarding immune evasion markers, we included the druggable targets PD-L1, PD-L2 and CTLA4, and the ATP/adenosine pathway proteins CD39 and CD73 which are involved in immune suppression and have shown promise as potential targetable markers for immunotherapy.[219], [220]

The choice of potential intracellular proteins involved in oncogenic signalling pathways is vast, thus we focused on markers that were implicated in HNSCC and/or were potential therapeutic candidates. Such pathways are activated by upstream binding of cytokines or growth factors to receptors on the cell surface and then act downstream via transcriptional regulation or protein phosphorylation to effect cell function. The janus kinase – signal transducer and activation of transcription (JAK-STAT) pathway is heavily implicated in solid tumour formation, playing a role in cell proliferation and differentiation, and also immune signalling.[222] The JAK-STAT pathway is a potential druggable target, with therapeutic agents approved in several cancer types, and early phase clinical trials underway in HNSCC.[224] Therefore, we chose STAT1, STAT3 and STAT5 to include in our panel. Signalling proteins ERK 1/2 and AKT were also included. The extracellular signal-related kinase (ERK) 1/2 and AKT pathways are activated by growth factors, including EGFR, and regulate cell proliferation and survival - both are implicated in tumour formation and progression/survival.[228], [231] The DNA damage repair protein family of poly ADP-ribose polymerases (PARP) have also shown potential as druggable targets in oncology.[226] In HNSCC, the PARP-inhibitor Olaparib is currently in phase I clinical trial status, thus we included PARP in our panel.

Several pathway proteins which are implicated in oncogenesis but have not yet been translated into clinical trials, or are not specific to HNSCC, were included in our panel. For example, the stress proteins p38 and pS6 proteins which also sit downstream and/or alongside the aforementioned pathways, are potential therapeutic targets.[235], [236] In addition, if we observed excellent antibody binding upon Transfix fixed cells during our optimisation experiments (discussed below) we endeavoured to include these markers. One such marker was pCREB (cAMP-regulatory element-binding protein), a transcription factor downstream of the AKT pathway (among others) which is implicated in gene downregulation in oncogenesis.[233]

Unfortunately, we were unable to include all the markers we wanted in our final panel. While we focused on markers that would allow us to fulfil the goal of the project, i.e. identify CTC sub-populations and investigate druggable target expression, some useful markers were omitted. For example, metastasis markers MMP9 and TIMP3, immune markers TIM3, LAG3 and TIGIT and proliferation markers MET, ATR and PI3K were all considered but not included. With the positive results of our pilot data below, these additional markers have been noted as potential candidates in a second iteration of our panel for any larger cohort studies.

7.3.2 Antibody optimisation

There were two key questions we needed to ask regarding antibody optimisation: *1) does the antibody clone successfully bind epitopes from a Transfix fixed cell and 2) what is the optimum strategy for cell permeabilisation to facilitate phospho-antibody binding in Transfix fixed cells.*

We were fortunate that Dr Naeem Khan, from the University of Birmingham Clinical Immunology service, kindly shared an optimised protocol for phospho-antibody staining for mass cytometry application. Therefore, we used this protocol as a starting point for antibody optimisation. In brief, this protocol involved fixing cells with paraformaldehyde before surface marker staining and then methanol permeabilization to enable intracellular (i.e. phospho-antibody) staining. Thus, our cells were subject to the harsh fixation of both Transfix and PFA prior to antibody staining, beyond the normal expectations of efficacy for antibodies developed for fixed cells/tissues - a factor which may explain the decreased binding of some antibodies, as discussed below.

As per previous antibody optimisation experiments (section 6.2), to test different antibody clones for cell surface marker binding we constructed a cell control mix of healthy donor PBMCs and HNSCC epithelial cells (FaDu cell line spiked at 5-10%) which were fixed in Transfix solution for 24hrs. We also included a primary dermal fibroblast cell line in initial experiments to test the FAP antibody clone. If we could access metal conjugated antibodies from collaborators, we performed mass cytometry assays for antibody testing. However, when we were testing multiple clones on the same markers we used flow cytometry assays with secondary antibody fluorophores, for the purposes of ease and time efficiency. Initial antibody clone choice for testing was governed either by availability from collaborators or evidence of Transfix compatibility, as obtained from the Cytomark website.[238] To test intracellular markers, including phosphorylated proteins, we used the above cell control mix, however the cells were stimulated with cell activation cocktail (PMA-ionomycin, Biolegend)

for 15 minutes prior to Transfix fixation, to increase expression of intracellular signalling pathway proteins.

To assess the efficacy of antibody binding we graded results for each clone on a subjective scale of 'unacceptable/-', 'acceptable/+' or 'good/+'. Figure 7.1 gives an example of these results from flow cytometry antibody clone testing experiments, for each grade of efficacy. Table 7.2 lists the results for all testing experiments, obtained using the standard mass cytometry staining protocol. Results from each experiment were independently assessed by 2 observers (the PhD applicant Dr Karl Payne and supervisor Dr Graham Taylor who has >20 years' experience of cytometry) and then verified by open discussion with Dr Naeem Khan (a mass cytometry specialist with >30 years cytometry experience). As this was a test panel to demonstrate proof-of-principle in a pilot cohort, inter-rater variability was not assessed. However, there were no instances of observer disagreement between unacceptable and acceptable antibody staining. It should be stressed that this was an iterative process conducted over several months of optimisation, with regular review of data by all supervisors of this thesis, to ensure a robust final panel was used in CTC analysis. It should also be noted that prior to implementation of this technique/antibody panel in a larger cohort trial and subsequent clinical translation, in-depth analysis of antibody staining reproducibility would need to be assessed and verified. Data from antibody optimisation experiments is presented in the Appendix.

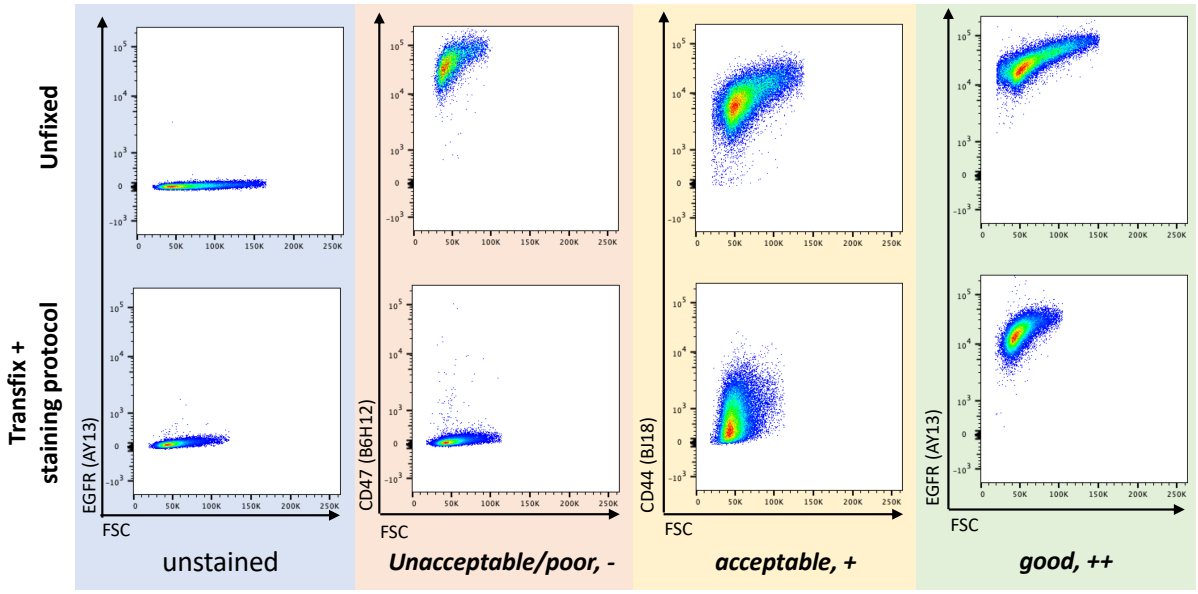


Figure 7.1: Antibody optimisation experiments demonstrate range of binding efficacies on Transfix fixed cells. Flow cytometry plots for FaDu cells, comparing antibody staining on unfixed cells (top row, conventional staining protocol with no PFA fixation) and Transfix fixed cells (bottom row, stained using mass cytometry staining protocol). A range of results, from unacceptable to good, is displayed as an example of subjective grading of staining.

Table 7.2 Results of antibody clone optimisation experiments. This table lists all markers that were tested, and the corresponding clones, for possible inclusion in the final mass cytometry panel. Staining quality is displayed as unacceptable (-), acceptable (+) or good (++).

MARKER	CELL MARKER	CLONE(S)	TEST METHOD	STAINING QUALITY
CELL LINEAGE				
EPCAM	Epithelial	9C4	FC	++
PAN-CYTOKERATIN	Epithelial	AE1	FC/MC	++
		C-11		++
E-CADHERIN	Epithelial	67A4	MC	++
EGFR	Epithelial	AY13	FC	++
CD45	pan- leukocyte	HI30	MC	++
CD3	T-cell	UCHT1	MC	++
CD4	CD4+ T-cell	RPA-T4	MC	++
CD8	CD8+ T-cell	SK1	MC	+
CD25	T-regs	2A3	MC	+
CD14	Monocytes	W052	MC	++
CD86	M1 Macrophage	FUN1	MC	++

CD163	M2 Macrophage	GHI61	MC	+
CD19	B-cells	HIB19	MC	++
CD56	NK cells	NCM162	MC	+
CD66B	Granulocytes	6/40c	MC	++
HLA-ABC	MHC class I	W632	MC	+
HLA-DR	MHC class II	L243	MC	++
CD31	Endothelial cells	clone7	FC	-
		WM59		++
FAP	Fibroblasts	F-1124	FC	++
EPITHELIAL-MESENCHYMAL TRANSITION				
VIMENTIN	Mesenchymal	V9	FC/MC	++
		RV202		+
		E5		++
N-CADHERIN	Mesenchymal	13A9	FC	-
		8C11		+
SNAIL1	EMT	G7	FC	+
TWIST	EMT	Twist2C1a	FC	+
STEMNESS/PROLIFERATION				
CD133	Stemness	S16015F	FC	+
CD24	Stemness/B-cell	ML5	MC	++
CD44	Stemness/proliferation	DF1485	FC	-
		BJ18		++
		IM7		++
Ki67	Proliferation	B56	MC	++
IMMUNE EVASION/SUPPRESSION				
CD47	Immune evasion	B6H12	FC	-
		CC2C6		++
PD-1	Immune response	EH122H7	MC	++
PD-L1	Immune suppression	MIH2	MC	++
PD-L2	Immune suppression	24F.10C12	MC	+
CTLA4	Immune suppression	14D3	MC	++
CD39	Immune suppression	A1	MC	++
CD73	Immune suppression	AD2	MC	++
PHOSPHORYLATED SIGNALLING PROTEINS				
pSTAT1	Immune surveillance	58D6	MC	++
pSTAT3	Proliferation	4/P-STAT3	MC	++
pSTAT5	Cell survival	47	MC	++
pPARP	DNA damage repair	F21852	MC	+
pAKT	Proliferation	D9E	MC	+
pERK	Proliferation	D13144E	MC	++
pCREB	Proliferation/survival	87G3	MC	++
p38	Proliferation/survival	D3F9	MC	+
pS6	Cell survival	N7548	MC	+

As a general observation the binding of antibodies to Transfix fixed cells using the mass cytometry staining protocol was 'good' for the majority of markers. For commonly investigated markers, multiple antibody clones were tested to achieve the best possible level of binding - for example, we tested 3 clones for Vimentin. However, with some less common markers, for example Twist and Snail1, we were limited with antibody choice and availability (in part due to the COVID-19 pandemic), thus settling for 'acceptable' binding to progress the project.

One area of further testing were the phosphoprotein-specific antibodies, which require careful optimisation due to phosphorylated-proteins being more labile and antibody binding potentially impacted by fixation, when compared to other proteins.[239] Several phosphoprotein markers in our provisional panel were impacted by Transfix fixation. The markers PARP, AKT, S6 and p38 were deemed to have 'acceptable' binding, and we were keen to see if this could be improved by modifying the method of cell permeabilisation for intracellular antibody staining. Fourteen of the markers in our panel were intracellular markers. We compared the standard method of methanol cell permeabilisation to protocols using 'true-phos perm buffer' (a reagent from Biolegend designed specifically for phospho-antibody staining) and the 'inside stain kit' (a protocol from Miltenyi Biotec for intracellular antibody staining). Despite numerous experiments comparing different permeabilization protocols we concluded that the previously optimised method of methanol cell permeabilization yielded the best results for cells treated with Transfix. While we were able to achieve marginally improved binding of phosphoprotein-specific antibodies with the true-phos buffer, the binding of antibodies to some non-phosphoprotein intracellular markers, notably Vimentin and Ki67, decreased. Thus, we were balancing the advantages and

disadvantages of all protocols, with no one perfect option. We decided against compromising the binding of CTC phenotype markers to improve a select few phospho-antibodies, thus continuing with the methanol permeabilisation protocol.

7.3.3 Mass cytometry panel design and antibody titration

As discussed, mass cytometry has the availability of 45 metals with which to construct an antibody panel. Certain metals produce a stronger signal and are therefore useful for markers known to have decreased expression or reduced antibody binding. In addition, some metals exhibit increased spillover into neighbouring channels, so marker choice needed to be carefully considered i.e. avoiding the use of channels susceptible to such spillover to detect markers present on the same cell type thus avoiding artefactual false positive results. One final factor in panel design is the oxidation of conjugated metals, whereby some metals are prone to oxidation when vaporised in the mass cytometer, producing spillover signal 16 atomic mass units (the mass of oxygen) above the metal's atomic mass. Where possible we purchased ready conjugated antibodies from Fluidigm. The remaining markers were purchased as primary antibodies and metal conjugated using the Maxpar antibody labelling kit (Fluidigm), as per the manufacturers protocol. Once the panel was assembled we tested the antibodies (as previously) with a synthetic sample comprising PBMCs from a healthy donor spiked with the HNSCC cell line FaDu. Based on these tests we identified that 3 markers CD73, Vimentin and CD44 produced overly strong signal causing spillover into other important detection channels (figure 7.2). We therefore moved these markers onto different metals and kept these channels empty to avoid future spill over.

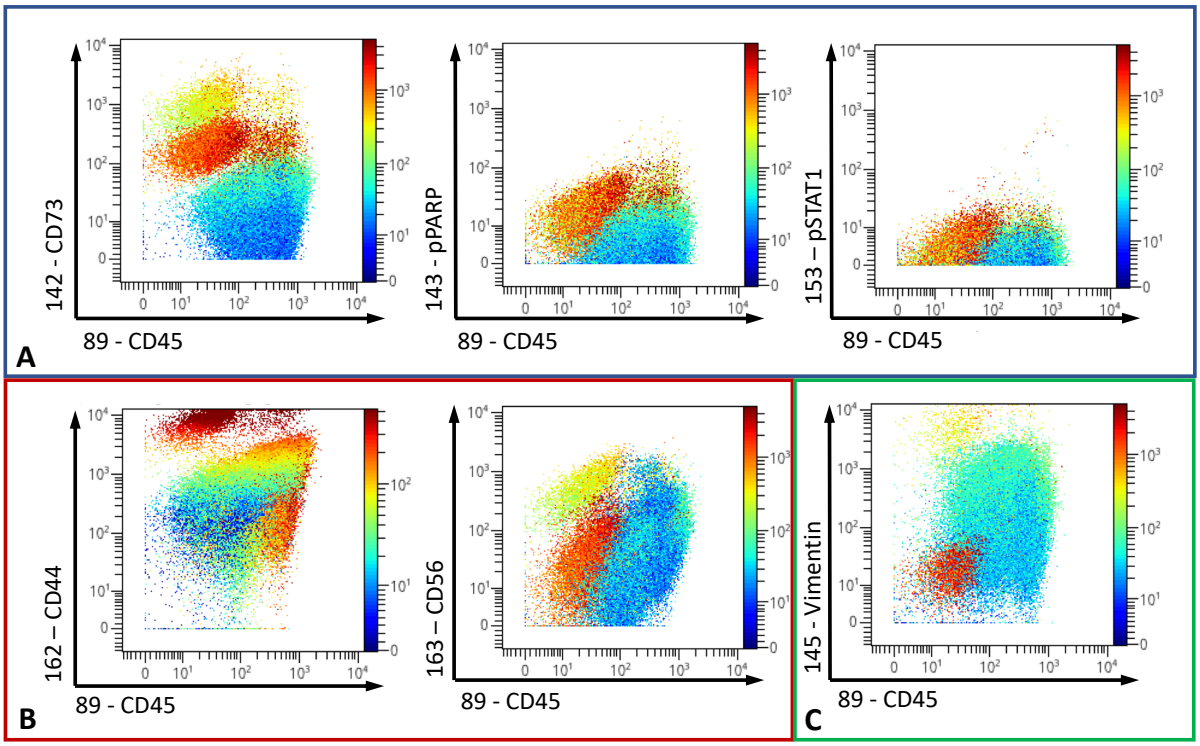


Figure 7.2: Assessing signal spillover for marker channels. Biaxial mass cytometry data plots are colour coded for EpCAM marker expression (right y-axis) to identify HNSCC cell line. A) Top left plot demonstrates high signal for CD73 on indium 142 metal that is spilling into the adjacent neodymium 143 channel used to detect pPARP (in the top middle plot). These cells were not stimulated, so we would expect to see pPARP with the same intensity (i.e. $<10^1$) as pSTAT1 (shown in top right plot as a reference). Such spill over may produce a false positive result. B) Bottom left plot shows very high signal for CD44 on dysprosium 162 is spilling into dysprosium 163 channel and causing a false positive signal for CD56 (bottom middle plot). C) Bottom right plot showing excessively high Vimentin signal on neodymium 145 making the identification of positive and negative cell populations challenging.

The final set of optimisation experiments was to perform antibody titrations to determine the minimum required concentration of antibody needed to detect physiological levels of markers on relevant cell types. We compared 0.25, 0.5 and 1 µg antibody volumes to provide the optimum separation of positive control cells within our testing cell cocktail. Conjugated

antibodies supplied by Fluidigm are supplied as multiples of 1 µl 'tests' – for these antibodies we tested titrations of 0.25, 0.5 and 1 µl. Multiple antibody titration experiments were performed, to achieve an n=3 for each marker. As discussed in section 7.3.2, the process of selecting concentrations was performed subjectively with assessment of data by 2 observers (Dr Karl Payne and Dr Graham Taylor) and corroborated by a third observer (Dr Naeem Khan). The final panel, together with concentrations is listed in table 7.3. As can be seen the majority of markers demonstrated satisfactory staining down to 0.25 µg, with the phospho-antibodies requiring 0.5 µl.

Table 7.3: Final mass cytometry panel. Markers included in the mass cytometry panel are shown, together with the corresponding metal label, antibody clone and supplier (F=Fluidigm, BL=Biolegend, SC=SantaCruz). Extracellular (purple) or intracellular (red) epitope location and concentration is listed as used in the cell staining protocol.

Metal	marker	clone	Supplier	Extracellular/ Intracellular	Concentration
089 Y	CD45	HI30	F		0.25
106 Cd	B2M, CD298	2M2, LnH-94	BL		0.25
110 Cd	CD14	W052	BL		0.25
111 Cd	Vimentin	E5	SC		0.25
112 Cd	CD66b	6/40c	BL		0.25
113 Cd	CD19	HIB19	BL		0.25
114 Cd	CD44	IM7	BL		0.25
115 In	CD3	UCHT1	BL		0.25
116 Cd	CD4	RPA-T4	BL		0.25
139 La	empty channel to avoid spill over				
141 Pr	EpCAM	9C4	BL		0.25
142 Nd	empty channel to avoid spillover				
143 Nd	PARP	F21852	F		0.5
144 Nd	CD31	WM59	F		0.25
145 Nd	empty to avoid spill over				

146 Nd	CD86	FUN1	BD		0.25
147 Sm	Twist	Twist2C1a	SC		0.5
148 Sm	PD-L2	24F.10C12	BL		0.25
149 Sm	CD25	2A3	F		0.25
150 Nd	pSTAT5	47	F		0.5
151 Eu	empty channel to avoid spill over				
152 Sm	pAKT	D9E	F		0.5
153 Eu	pSTAT1	58D6	F		0.5
154 Sm	E-cadherin	67A4	BL		0.25
155 Gd	PD1	EH122H7	F		0.25
156 Gd	p38	D3F9	F		0.5
158 Eu	pSTAT 3	4/P-STAT3	F		0.5
159 Tb	CD133	S16015F	BL		0.25
160 Gd	CD39	A1	F		0.25
161 Dy	CTLA4	14D3	F		0.25
162 Dy	empty channel to avoid spill over				
163 Dy	CD56	NCAM162	F		0.25
164 Dy	pan-keratin	AE1	SC		0.5
165 Ho	PD-L1	MIH2	BL		0.25
166 Er	CD24	ML5	F		0.25
167 Er	FAP	F11-24	SC		0.25
168 Er	Ki-67	B56	F		0.5
169 Tm	N-cadherin	8C11	BL		1
170 Er	EGFR	AY13	F		0.25
171 Yb	pERK	D13144E	F		0.5
172 Yb	CD163	GHI61	BL		0.5
173 Yb	empty channel to avoid spill over				
174 Yb	Snail 1	G-7	SC		0.5
175 Lu	pS6	N7548	F		0.5
176 Yb	pCREB	87G3	F		0.5
194 Pt	CD73	AD2	BL		0.25
195 Pt	CD8	SK1	BL		0.25
196 Pt	HLA-DR	L243	BL		0.25
198 Pt	HLA-ABC	W632	BL		0.25
209 Bi	CD47	CC2C6	F		0.25

To confirm metal conjugation of the antibodies in our panel we used compensation beads as a positive control. Each marker was incubated separately with a small aliquot of beads, then all combined together and the mixture run as one sample. Analysis of the data using t-distributed stochastic neighbour embedding (tSNE) dimensionality reduction separated bead populations into clearly defined groups (figure 7.3), thus confirming a mass cytometry compatible antibody panel for further experiments.

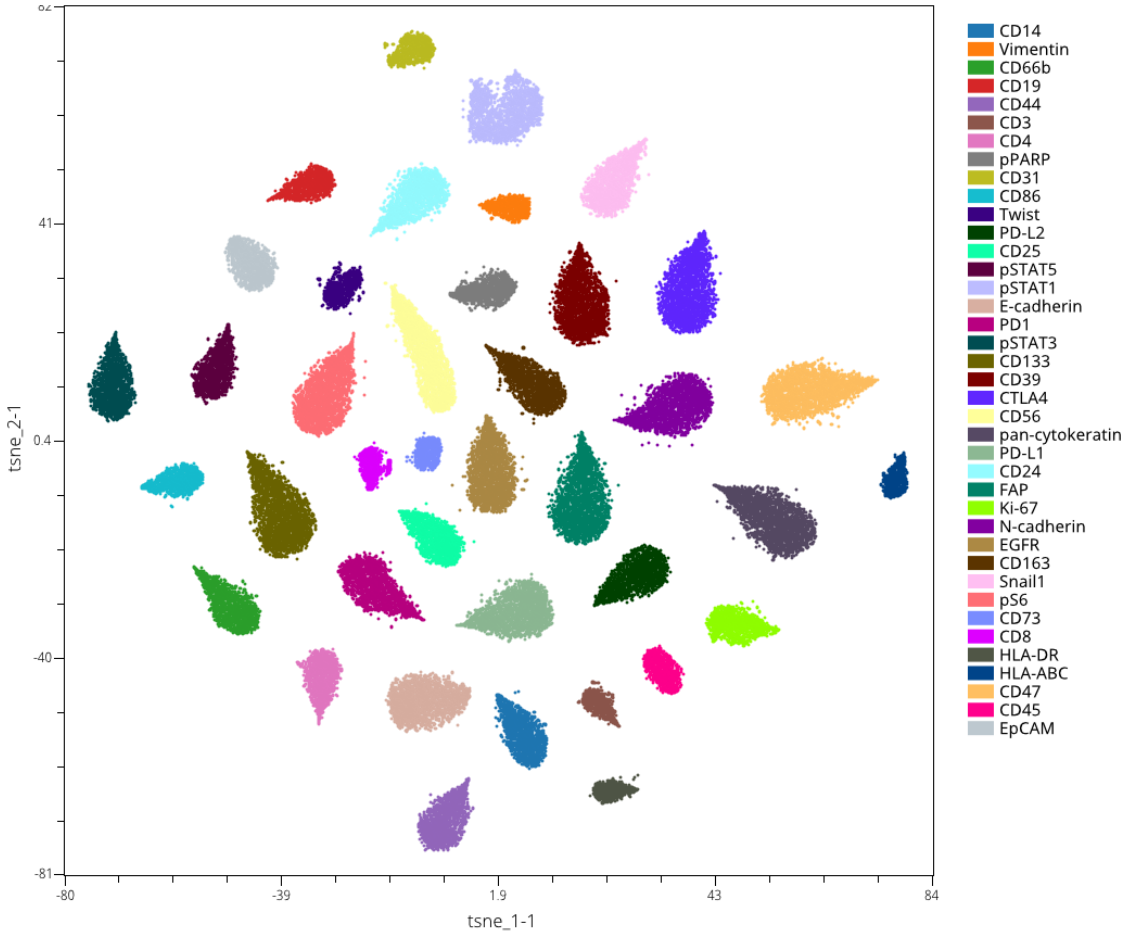


Figure 7.3: Compensation bead antibody validation. tSNE plot demonstrates populations of compensation beads bound to antibodies. Clear separation and signal confirms metal conjugation. Note – pERK and pAKT antibodies were not from mice and thus not compatible with the compensation beads. As both antibodies were purchased already metal-conjugated from Fluidigm, and subject to manufacturer batch control, analysis using beads was deemed unnecessary.

7.3.4 Discussing the subjective and iterative process of antibody optimisation, titration and panel design in mass cytometry

Designing a mass cytometry panel is an iterative process requiring the adjustment of antibody concentrations and the location of markers on specific metal isotopes in the panel. Several factors, such as metal spillover, are balanced to achieve the 'optimum' but often not perfect panel. As the requirement of each panel and target cell population is different between studies, one cannot apply a 'one-size fits all' approach. Experience and in-depth knowledge of mass cytometry is crucial. Therefore, I was fortunate to have the assistance of supervisor Dr Graham Taylor and mass cytometry specialist Dr Naeem Khan, who together have >50yrs cytometry experience. Our optimisation and panel design stage were significantly hampered by the ongoing global COVID pandemic. Access to antibodies, especially rare clones, was limited and experiments were disrupted. To make progress and speed up the optimisation process, where possible we endeavoured to source metal-labelled antibodies from colleagues. For this reason, we needed to conduct multiple test panels using both flow and mass cytometry across a period of several months, further exacerbated by laboratory shutdowns during various national lockdowns. Data from optimisation, spillover and titration experiments is presented in the Appendix.

One of the primary concerns when optimising antibody staining or titration is the ability of an antibody to identify expected biological variance within target cell populations, and less so the precise signal intensity of each marker. Therefore, we assessed antibody concentrations on positive control target cell populations. In this regard the precise signal intensity during optimisation experiments is not the main outcome as this will be reproducible in-patient

sample experiments by virtue of a consistent antibody concentration and metal isotope being used. For all antibody optimisation and titration experiment an n=3 was achieved for each marker, but this was achieved across multiple iterations of panel design and analyses. For this reason the supervisory team with extensive cytometry experience was crucial to ensure this phase of the project was as well-designed, robust and reproducible as allowed by the research environment at the time. Of greater importance is a standardised positive and negative control used for antibody assessment, thus data is reproducible between experiments. We used PBMCs from a single healthy donor which were frozen in aliquots of 1 million cells for each experiment. In addition, the FaDu HNSCC cell line was used as a constant epithelial (CTC) positive control in all experiments. This cell line, together with the CAL27 cell line, was included in the final panel test to demonstrate positive antibody staining (discussed below).

It should be noted that in-depth discussion of antibody staining (sometimes referred to as binding), affinity and avidity kinetics is outside the scope of this thesis. In brief, antibody staining is a product of affinity and avidity. Affinity is measured by K_D (the equilibrium dissociation constant between the antibody and its antigen) and avidity refers to the additive effects of multiple binding sites within the antibody molecule which varies depending on the lass of antibody (IgG<IgA<IgM).[240] Data to demonstrate the correlation of antibody staining and signal intensity (either from flow cytometry fluorophore or mass cytometry metal tagged signal) is quality checked by the supplier (on a positive control) and this approach has previously been validated by comparing antibody signal intensity to surface plasmon resonance generated K_D values.[241] As discussed above, this signal intensity in mass cytometry is also dependent on the brightness of the antibody tagged metal used, the relationship to marker intensity and subsequent spillover between channels when these are

combined. Thus, the purpose of our optimisation strategy was to initially assess antibody staining on fixed control cells (figure 7.1 and table 7.2) and then assess spillover of metal signal intensity in the final constructed panel (figure 7.2 and table 7.3).

7.3.5 Pilot test on Parsortix enriched cell line spiked blood sample

While multiple 'mock CTC' cell line spiked experiments were used throughout the above discussed optimisation process, we present data below of a final test before proceeding to patient samples - in order to highlight our data analysis strategy. We processed and analysed a 9ml blood sample from a healthy donor into which was spiked 1,000 epithelial cells (a 1:1 ratio of the FaDu and CAL27 HNSCC cell lines). Thus, this 'synthetic' sample mimicked the volume of blood available from a patient (9ml) and the likely heterogeneity of CTC sub-groups. The sample was processed as per methods section 4.9, with barcoded PBMCs from the same donor added to act as carrier cells. Normalised mass cytometry data files underwent clean-up gating as described by Bagwell *et al*.[242] using gaussian discrimination parameters (figure 7.4). Following this step, barcoded PBMC carrier cells were removed from the Parsortix enriched cell population (figure 7.5a). Dimensionality reduction using tSNE, on cell lineage markers, was used to define the group of spiked epithelial cells - visible as an EpCAM positive cluster spatially separated from the contaminating Parsortix enriched PBMCs (figure 7.5b). Principal component analysis (PCA) of the EpCAM positive population, performed using CTC phenotype and signalling pathway marker input variables, separated the two cell lines into clearly defined groups (figure 7.5c). Using colour coded heat intensity for individual markers on the PCA plot, we were able to demonstrate how the high dimensional data from mass

cytometry identified heterogeneity within the cell lines at the single-cell level (figure 7.5c). This strategy formed the basis of our approach to analysing patient samples.

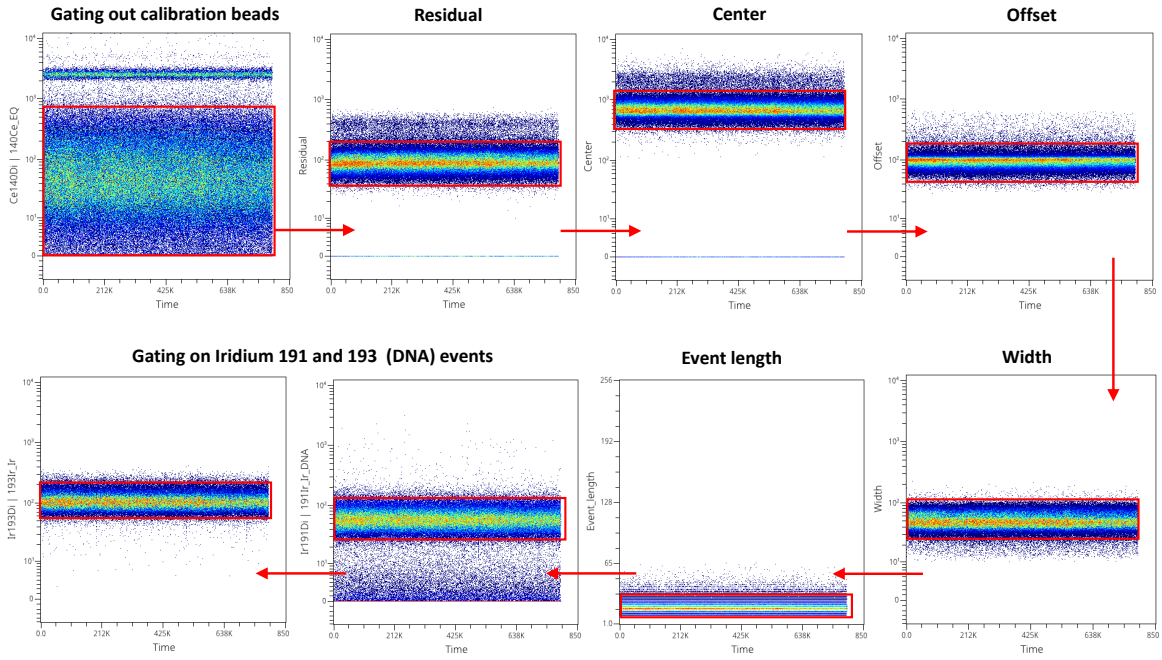


Figure 7.4: Mass cytometry data clean-up strategy. Clean-up gating strategy for mass cytometry datasets as described by Bagwell et al. The final population of iridium DNA positive events (bottom left) can then be gated into cell populations (as per figure 7.5).

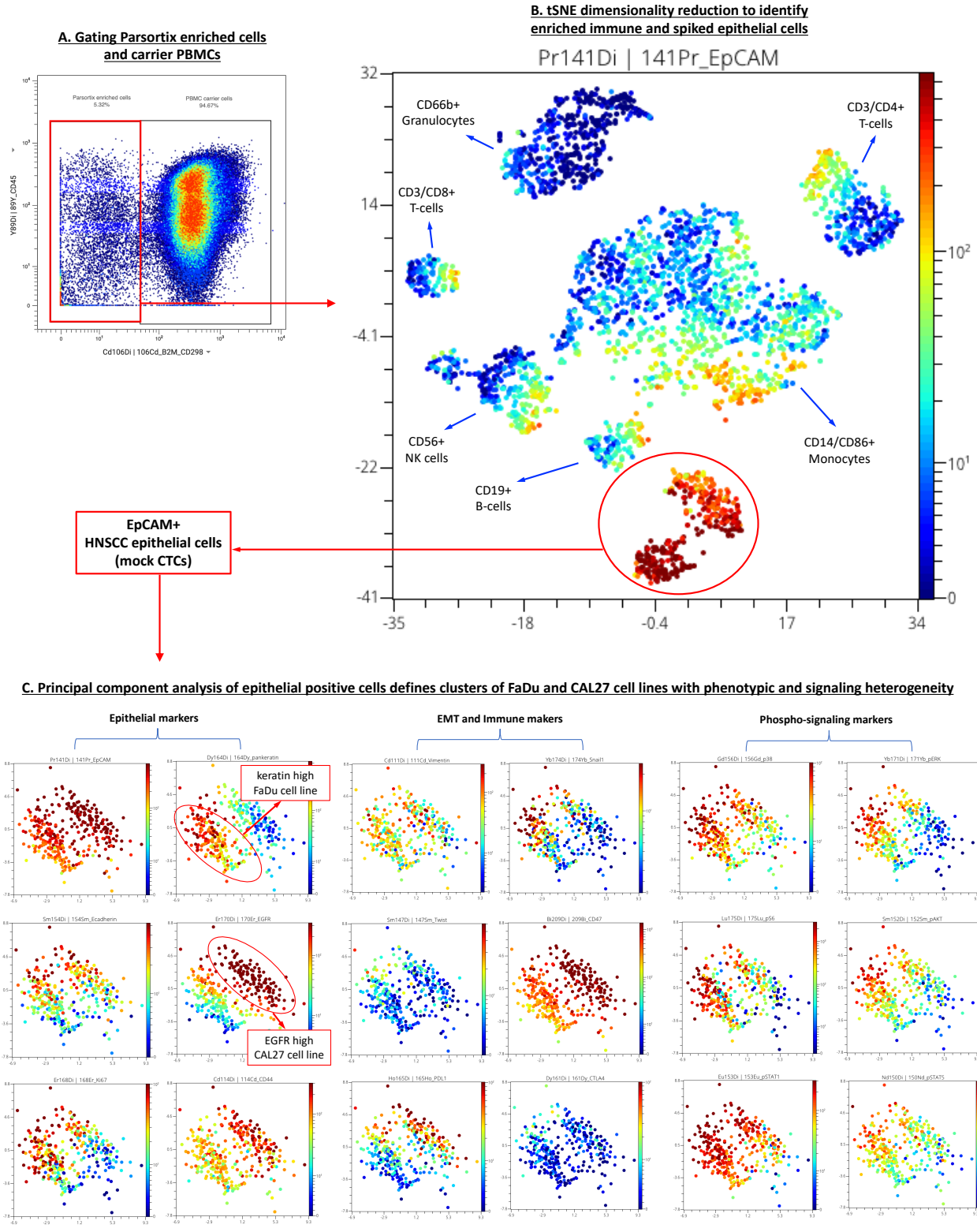


Figure 7.5: Mass cytometry successfully identifies spiked HNSCC cell lines from Parsortix enriched whole blood. FaDu and CAL27 HNSCC cell lines, spiked into healthy donor whole blood (acting as a 'synthetic' CTC sample), were Parsortix enriched and underwent mass cytometry characterisation as per the described protocol. A) Parsortix enriched cells are gated separately from the carrier PBMCs. B) tSNE dimensionality reduction of

Parsortix enriched cells separates cells into phenotypic groups. EpCAM high epithelial cells (spiked in mock CTCs) can be clearly identified as shown by the EpCAM colour map within the tSNE plot (other immune cell populations are labelled but colour maps not shown). C) PCA plots of the epithelial cells identified in (B) are colour mapped for epithelial/EMT, phenotype and phosphoprotein markers – within this PCA plot the FaDu and CAL27 cell lines cluster separately with differing marker expression profiles i.e. we can identify potential heterogeneity within the target cell population).

7.4 Mass cytometry characterisation of Parsortix enriched patient blood samples

7.4.1 Patient sample processing and data analysis

Following successful optimisation, we proceeded to process and analyse patient samples. Blood and tumour tissue samples were prospectively collected from 12 HNSCC patients (methods section 4.13) with demographic and clinicopathological data shown in table 7.4. Pre-treatment blood samples were collected in Transfix collection tubes and processed as described (methods section 4.9). Mean patient age was 64 years (range 52-77), with 8 of the 12 patients being male. The majority of patients were advanced stage of disease (10/12), with 5 patients positive for nodal metastases. The oral cavity tumour subsite predominated (11/12).

After processing 5 samples we paused to analyse the datasets. Having found interesting variability of signalling pathway proteins in CTC sub-groups, we added a phospho-antibody control to our protocol – using osmium barcoded stimulated donor PBMCs (methods section 4.9), providing internal validation of positive phospho-antibody staining in each experiment.

Despite 'acceptable' results in optimisation experiments (table 7.2) we observed that N-cadherin and pS6 antibodies did not perform well in Transfix fixed patient samples, the reason for this was unknown. We were unable to go back and optimise additional antibody clones, therefore these were removed from our panel in subsequent analysis.

Table 7.4. Demographic and clinicopathological data for the twelve patients in our mass cytometry cohort. (S = surgery, CRT = chemoradiotherapy, RT = radiotherapy).

PATIENT NUMBER	AGE	GENDER	T	N	M	STAGE	TUMOUR SITE	TREATMENT
1	77	M	4a	1	0	IV	Oral	S + CRT
2	65	M	4a	0	0	IV	Larynx	S
3	73	M	3	2b	0	IV	Oral	S + RT
4	75	M	3	2b	0	IV	Oral	S + RT
5	53	M	3	0	0	III	Oral	S + CRT
6	63	M	4a	0	0	IV	Oral	S + RT
7	43	F	4a	2b	0	IV	Oral	S + CRT
8	66	F	4a	2b	0	IV	Oral	S + RT
9	57	M	3	0	0	III	Oral	S
10	52	F	2	0	0	II	Oral	S
11	73	M	4a	0	0	IV	Oral	S + RT
12	76	F	1	0	0	I	Oral	S

Initial data analysis followed the pipeline as described in figure 7.4 and 7.5. However, we observed that when using tSNE dimensionality reduction on the entire Parsortix enriched population (as was the first step in data analysis) CTCs did not cluster as clearly. We were concerned we were potentially missing CTCs with downregulated epithelial (EpCAM, cytokeratin, E-cadherin) expression due to a mesenchymal-transition state. Therefore, we

initially plotted the Parsortix enriched cells to gate on a CD45 low and pan-cytokeratin high population, using the CD45 high PBMCs as a negative control. Dimensionality reduction on this population revealed a large population of CD66b high granulocytes, with smaller poorly defined populations of CD66b low cells, among which were presumed to be our CTCs. On this CD66b low population we manually gated out PBMC lineage marker positive cells. We considered this a 'triple confirmation' of CTC identification – defining CTCs as *CD45 low, epithelial marker intermediate/high and PBMC lineage marker low* cells. The CTCs were then analysed using PCA (with EMT, phenotype and phospho-signalling input variables) to examine sub-groups within the CTC population (figure 7.6).

As per our protocol with flow cytometry CTC characterisation (chapter 6), we processed 5 healthy donor control samples to define a confidence threshold with which to confirm CTC presence. We identified events meeting our above CTC criteria in 2 of 5 healthy donor samples at a level of 2 and 3 events. Thus, we defined >3 events as a threshold for CTC detection.

CTC count ranged from 18 – 183 CTCs from a 9ml blood sample, or 2 – 20.3 CTCs/ml whole blood. There was no positive correlation evident when comparing T-stage, positive nodal metastasis or overall stage of disease and total CTC count. One patient with a T1N0 (stage I) tumour demonstrated a total of 4 events meeting our CTC criteria. We deemed this to be too close to our confidence threshold of 3 events set by healthy control samples, therefore we labelled this patient as 'CTC negative'.

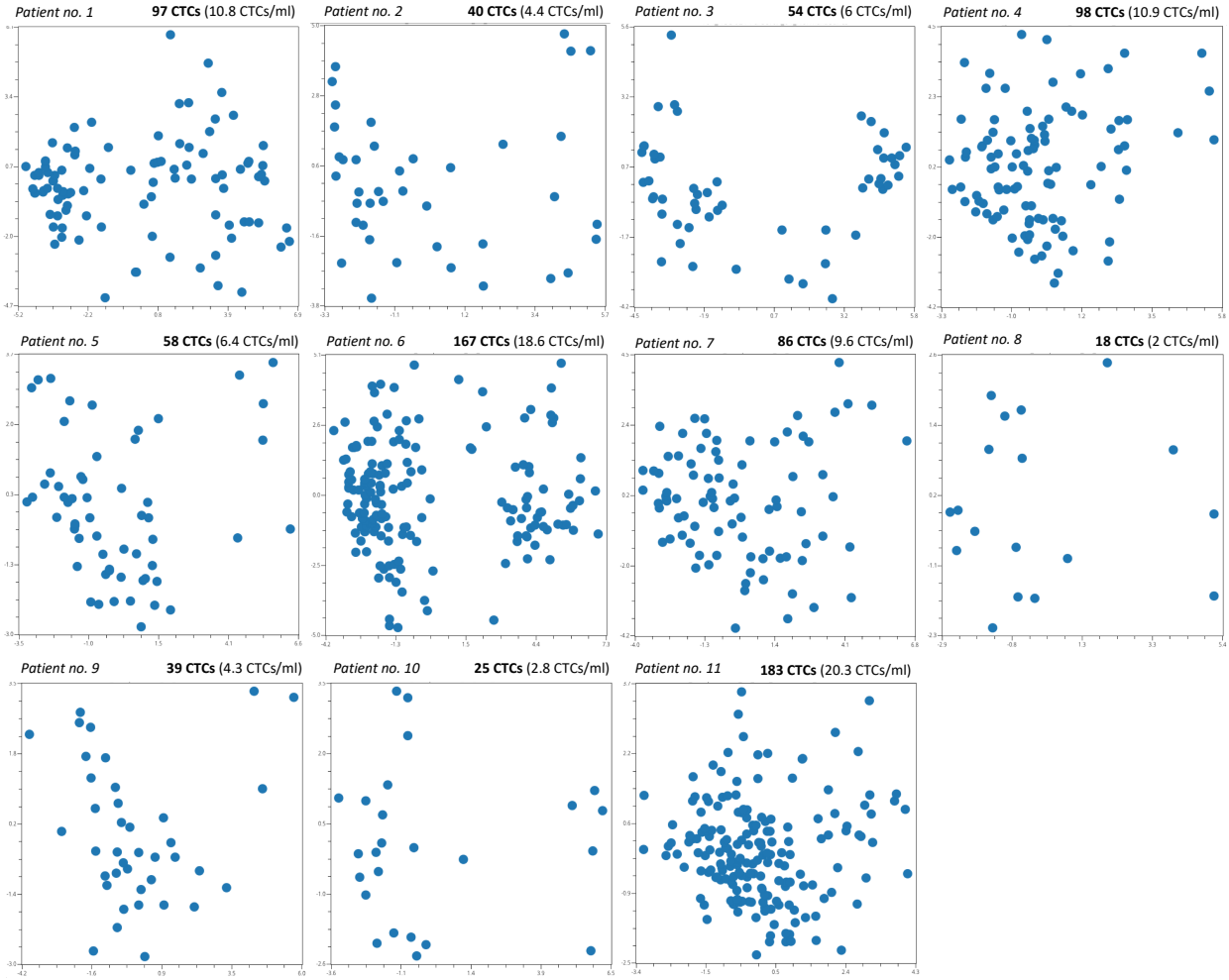


Figure 7.6: CTCs visualised on PCA plots. Results of CTC identification using mass cytometry for 11 (CTC positive) HNSCC patients, showing total number of enriched CTCs and CTCs per ml of whole blood. PCA plots (x-axis = principal component 1, y-axis = principal component 2), using input variables of EMT, phenotype and phospho-signalling markers, are used to visualise marker expression in a 2-dimensional space. In some patients CTCs can be observed to co-locate into clearly defined sub-groups. These PCA plots can then be used to investigate individual marker expression using overlaid signal intensity colour mapping.

One potential limitation of our protocol was the need to process and antibody stain samples soon after collection (i.e. after 24hrs Transfix incubation). It is preferable to collect samples into larger batches for simultaneous processing to reduce the impact of so-called 'batch-effect' of non-biological variables between individual sample experiments. We endeavoured

to reduce cytometer related batch effect by freezing processed samples and running them in batches of 3-4 at a time. Samples were frozen as per the mass cytometer manufacturer (Fluidigm) recommendations [243] - with strong published evidence that this protocol produces very minimal cell loss or signal intensity change and is an accepted protocol for sample storage within the cytometry community.[244][245] We use a prolonged 'harsh' fixation time of 12hrs to rigidly fix cells and antibodies together, which further diminishes any freeing impact. In addition, antibody signal is determined by metal label binding to the cell, which is inert and not interfered with by freezing. Patient samples 1 and 2 were run on the cytometer immediately (i.e. un-frozen) and the remaining samples frozen and run in batches – as can be seen from the heatmap in figure 7.9, no obvious single intensity variation was observed across target markers in these un-frozen and frozen samples. While all antibody optimisation experiments (including the spiked test case) were conducted on unfrozen cells, given the above manufacturer guidance and evidence we think this is acceptable practice.

To build confidence in the consistency of inter-sample antibody staining we compared the median signal intensity of 5 markers from within our carrier PBMC population. We chose markers from easily defined cell populations with known stable protein expression – the T-cell makers CD3, CD4 and CD8, monocyte marker CD14 and marker HLA-ABC. Comparing the mean (4.0, 3.6, 2.6, 2.6 and 4.2) and standard deviation (0.13, 0.21, 0.18, 0.26 and 0.26) of the above markers respectively, it was reassuring to observe minimal variability in median signal intensity across our 12 patient samples (figure 7.7).

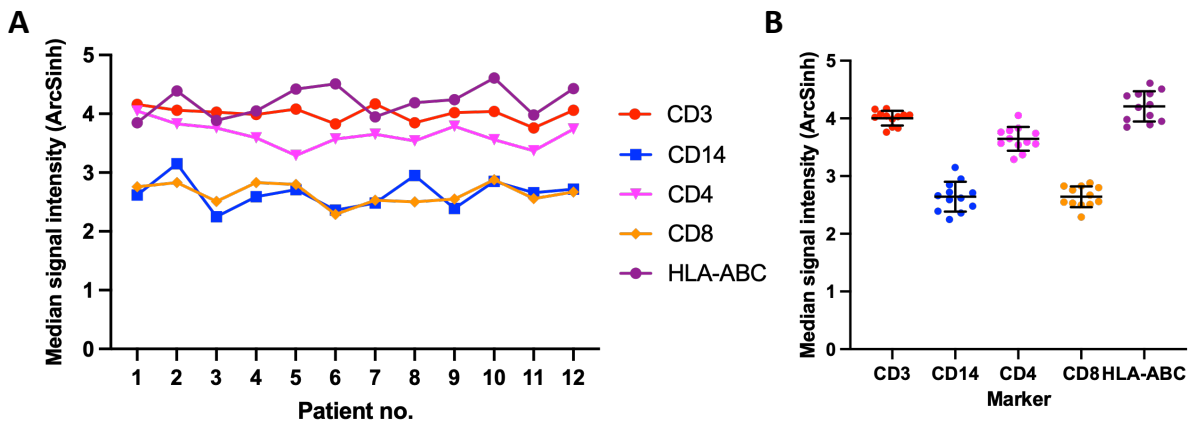


Figure 7.7: Antibody signal intensity is consistent between patient samples. A) Median signal intensity for 5 markers plotted as a continuous line across 12 samples. B) Dot plots of signal intensity for all 12 samples across 5 markers (middle line denotes mean and whiskers denote 1 standard deviation). Across all markers the inter-sample variance, as evidenced by a narrow standard deviation, was minimal and acceptable for data analysis.

Rationalising our data analysis pipeline in response to considerable and unexpected CTC heterogeneity

Instead of gating subjectively on groups identified visually on PCA plots, as an additional step to the pipeline shown in figure 7.5, we performed formal clustering analysis using FlowSOM - so as to objectively define CTC sub-groups based upon EMT, phenotypic and signalling pathway marker expression. The primary reason for the addition of this step was that CTCs demonstrated heterogeneity far beyond that seen in homogenous cell lines in our test case. Therefore, we needed to adapt our analytical pipeline to account for this biological variation, which was both unexpected and novel (i.e. not reported by other groups).

FlowSOM is a powerful algorithm optimised for unsupervised metaclustering of cytometry datasets.[246] Performing manual clustering of high-dimensional data sets is ultimately

impossible given the limitations of visualising such data in a 2 or 3-dimensional space. Thus, utilising the computational method of FlowSOM clusering allowed a higher level of data analysis and an *objective* and reproducible means of unsupervised clustering which removed the potential bias of manual clustering.

While FlowSOM will automatically define a 'k' value deemed to be the optimum number of significantly different clusters using an 'elbow' estimation method, the user can also set predetermined k values, i.e. telling FlowSOM how many clusters to find. We set k values from 2-10 for each sample analysis, setting a cut-off that a CTC sub-group (FlowSOM cluster) should be made up of 5 or more cells, so as to be of biological significance and decreasing the risk of false discovery (in one sample we accepted a cluster size of 4 cells due to clear separation of cell groups on PCA and heatmap plots).

Assigning biological significance to computationally derived CTC clusters

The benefit of an unsupervised clustering method such as FlowSOM is that clustering is standardised and reproducible between samples.[246] However, such computational algorithms cannot recognise the biological significance and potential translational potential of specific clusters or patterns of expression between clusters and/or patients. To manually interrogate marker expression in clustered groups, it is common practice in cytometry to construct heatmaps of median signal intensity.[247], [248]

We created heatmaps of median signal intensity of ArcSinh transformed values for individual markers plotted for each FlowSOM cluster and displayed them on an individual patient basis (figure 7.9). Carrier PBMC and stimulated PBMC populations were used as control references. ArcSinh is the standard transformation algorithm used in mass cytometry data analysis - it allows populations to be visualised on a linear scale when signal intensity may be up to x10,000 greater in positive events. Thus, following FlowSOM clustering analysis, we manually interrogated heatmaps of marker expression for each cluster to identify common patterns of expression and thus define CTC sub-groups verified as being present in multiple patient samples. A flow diagram of the thought process of marker expression criteria for cluster categorisation is shown in figure 7.8 and annotated heatmaps for EMT, phenotype and phospho-protein marker expression are shown in figure 7.9. This process was conducted by Dr Karl Payne and Dr Graham Taylor and the rationale for sub-group labelling is discussed in-depth below.

7.4.2 Clustering analysis identifies novel phenotypic sub-groups

We initially categorised clusters into sub-groups based upon EMT status (figure 7.8). As per our gating strategy, all clusters demonstrated intermediate/high pan-cytokeratin expression and clusters with additional high expression of the epithelial markers EpCAM and E-cadherin were labelled as ‘epithelial’ cell groups. Clusters with high expression of the mesenchymal marker Vimentin were labelled as ‘EMT’ cell groups. Within this EMT group, we observed clusters with decreased E-cadherin expression, a hallmark of EMT, which we labelled as ‘advanced EMT’. In the advanced EMT sub-group we also observed decreased EpCAM expression, although to a lesser extent than E-cadherin. In addition, a sub-group of the

Vimentin expressing EMT clusters demonstrated high expression of *both* epithelial and mesenchymal markers, which we labelled as 'early EMT'.

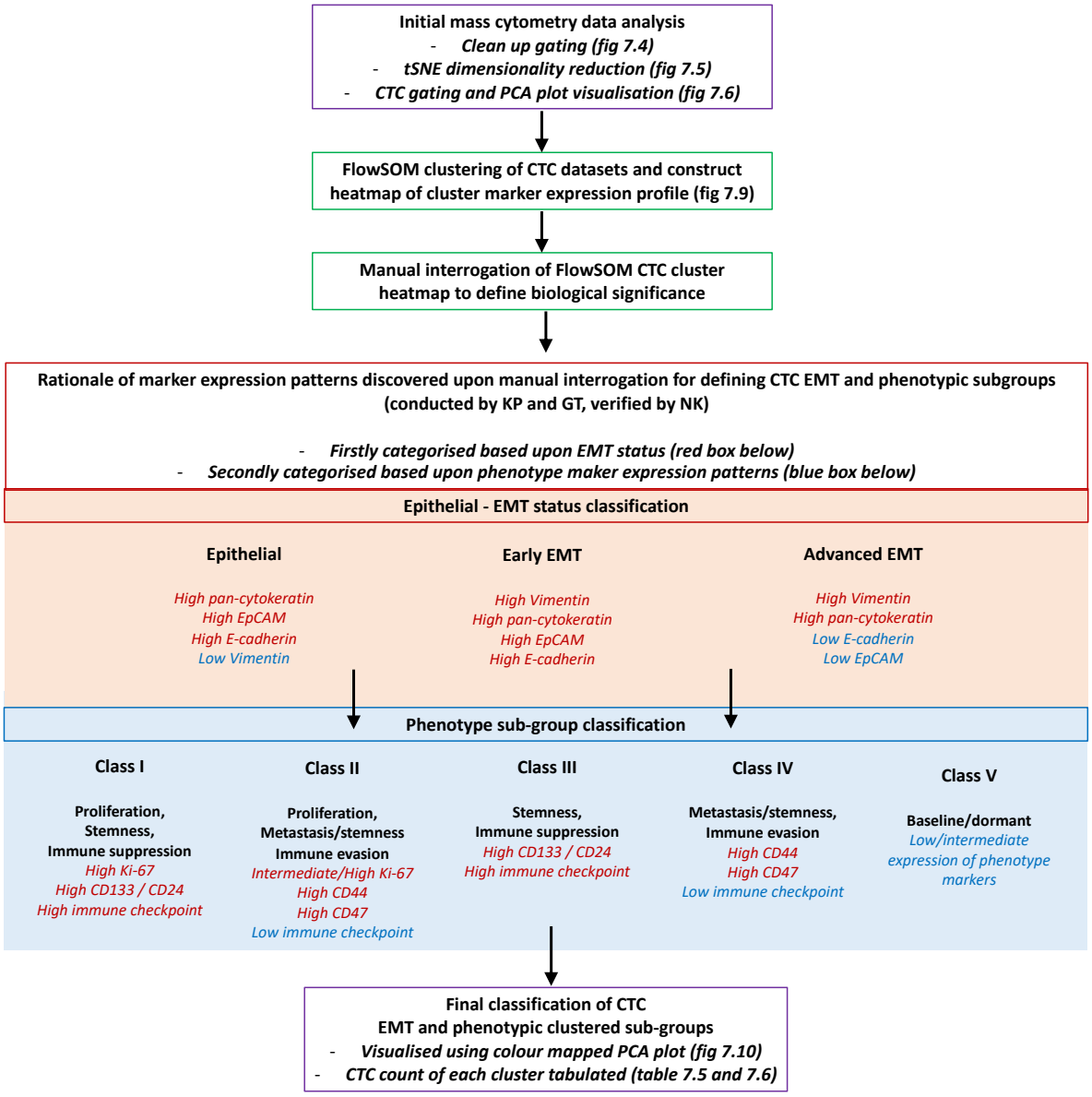


Figure 7.8: CTCs cluster into sub-groups based upon epithelial-EMT status and functional phenotype marker expression patterns. The above flow diagram demonstrates the data analysis pipeline and patterns of marker expression that defined EMT and phenotypic CTC sub-groups. Each cluster identified using FlowSOM was first categorised based upon EMT marker expression and then functional phenotype marker expression. Note, we labelled clusters separately based upon CD47 and immune checkpoint expression - high CD47 clusters being 'immune evasive' and immune checkpoint expression labelled as 'immune (suppression) low/high'. KP – Dr Karl Payne, GT – Dr Graham Taylor, NK – Dr Naeem Khan.

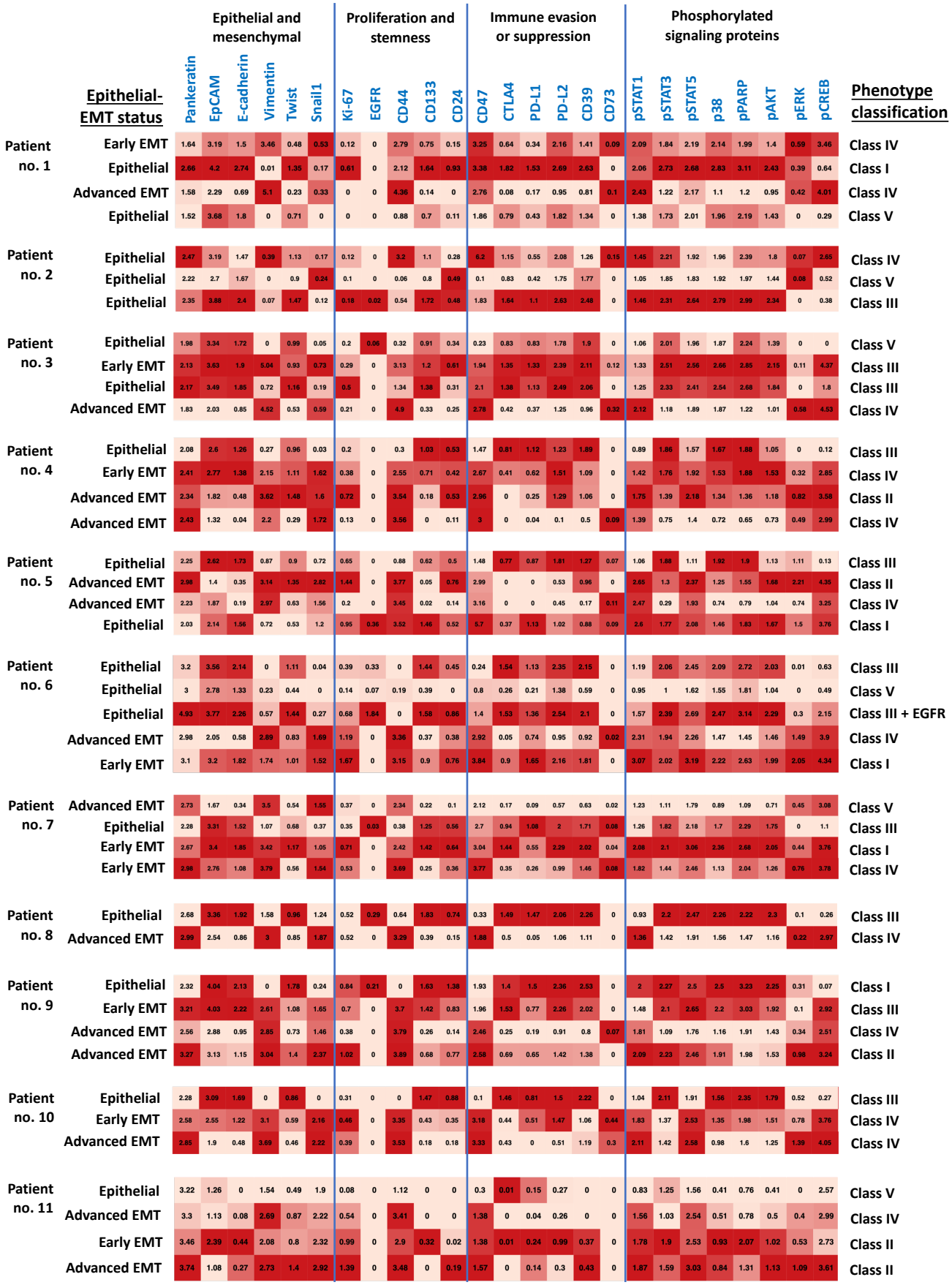

Epithelial and mesenchymal
Proliferation and stemness
Immune evasion or suppression
Phosphorylated signaling proteins
Epithelial-EMT status
Pankeratin
EpCAM
E-cadherin
Vimentin
Twist
Snail1
Ki-67
EGFR
CD44
CD133
CD24
CD47
CTLA4
PD-L1
PD-L2
CD39
CD73
pSTAT1
pSTAT3
pSTAT5
p38
pPARP
pAKT
pERK
pCREB
Phenotype classification
Patient no. 1
Early EMT 1.64 3.19 1.5 3.46 0.48 0.53 0.12 0 2.79 0.75 0.15 3.25 0.64 0.34 2.16 1.41 0.09 2.09 1.84 2.19 2.14 1.99 1.4 0.59 3.46 Class IV
Epithelial 2.66 4.2 2.74 0.01 1.35 0.17 0.61 0 2.12 1.64 0.93 3.38 1.82 1.53 2.69 2.63 0 2.06 2.73 2.68 2.83 3.11 2.43 0.39 0.64 Class I
Advanced EMT 1.58 2.29 0.69 5.1 0.23 0.33 0 0 4.36 0.14 0 2.76 0.08 0.17 0.95 0.81 0.1 2.43 1.22 2.17 1.1 1.2 0.95 0.42 4.01 Class IV
Epithelial 1.52 3.68 1.8 0 0.71 0 0 0 0.88 0.7 0.11 1.86 0.79 0.43 1.82 1.34 0 1.38 1.73 2.01 1.96 2.19 1.43 0 0.29 Class V
Patient no. 2
Epithelial 2.47 3.19 1.47 0.39 1.13 0.17 0.12 0 3.2 1.1 0.28 6.2 1.15 0.55 2.08 1.26 0.15 1.45 2.21 1.92 1.96 2.39 1.8 0.07 2.65 Class IV
Epithelial 2.22 2.7 1.67 0 0.9 0.24 0.1 0 0.06 0.8 0.49 0.1 0.83 0.42 1.75 1.77 0 1.05 1.85 1.83 1.92 1.97 1.44 0.08 0.52 Class V
Epithelial 2.35 3.88 2.4 0.07 1.47 0.12 0.18 0.02 0.54 1.72 0.48 1.83 1.64 1.1 2.63 2.48 0 1.46 2.31 2.64 2.79 2.99 2.34 0 0.38 Class III
Patient no. 3
Epithelial 1.98 3.34 1.72 0 0.99 0.05 0.2 0.06 0.32 0.91 0.34 0.23 0.83 0.83 1.78 1.9 0 1.06 2.01 1.96 1.87 2.24 1.39 0 0 Class V
Early EMT 2.13 3.63 1.9 5.04 0.93 0.73 0.29 0 3.13 1.2 0.61 1.94 1.35 1.33 2.39 2.11 0.12 1.33 2.51 2.56 2.66 2.85 2.15 0.11 4.37 Class III
Epithelial 2.17 3.49 1.85 0.72 1.16 0.19 0.5 0 1.34 1.38 0.31 2.1 1.38 1.13 2.49 2.06 0 1.25 2.33 2.41 2.54 2.68 1.84 0 1.8 Class III
Advanced EMT 1.83 2.03 0.85 4.52 0.53 0.59 0.21 0 4.9 0.33 0.25 2.78 0.42 0.37 1.25 0.96 0.32 2.12 1.18 1.89 1.87 1.22 1.01 0.58 4.53 Class IV
Patient no. 4
Epithelial 2.08 2.6 1.26 0.27 0.96 0.03 0.2 0 0.3 1.03 0.53 1.47 0.81 1.12 1.23 1.89 0 0.89 1.86 1.57 1.67 1.88 1.05 0 0.12 Class III
Early EMT 2.41 2.77 1.38 2.15 1.11 1.62 0.38 0 2.55 0.71 0.42 2.67 0.41 0.62 1.51 1.09 0 1.42 1.76 1.92 1.53 1.88 1.53 0.32 2.85 Class IV
Advanced EMT 2.34 1.82 0.48 3.62 1.48 1.6 0.72 0 3.54 0.18 0.53 2.96 0 0.25 1.29 1.06 0 1.75 1.39 2.18 1.34 1.36 1.18 0.82 3.58 Class II
Advanced EMT 2.43 1.32 0.04 2.2 0.29 1.72 0.13 0 3.56 0 0.11 3 0 0.04 0.1 0.5 0.09 1.39 0.75 1.4 0.72 0.65 0.73 0.49 2.99 Class IV
Patient no. 5
Epithelial 2.25 2.62 1.73 0.87 0.9 0.72 0.65 0 0.88 0.62 0.5 1.48 0.77 0.87 1.81 1.27 0.07 1.06 1.88 1.11 1.92 1.9 1.13 1.11 0.13 Class III
Advanced EMT 2.98 1.4 0.35 3.14 1.35 2.82 1.44 0 3.77 0.05 0.76 2.99 0 0 0.53 0.96 0 2.65 1.3 2.37 1.25 1.55 1.68 2.21 4.35 Class II
Advanced EMT 2.23 1.87 0.19 2.97 0.63 1.56 0.2 0 3.45 0.02 0.14 3.16 0 0 0.45 0.17 0.11 2.47 0.29 1.93 0.74 0.79 1.04 0.74 3.25 Class IV
Epithelial 2.03 2.14 1.56 0.72 0.53 1.2 0.95 0.36 3.52 1.46 0.52 5.7 0.37 1.13 1.02 0.88 0.09 2.6 1.77 2.08 1.46 1.83 1.67 1.5 3.76 Class I
Patient no. 6
Epithelial 3.2 3.56 2.14 0 1.11 0.04 0.39 0.33 0 1.44 0.45 0.24 1.54 1.13 2.35 2.15 0 1.19 2.06 2.45 2.09 2.72 2.03 0.01 0.63 Class III
Epithelial 3 2.78 1.33 0.23 0.44 0 0.14 0.07 0.19 0.39 0 0.8 0.26 0.21 1.38 0.59 0 0.95 1 1.62 1.55 1.81 1.04 0 0.49 Class V
Epithelial 4.93 3.77 2.26 0.57 1.44 0.27 0.68 1.84 0 1.58 0.86 1.4 1.53 1.36 2.54 2.1 0 1.57 2.39 2.69 2.47 3.14 2.29 0.3 2.15 Class III + EGFR
Advanced EMT 2.98 2.05 0.58 2.89 0.83 1.69 1.19 0 3.36 0.37 0.38 2.92 0.05 0.74 0.95 0.92 0.02 2.31 1.94 2.26 1.47 1.45 1.46 1.49 3.9 Class IV
Early EMT 3.1 3.2 1.82 1.74 1.01 1.52 1.67 0 3.15 0.9 0.76 3.84 0.9 1.65 2.16 1.81 0 3.07 2.02 3.19 2.22 2.63 1.99 2.05 4.34 Class I
Patient no. 7
Advanced EMT 2.73 1.67 0.34 3.5 0.54 1.55 0.37 0 2.34 0.22 0.1 2.12 0.17 0.09 0.57 0.63 0.02 1.23 1.11 1.79 0.89 1.09 0.71 0.45 3.08 Class V
Epithelial 2.28 3.31 1.52 1.07 0.68 0.37 0.35 0.03 0.38 1.25 0.56 2.7 0.94 1.08 2 1.71 0.08 1.26 1.82 2.18 1.7 2.29 1.75 0 1.1 Class III
Early EMT 2.67 3.4 1.85 3.42 1.17 1.05 0.71 0 2.42 1.42 0.64 3.04 1.44 0.55 2.29 2.02 0.04 2.08 2.1 3.06 2.36 2.68 2.05 0.44 3.76 Class I
Early EMT 2.98 2.76 1.08 3.79 0.56 1.54 0.53 0 3.69 0.25 0.36 3.77 0.35 0.26 0.99 1.46 0.08 1.82 1.44 2.46 1.13 2.04 1.26 0.76 3.78 Class IV
Patient no. 8
Epithelial 2.68 3.36 1.92 1.58 0.96 1.24 0.52 0.29 0.64 1.83 0.74 0.33 1.49 1.47 2.06 2.26 0 0.93 2.2 2.47 2.26 2.22 2.3 0.1 0.26 Class III
Advanced EMT 2.99 2.54 0.86 3 0.85 1.87 0.52 0 3.29 0.39 0.15 1.88 0.5 0.05 1.06 1.11 0 1.36 1.42 1.91 1.56 1.47 1.16 0.22 2.97 Class IV
Patient no. 9
Epithelial 2.32 4.04 2.13 0 1.78 0.24 0.84 0.21 0 1.63 1.38 1.93 1.4 1.5 2.36 2.53 0 2 2.27 2.5 2.5 3.23 2.25 0.31 0.07 Class I
Early EMT 3.21 4.03 2.22 2.61 1.08 1.65 0.7 0 3.7 1.42 0.83 1.96 1.53 0.77 2.26 2.02 0 1.48 2.1 2.65 2.2 3.03 1.92 0.1 2.92 Class III
Advanced EMT 2.56 2.88 0.95 2.85 0.73 1.46 0.38 0 3.79 0.26 0.14 2.46 0.25 0.19 0.91 0.8 0.07 1.81 1.09 1.76 1.16 1.91 1.43 0.34 2.51 Class IV
Advanced EMT 3.27 3.13 1.15 3.04 1.4 2.37 1.02 0 3.89 0.68 0.77 2.58 0.69 0.65 1.42 1.38 0 2.09 2.23 2.46 1.91 1.98 1.53 0.98 3.24 Class II
Patient no. 10
Epithelial 2.28 3.09 1.69 0 0.86 0 0.31 0 0 1.47 0.88 0.1 1.46 0.81 1.5 2.22 0 1.04 2.11 1.91 1.56 2.35 1.79 0.52 0.27 Class III
Early EMT 2.58 2.55 1.22 3.1 0.59 2.16 0.46 0 3.35 0.43 0.35 3.18 0.44 0.51 1.47 1.06 0.44 1.83 1.37 2.53 1.35 1.98 1.51 0.78 3.76 Class IV
Advanced EMT 2.85 1.9 0.48 3.69 0.46 2.22 0.39 0 3.53 0.18 0.18 3.33 0.43 0 0.51 1.19 0.3 2.11 1.42 2.58 0.98 1.6 1.25 1.39 4.05 Class IV
Patient no. 11
Epithelial 3.22 1.26 0 1.54 0.49 1.9 0.08 0 1.12 0 0 0.3 0.01 0.15 0.27 0 0 0.83 1.25 1.56 0.41 0.76 0.41 0 2.57 Class V
Advanced EMT 3.3 1.13 0.08 2.69 0.87 2.22 0.54 0 3.41 0 0 1.38 0 0.04 0.26 0 0 1.56 1.03 2.54 0.51 0.78 0.5 0.4 2.99 Class IV
Early EMT 3.46 2.39 0.44 2.08 0.8 2.32 0.99 0 2.9 0.32 0.02 1.38 0.01 0.24 0.99 0.37 0 1.78 1.9 2.53 0.93 2.07 1.02 0.53 2.73 Class II
Advanced EMT 3.74 1.08 0.27 2.73 1.4 2.92 1.39 0 3.48 0 0.19 1.57 0 0.14 0.3 0.43 0 1.87 1.59 3.03 0.84 1.31 1.13 1.09 3.61 Class II

Figure 7.9: Signal intensity heatmapping defines CTC sub-groups within patient samples. Heatmap (red = high intensity, white = low intensity) of median signal intensity for target markers (y-axis) plotted against CTC cluster (x-axis). Each row represents a separate cell group as identified from FlowSOM clustering. Rows are categorised based on epithelial-EMT status (left hand side end column) and phenotypic sub-group classification (right hand side end column). Note – to better demonstrate variance in expression between different markers and patients, the red signal intensity colouring is mapped separately between patient samples AND columns.

Despite being a transcription factor implicated in EMT, it was notable that Twist was predominantly over-expressed in the epithelial group, significantly greater than Snail expression (t-test, p=0.0054). As expected, EMT sub-groups expressed high levels of both Twist and Snail, although conversely Snail expression was proportionately higher across these EMT sub-groups (t-test, p=0.0141).

Following assessment of EMT markers, we evaluated markers of proliferation, stemness and immune evasion/suppression to assess the functional impact of EMT upon the phenotypic characteristics of CTCs. There was a clear separation of clustered sub-groups with high or low Ki-67 expression, a marker of cell proliferation. While this was not observed to be exclusive to either epithelial or EMT groups, the epithelial groups generally demonstrated intermediate expression levels and EMT groups had higher levels of Ki-67 expression. In addition, there appeared to be a correlation between high Twist and Ki-67 expression - with an EMT sub-group cluster observed to have markedly reduced Twist expression, which correlated with a decrease in Ki-67 expression. In all but one sample we observed minimal EGFR expression by CTCs. In addition, immune-related marker CD73 was expressed at a very low level.

The marker CD44 is regarded primarily as a cancer stem cell marker, but has also been shown to be involved in cell proliferation and survival, thus contributing to tumour growth and metastasis.[249] CD44 was highly expressed in all EMT sub-groups, almost exclusively compared to the epithelial cell groups which did not appear to express CD44 (t-test, $p<0.0001$).

The stemness markers CD133 and CD24 appeared to be co-expressed – a finding observed in both epithelial and EMT sub-groups. Over-expression of stemness markers was seen more frequently, and to a higher level, in epithelial sub-groups compared to EMT sub-groups. In addition, in the epithelial groups the expression of these stemness markers frequently correlated with increased Ki-67 expression. Within the EMT sub-groups, high stemness marker expression was observed almost exclusively in early compared to advanced EMT sub-groups.

Immune-related markers in our panel included CD47 and the immune checkpoint markers CTLA4, PD-L1, PD-L2, CD39 and CD73. On the whole, CD47 was observed to be separate and inversely expressed compared to the aforementioned immune-checkpoint markers and so was categorised separately. In part, this finding was due to a general decrease in expression of immune checkpoint makers in EMT compared to epithelial sub-groups. We labelled clusters separately based upon CD47 and immune checkpoint expression - high CD47 clusters (which was co-expressed with CD44) being ‘immune evasive’ and immune checkpoint expression labelled as ‘immune (suppression) low/high’. In analysis and discussion below CD47 expression is explicitly stated as such, and immune suppressing checkpoint inhibitors are grouped as ‘immune’. While some epithelial sub-groups also displayed decreased immune

checkpoint expression, it was notable that the advanced EMT subgroup had markedly reduced immune checkpoint expression compared to other sub-groups. In the epithelial cell groups, we observed a sub-group with high Ki-67, stemness and immune-related marker expression, i.e. a proliferative and immune-evasive (high CD47) CTC phenotype. There was no clear relationship between immune-related marker expression and proliferation marker expression in EMT sub-groups.

Using manual interrogation of median signal intensity for CTC phenotype markers from clustered heatmaps, we identified recurring patterns of marker expression that were common between patients. We defined 5 phenotypic sub-groups, labelled class I-V (figure 7.8). The Class I and class II sub-groups were proliferative cells with high Ki-67, separated by patterns of stemness (CD133 and CD24) marker expression (class I) or metastasis and immune evasion (CD47 and CD44) marker expression (class II). The class III sub-group was characterised by expression of stemness and immune checkpoint markers. Class IV was similar to class II, with high CD44 and CD47, but did not demonstrate increased Ki-67 proliferative marker expression. Finally, class V was labelled as a 'baseline or dormant' cell group, demonstrating no obvious increase in any phenotype or immune-related marker. Of note, classes I and III were 'immune (suppression) high' with expression of multiple immune checkpoint markers (CTLA4, PD-L1, PD-L2), and classes II, IV and V were 'immune (suppression) low'.

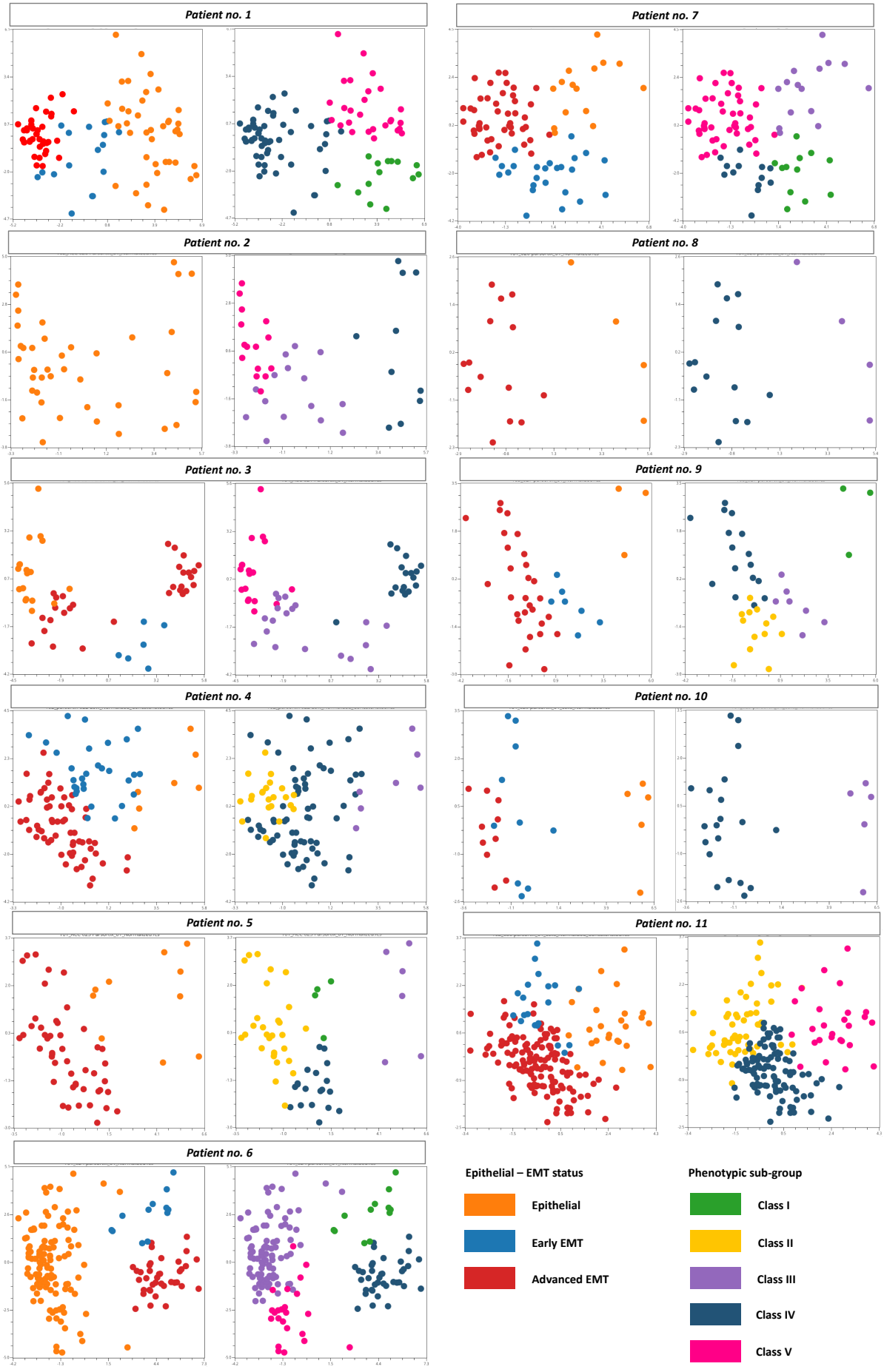

Patient no. 1
Patient no. 7
Patient no. 2
Patient no. 8
Patient no. 3
Patient no. 9
Patient no. 4
Patient no. 10
Patient no. 5
Patient no. 11
Patient no. 6
Epithelial – EMT status
Epithelial
Early EMT
Advanced EMT
Phenotypic sub-group
Class I
Class II
Class III
Class IV
Class V

Figure 7.10: Relationship between epithelial-EMT status and phenotypic classification for clustered CTC sub-groups. The above plots display clusters from FlowSOM analysis colour coded based upon epithelial-EMT status (left side) and phenotype classification (right side) onto PCA plots from figure 7.3. Relationships such as a predominance of advanced EMT class IV sub-groups or epithelial class III sub-groups can more easily be visualised.

Table 7.5: Epithelial–EMT status and phenotype classification correlate together. Numbers of CTC sub-groups (as identified from FlowSOM clustering and heatmap analysis in figure 7.7) plotted against epithelial-EMT status (rows) and phenotype classification (columns) are displayed.

	CLASS I	CLASS II	CLASS III	CLASS IV	CLASS V
EPITHELIAL	3	0	9	2	5
EARLY EMT	2	1	2	4	0
ADVANCED EMT	0	4	0	9	1

In all but 2 patients we observed the presence of all three EMT status sub-groups (epithelial, early and advanced EMT). One patient demonstrated epithelial only CTC sub-groups and one patient demonstrated epithelial and advanced EMT sub-groups. To highlight the relationship between the EMT status and phenotype classification of clustered sub-groups, we colour mapped the PCA plots from figure 7.6 (as shown in figure 7.10) and also tabulated correlations (table 7.5). Class III (stemness/immune suppression) and class V (baseline/dormant) sub-groups were almost exclusively found in epithelial cells. In contrast, the class IV sub-group (metastasis/immune evasion with low immune checkpoint) was clearly correlated to an EMT cell type, of which the majority were the advanced EMT sub-group. Of note, the advanced

EMT class IV sub-group was present in all but one patient sample. Epithelial CTCs were not present with a class II phenotype (proliferative with low immune checkpoint), in contrast to advanced EMT CTCs where the class I phenotype was not observed (proliferative with high immune checkpoint).

In a small cohort, skewed toward T4 advanced stage disease, formal statistical analysis of stage and CTC sub-group was not possible. In addition, there was no clear correlation between class I and II proliferative CTC sub-groups and the presence of nodal metastases. All patients (5/11) with a 'high' CTC burden (CTC count ≥ 9 cells/ml blood) demonstrated proliferative CTC sub-groups. In contrast those patients (3/11) with a 'low' CTC burden (CTC count <3 cells/ml blood) did not exhibit proliferative sub-group CTCs.

To objectively quantify the EMT status of CTCs in each sample, we calculated the proportion of EMT and epithelial CTCs within the total CTC population, presenting the value as a percentage (table 7.6). The proportion of EMT CTCs ranged from 0 to 94.2% (mean 66.9%), with the majority of patients exhibiting >50% EMT CTCs (8/11). No correlation was evident between increasing T stage or N stage and EMT ratio (Pearson's R=-0.327, p=0.328 and R=0.233, p=0.491 respectively). However, the number of proliferative CTCs (i.e. class I and class II subgroups) was positively correlated with EMT burden (Pearson's R=0.8865, p=0.000277). This finding is explained by the close relationship between EMT and class I/II sub-groups (table 7.5).

Table 7.6: Proportion of EMT, epithelial, proliferative and immune checkpoint high CTCs. For each patient the epithelial-EMT group and phenotype classification is shown along with the total number of CTCs within that group (total CTCs from each sample are shown in bold underneath each sample). The proportion of EMT, epithelial, proliferative (class I and II) and immune checkpoint high (class I and III) CTCs are shown as a percentage of the total CTC count.

Patient no.	*Immune high (class I/III) %*	*Proliferative (class I/II) %*	*Epithelial CTC %*	*EMT CTC %*	*Epithelial-EMT group*	*Class*	*CTC count*
1	14.4	14.4	42.3	57.7	Early-EMT	IV	17
					Epithelial	I	14
					Adv-EMT	IV	39
					Epithelial	V	27
					Total CTC count		**97**
2	35	0	100	0	Epithelial	IV	10
					Epithelial	V	16
					Epithelial	III	14
					Total CTC count		**40**
3	35.2	0	57.4	42.6	Epithelial	V	18
					Early-EMT	III	6
					Epithelial	III	13
					Adv-EMT	IV	17
					Total CTC count		**54**
4	7.1	22.4	7.1	92.9	Epithelial	III	7
					Early-EMT	IV	30
					Adv-EMT	II	22
					Adv-EMT	IV	39
					Total CTC count		**98**
5	17.2	56.9	17.2	82.8	Epithelial	III	6
					Adv-EMT	II	29
					Adv-EMT	IV	19
					Epithelial	I	4
					Total CTC count		**58**
6	47.3	7.2	70.1	29.9	Epithelial	III	67
					Epithelial	V	23
					Epithelial	III	27
					Adv-EMT	IV	38
					Early-EMT	I	12
					Total CTC count		**167**
7	22.4	14	5.8	94.2	Adv-EMT	V	46
					Epithelial	III	5
					Early-EMT	I	12

					Early-EMT	IV	13
					Total CTC count		**76**
8	22.2	0	22.2	77.8	Epithelial	III	4
					Adv-EMT	IV	14
					Total CTC count		**18**
9	25.6	41	7.7	92.3	Epithelial	I	3
					Early-EMT	III	7
					Adv-EMT	IV	16
					Adv-EMT	II	13
					Total CTC count		**39**
10	20	0	20	80	Epithelial	III	5
					Early-EMT	IV	10
					Adv-EMT	IV	10
					Total CTC count		**25**
11	0	30.6	14.2	85.8	Epithelial	V	26
					Adv-EMT	IV	101
					Early-EMT	II	20
					Adv-EMT	II	36
					Total CTC count		**183**

7.4.3 Phosphorylated signalling proteins underpinning CTC sub-groups

We then compared the CTC sub-group classification to the expression of phospho-signalling proteins, thus exploring the cell signalling activity associated with different CTC groups. There were clear patterns within each patient sample that certain sub-groups were separated due to expression of higher amounts of phospho-proteins. Across all samples, there was a trend that classes I-III demonstrated increased intracellular pathway activation. However, notable correlations in relation to individual markers and CTC sub-groups were evident. In all early and advanced EMT sub-groups, irrespective of phenotype classification, pCREB was highly expressed when compared to epithelial groups. In contrast, no clear relationship between the epithelial sub-group status and pathway activation was evident, with heterogenous

expression linked more closely to the phenotype sub-group classification. All class IV sub-groups demonstrated increased pSTAT1 expression, with a trend that class IV sub-groups also expressed high levels of pSTAT5. pSTAT1 was also highly expressed in the proliferative sub-group classes I and II. Compared to other phospho-protein levels, pERK expression appeared to vary between patients, but was highly expressed in class II and IV sub-groups – both of which were immune checkpoint low and almost exclusively early or advanced EMT groups. pSTAT3, pSTAT5, p38, pPARP and pAKT were observed to be co-expressed in individual sub-groups – more often classes I, II and III.

7.5 Comparison of mass cytometry to RNAseq CTC and tumour characterisation

7.5.1 Gene expression profiling of Parsortix enriched samples

To demonstrate the potential benefits of mass cytometry CTC characterisation, we compared the above findings to conventional methods of CTC gene expression profiling using bulk RNAseq analysis. Pre-treatment Transfix blood samples (taken at the same time as the mass cytometry analysed samples) were available for 10 of the patients analysed by mass cytometry. These samples were Parsortix enriched and underwent quantitative mRNA expression analysis (Lexogen QuantSeq) as per sections 4.10 and 5.6. For each patient, PBMC carrier cells were also sequenced as a negative control. Differential gene expression analysis compared the Parsortix enriched cell fraction from each individual patient sample to the group of 10 PBMC control samples. The rationale behind this approach, as employed by previous studies, is that the presence of elevated transcripts in the sample when compared

to the PBMC control, are presumed to be originating from CTCs in the enriched cell population. For each marker included in the mass cytometry panel, the results of differential gene expression were analysed and plotted as a heatmap, to allow the results of mass cytometry and transcriptomic analysis to be compared (figure 7.11).

Epithelial gene expression (keratin, EpCAM or E-cadherin), thus indicating the presence of CTCs, was observed in 9/10 samples. In all of these samples elevated keratin gene expression was observed, with E-cadherin and EpCAM elevated in 3 and 1 samples respectively. Three samples demonstrated Vimentin expression, thus identifying the presence of mesenchymal-transitioning CTCs. However, the EMT transcription factors Twist or Snail were not significantly expressed in any samples, and neither were the stemness markers CD133 or CD24, or target marker EGFR expressed in any sample. Expression of immune-checkpoint genes in RNAseq data was low, with only one sample demonstrating PD-L1 expression, and no significantly elevated CTLA4 or PD-L2 gene expression. Signalling protein gene expression was variable, with every gene expressed in multiple samples and no apparent pattern between signalling pathways. The AKT1 and ERK1 genes were observed to be more frequently expressed than other signalling genes (in 8/9 and 9/10 samples respectively).

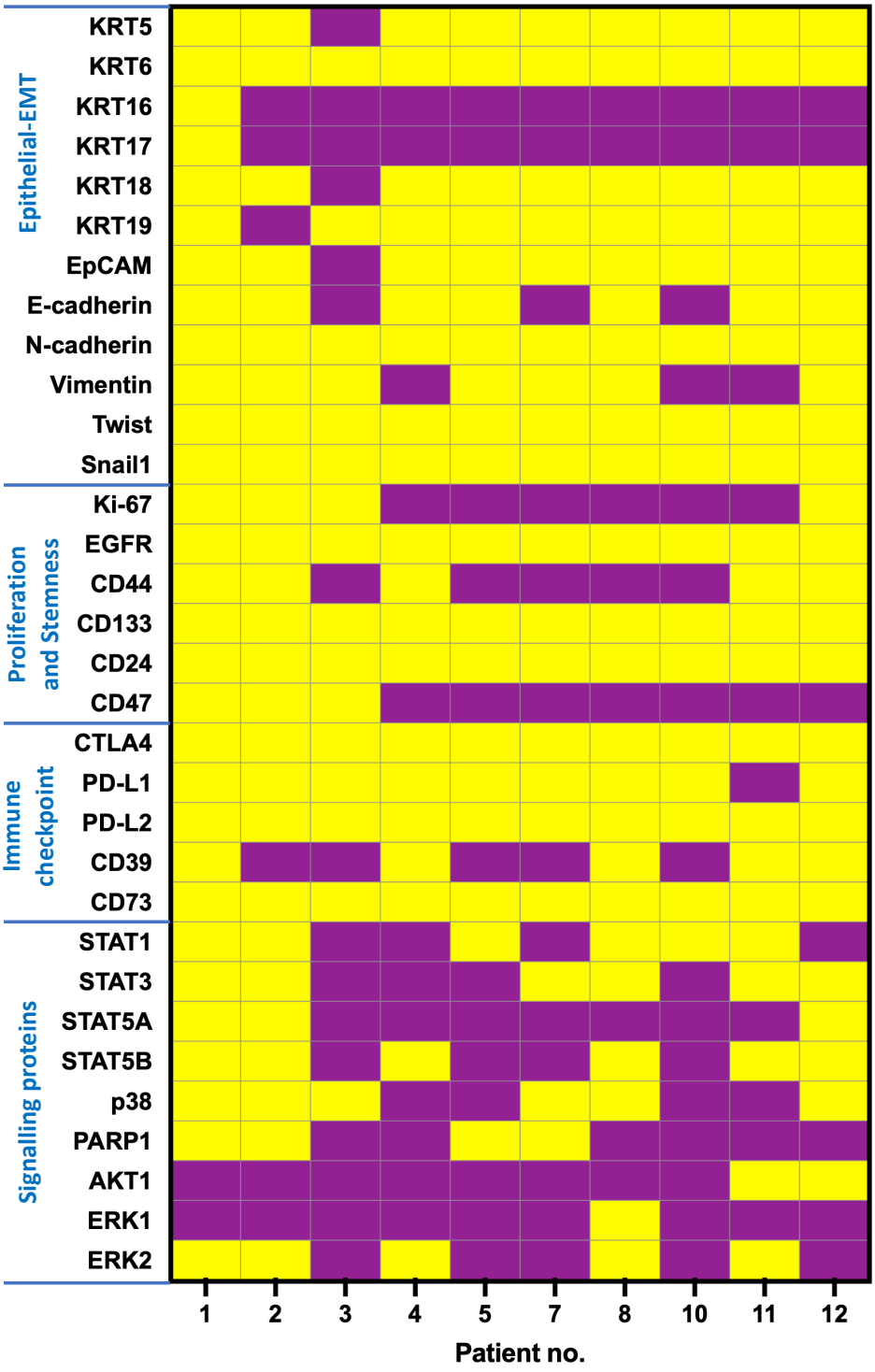


Figure 7.11: Quantitative mRNA expression analysis identifies CTC specific transcripts in patient samples. Differential gene expression (DGE) analysis comparing RNA extracted from Parsortix enriched blood samples to PBMC negative controls identifies genes with significantly elevated transcript counts in patient samples. Results from CTC target marker genes corresponding to the mass cytometry protein panel are shown listed in the left column (purple = significant DGE, yellow = not significant DGE).

When comparing these results to mass cytometry analysis we observe that the presence of CTCs in RNAseq correlated in 8/10 patients, with RNAseq failing to identify the presence of CTCs in one sample (patient 1). Of note, RNAseq identified elevated keratin gene expression (an epithelial CTC marker) in the sample from patient 12 – a patient with a T1N0 tumour in whom we identified only 4 cells meeting CTC criteria which was deemed to be too close to our cut-off confidence threshold and thus labelled as CTC negative. The expression of EMT related genes was low, with RNAseq failing to identify the presence of EMT CTCs in 5 patients when compared to mass cytometry EMT sub-groups.

Expression patterns (elevated or absent) of the proliferation marker Ki-67 correlated with the presence of class I or II proliferative CTC sub-groups in 6 patients. All three samples with elevated EMT marker Vimentin (patients 4, 10 and 11) also demonstrated elevated Ki-67 and CD47 – which would correlate with the class II CTC sub-group seen in mass cytometry analysis of patients 4 and 11. Patterns of co-expression of CD44 and CD47, as defining class II and IV phenotypic sub-groups in mass cytometry data, were also maintained in RNAseq data. Of those RNAseq samples with elevated CD44 or CD47 expression (8/10); all demonstrated a class IV sub-group in corresponding mass cytometry analysis (CTC positive patients excluding patient 12). Samples with elevated Ki-67, CD44 and CD47 gene expression appeared to demonstrate significantly elevated signalling protein gene expression, particularly the STAT, p38 and PARP pathways – similar to the pattern of ‘cell activation’ seen in mass cytometry analysis.

7.5.2 Correlating tumour EMT profile and CTC sub-groups

We performed mRNA quantitative expression profiling of available core and margin tumour samples from 10 patients in our cohort (of these, one patient had only core tissue available). Thirty-five genes were significantly differentially expressed when comparing margin and core tumour samples. Functional enrichment analysis of gene ontology gene sets revealed significant enrichment of the following domains, all involved in epidermal and keratinocyte pathways: *cornification, epidermis development, keratinization, keratinocyte development and epidermal cell differentiation*. Such a finding may be expected, given the tumour margin is a proliferating and invading front, in contrast to the poorly differentiated tumour core.

As per section 6.4.6, we clustered samples based upon the expression of the HSNCC EMT signature derived by Jung *et al.* Out of 19 samples, 7 clustered into a mesenchymal group (figure 7.12). Within this mesenchymal group were the core *and* margin samples from patients 5, 6 and 9, and the margin sample from patient 3. Neither CTC burden (total CTC count) nor EMT CTC burden (percentage of EMT CTCs) correlated with tumour mesenchymal status (t-test, p=0.485 and p=0.365 respectively). In addition, there was no obvious correlation between the presence of early and/or advanced EMT sub-groups with tumour mesenchymal status.

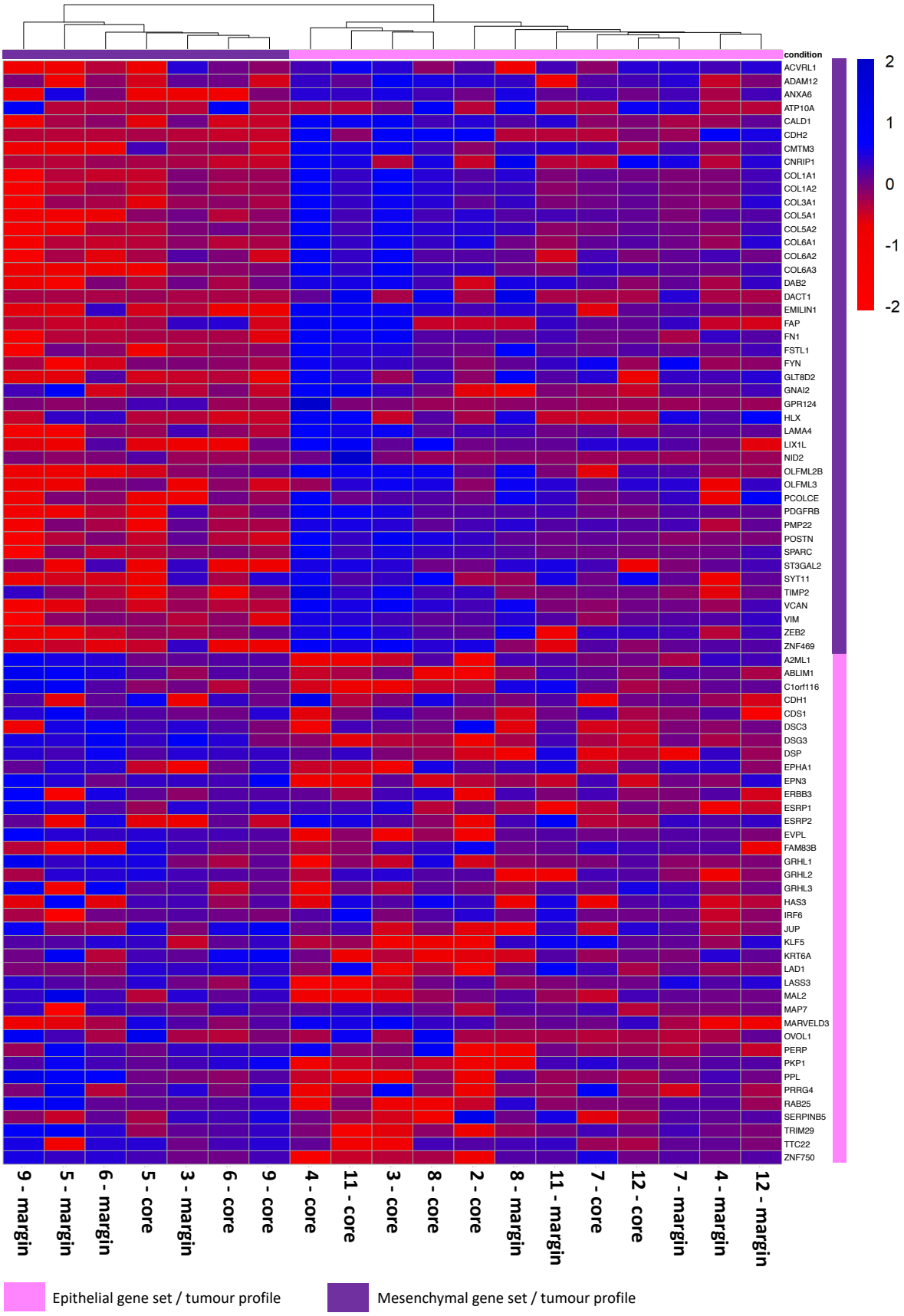

condition
ACVRL1
ADAM12
ANXA6
ATP10A
CALD1
CDH2
CMTM3
CNRIP1
COL1A1
COL1A2
COL3A1
COL5A1
COL5A2
COL6A1
COL6A2
COL6A3
DAB2
DACT1
EMILIN1
FAP
FN1
FSTL1
FYN
GLT8D2
GNAI2
GPR124
HLX
LAMA4
LIX1L
NID2
OLFML2B
OLFML3
PCOLCE
PDGFRB
PMP22
POSTN
SPARC
ST3GAL2
SYT11
TIMP2
VCAN
VIM
ZEB2
ZNF469
A2ML1
ABLIM1
C1orf116
CDH1
CDS1
DSC3
DSG3
DSP
EPHA1
EPN3
ERBB3
ESRP1
ESRP2
EVPL
FAM83B
GRHL1
GRHL2
GRHL3
HAS3
IRF6
JUP
KLF5
KRT6A
LAD1
LASS3
MAL2
MAP7
MARVELD3
OVOL1
PERP
PKP1
PPL
PRRG4
RAB25
SERPINB5
TRIM29
TTC22
ZNF750
2
1
0
-1
-2
9 - margin
5 - margin
6 - margin
5 - core
3 - margin
6 - core
9 - core
4 - core
11 - core
3 - core
8 - core
2 - core
8 - margin
11 - margin
7 - core
12 - core
7 - margin
4 - margin
12 - margin
Epithelial gene set / tumour profile
Mesenchymal gene set / tumour profile

Figure 7.12: Tumour samples cluster into mesenchymal and epithelial profiles. Heatmap of gene expression of the HNSCC EMT signature for core and margin samples. Samples are hierarchically clustered to reveal mesenchymal versus epithelial profile tumours. y-axis - mesenchymal genes are above and epithelial genes below, bottom x-axis – patient number and core or margin tumour site denoted, top x-axis – samples clustered based upon mesenchymal (left) and epithelial (right) expression patterns.

However, there was a notable trend that patients with mesenchymal profile tumours exhibited proliferative CTC sub-groups, when compared to epithelial profile tumours. Those patients with a mesenchymal profile in both core and margin samples all had proliferative class I or class II CTC sub-groups present. There were only 2/12 patients who demonstrated both class I and class II proliferative sub-groups (patients 5 and 9) and both of these were mesenchymal profile tumours. Tumours with a high mesenchymal profile (both core and margin) had a significantly increased proliferative CTC burden (percentage of proliferative CTCs) compared to epithelial tumours (t-test, p=0.0237). Tumours with a mesenchymal profile in either core or margin samples exhibited a significantly higher proportion of 'immune high' CTCs (t-test, p=0.0470) compared to epithelial tumours.

7.6 Exploring pan-proteomic CTC analysis

As a final avenue of research, we explored the idea of using mass spectrometry to characterise Parsortix enriched CTCs. While mass spectrometry doesn't achieve single-cell data resolution, it can assess the entirety of the proteome, far beyond the depth of information that mass cytometry can provide if one is analysing bulk CTC profiles. A select few studies have highlighted the role mass spectrometry may have in CTC characterisation, demonstrating success in spiked cell line models.[250], [251] However, no study has demonstrated mass

spectrometry of fixed CTC blood samples. Fixation of proteins is known to impact the quality of data from mass spectrometry proteomic studies.[252] However, we favour immediate fixation of blood samples to 'freeze' the CTC phenotype and prevent altered expression due to sample storage or processing. Therefore, we collaborated with Dr Tony Ly and Professor Doreen Cantrell from Dundee University – distinguished experts in mass spectrometry - who have optimised a protocol to perform mass spectrometry on fixed cells.[252]

No published reports detailing the compatibility of Transfix fixation and mass spectrometry exist. Therefore, as a test case unfixed PBMCs and Transfix fixed PBMCs from the same healthy donor (n=2) (methods section 4.12) were compared. Despite fixation, the number of identified proteins and the correlation of protein abundance (i.e. the count for individual proteins) was excellent across both replicates (figure 7.13a, Pearson correlation 0.75 and 0.73, p=<0.0001), thus demonstrating that mass spectrometry analysis was possible on Transfix fixed cells. Following this result, using our previous CTC model, healthy donor blood (collected in a Transfix blood tube) was spiked with the FaDu cell line and then enriched using the Parsortix device. Using mass spectrometry, the protein abundances from this test sample was compared to three controls: an unfixed FaDu control, Transfix PBMC control and a Parsortix enriched Transfix blood sample from the same donor without spiked epithelial cells (Parsortix control). While protein levels were decreased compared to the FaDu positive control (due to the contaminating PBMCs within the enriched cell population) we were able to successfully detect epithelial cell specific proteins not present in the PBMC and Parsortix control samples (figure 7.13b). We present an example of three proteins- keratin 18, EpCAM and Ki-67, that could be used for CTC detection and characterisation in a patient sample. Thus, this exploratory experiment served as proof-of-principle for potential follow-up research.

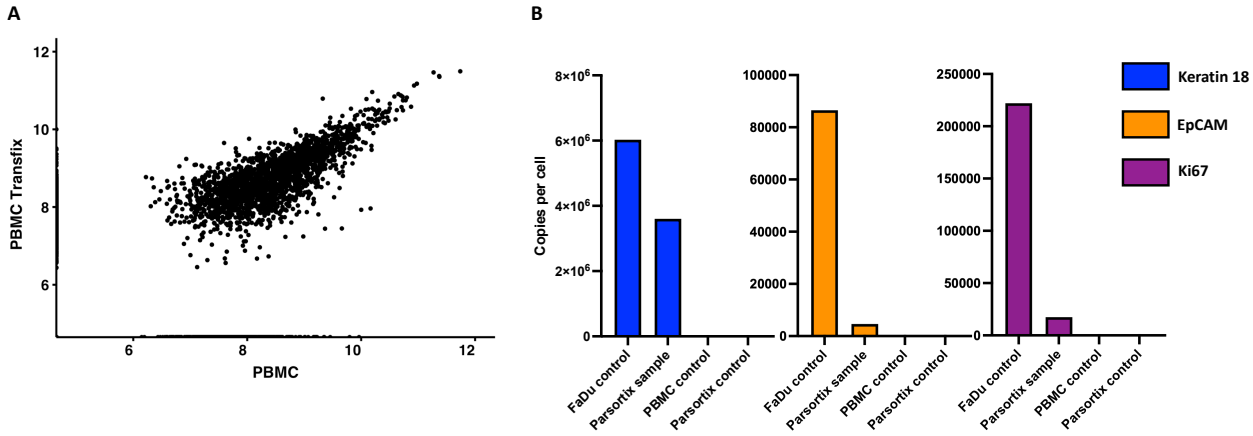


Figure 7.13: Mass spectrometry can identify proteins from Transfix fixed spiked cells enriched using Parsortix.

A) Dot plot of individual protein abundance in unfixed PBMC (x-axis) and Transfix fixed PBMC (y-axis) samples (from replicate 1) – demonstrating successful mass spectrometry on Transfix fixed cells. B) Bar chart plotting normalised protein levels as 'copies per cell' for CTC target markers keratin 18, EpCAM and Ki-67 – CTC markers are detected in an epithelial cell (FaDu) spiked Transfix donor blood sample which was enriched with Parsortix (Parsortix sample), while not being detectable in PBMC and Parsortix control samples.

7.7 Summary of key findings

- *In a test cohort, mass cytometry was able to identify CTCs from HNSCC patient blood samples and provide a means of multi-parameter characterisation down to the single cell level*
- *While there is considerable heterogeneity of marker expression, CTCs appear to cluster into distinct sub-groups based upon epithelial-EMT status and phenotype marker expression*
- *Patterns exist with regard the relationship between epithelial-EMT status and functional phenotype marker expression that define sub-groups of CTCs common between patients*
- *Mass cytometry can identify post-transcriptional modification of signalling proteins, in the form of phosphorylation, that may provide insight into the activated pathways within CTC sub-groups*
- *In a test cohort, mass cytometry allowed a higher level of multi-parameter data analysis than bulk gene expression profiling of CTCs – allowing more accurate CTC identification and cell phenotype/signalling pathway characterisation*
- *In a test cohort, the tumour mesenchymal status did not appear to correlate with CTC EMT burden – however such a finding would need corroboration in a larger cohort well designed clinical study*
- *In a test cohort, a tumour mesenchymal gene expression profile was significantly associated with a higher proportion of proliferative and immune-checkpoint high CTCs.*

CHAPTER 8: DISCUSSION

8.1 Summary of thesis findings

The final chapter of this thesis will summarise and discuss our findings. Finally, we will present a forward look of potential avenues of future research with regard CTC characterisation - discussing the technical feasibility and cost analysis of translating CTC-based biomarkers to clinic. In addition, contrasting CTCs to an alternative liquid biopsy compartment, that of circulating tumour DNA (ctDNA).

Summary

- *Parsortix microfluidic enrichment achieves a capture rate of >50% of CTC-like cells in a HNSCC model*
- *Immediate blood sample fixation increases cell capture and allows sample storage up to 72hrs, while protecting against phenotype alterations from microfluidic enrichment.*
- *Cytometric analysis of enriched blood samples can characterise and quantify CTC phenotypic sub-groups in HNSCC*
- *The multi-parameter characterisation afforded by mass cytometry outperforms flow cytometry and allows high dimensional CTC characterisation and interrogation of activated signalling pathways*
- *CTCs are present with heterogenous and diverse patterns of EMT and functional phenotype marker expression which point to novel CTC sub-groups common between patient samples*

- *In a small test cohort, mass cytometry was superior to bulk gene expression profiling of enriched CTCs, improving CTC detection rates and characterisation of phenotypic sub-groups*
- *In a small test cohort, the EMT profile of CTCs does not appear to correlate with tumour mesenchymal status, however patients with a mesenchymal tumour profile have increased proliferative and immune-checkpoint high CTCs – such findings require corroboration in a larger cohort clinical study*
- *The identification of proliferative and EMT CTC sub-groups with high expression of immune checkpoint markers creates the potential for therapeutic targeting of these highly metastatic cells*

8.2 Parsortix demonstrates clinical utility in a HNSCC cell line model

Due to the aforementioned advantages of marker-independent enrichment, microfluidic technology has shown promise to provide a reliable and standardised method of CTC enrichment - with multiple commercial platforms available.[79] Our findings from cell line optimisation experiments have demonstrated the potential of the Parsortix platform for CTC enrichment in HNSCC – adding to previous reports.[177] In addition, we have provided the first evidence of using fixation/preservative BCTs to improve microfluidic CTC capture rates in HNSCC. Our cell line optimisation model demonstrated a mean capture rate of 53.5% for unfixed cells, which is comparable to similar studies in other cancer types. [170] Of note, in the study by Hvichia *et al*., [170] the smaller bladder cancer cell line (18μm cell size) demonstrated the lowest capture rate of 42%, which is less than our capture rate, despite a

mean cell size of 12.2μm across our 2 cell lines. This disparity may be attributed to the differing bio-physical properties of cancers and cell lines and clearly demonstrates why specific optimisation and validation is required for each cancer type to be investigated.

CTCs are reported to be fragile cells outside of their native environment, with an estimated half-life in the region of 30 mins in the circulation.[91] Therefore, the method of cell preservation (i.e. blood sample collection) and time to processing is an important consideration in CTC research. Numerous BCTs have been utilised in research to-date, as discussed in section 1.5.4, with no clear evidence of the benefit of preservative BCTs (other than the Transfix BCT) to improve cell recovery rates.[253] To-date, only one previous paper has assessed Parsortix CTC enrichment and a CTC preserving BCT. In a breast cancer cell line model, Koch *et al.* compared capture rates of Parsortix enriched spiked blood samples in EDTA BCTs (processed at 0hrs) and CellSave, Streck and Transfix BCTs (processed at 24hrs).[254] They observed that the Transfix BCT was comparable to baseline EDTA collected samples (64% vs 60.7%), but far superior to the CellSave and Steck preservative BCTs (16.7% and 22.7% respectively). However, this group did not evaluate EDTA and Transfix BCTs at timepoints longer than 0 and 24hrs respectively, and we argue that examining capture rates at different timepoints for different BCTs creates inherent data bias.

We demonstrate that across all timepoints Transfix BCTs returned a significantly increased cell capture rate compared to EDTA BCTs, with a marked difference at 72hrs (73% vs 6%, p=0.0016). Such data is of crucial importance if samples are being collated (or collected) as part of multicentre clinical trials and time to processing is increased due to sample transport. Storage of unfixed samples has the potential to introduce significant bias if CTC enumeration

is used as a biomarker. Furthermore, it is unknown if CTC phenotypic subsets (i.e. epithelial vs mesenchymal) have differential sensitivities to cell storage, which would further impact data reliability in unfixed stored samples. We have observed in the literature that such key data of sample collection and time to processing is often overlooked or underreported. In the majority of studies, a non-fixative BCT is used and the time to processing is either not stated or defined as a range, for example 'within 24hrs of collection', with no evidence of validation for timepoints within this range. Clearly there are practical elements to consider with regard sample collection, transport and processing within larger hospitals and research institutions. However, if our cell line data is extrapolated to CTC counts then the drop-off in cell preservation and capture rate in EDTA BCTs between 0hrs and 24hrs has significant implications when using evidence from previous reports to inform future translational studies. In our experiments approximately 50% of non-fixed spiked cells were not captured, in contrast to a 30% cell loss in fixed samples. The question arises as to whether these non-fixed cells are not captured due to the inherent sensitivity of the Parsortix platform, or are they are lost due to microfluidic stress and trauma? We do not have the data to confirm this and most likely it results from a combination of both factors. However, given our data of a cell stress response in unfixed cells before and after microfluidic enrichment (section 5.4) and the observation of epithelial 'CTC fragments' in patient samples (section 6.4) we would hypothesise that loss of unfixed cells due to the physical forces of microfluidic enrichment is a major contributing factor. The phenomenon of cell loss in microfluidic enrichment has not been previously investigated in the literature. However, it should be noted that despite a cell loss of 30% with Parsortix, if a consistent and reproducible protocol of CTC enrichment and characterisation is used, then these margins of error are standardised between samples and could be corrected for in any future clinical trial/translation.

RNA degradation within CTCs has been shown to occur within a matter of hours.[99] Furthermore, physiological stress will alter the gene expression of cells and what is thought to be a snapshot of a CTC expression profile may actually be a biased result due to the sample handling/processing protocol. While multiple studies have evaluated RNA yield and gene expression of CTCs at various processing timepoints, very few have compared gene expression before and after CTC enrichment. Powell *et al*. assessed the impact of the MagSweeper immunomagnetic CTC enrichment protocol on the expression profile of 15 target genes in a breast cancer cell line;[101] observing no variation in gene expression before and after enrichment. To our knowledge, our study is the first to present gene expression data before and after microfluidic enrichment. Our findings indicate that during microfluidic enrichment, unfixed cells (i.e. those collected in an EDTA BCT) are downregulating RNA and ribosomal protein activity - moving from a rapidly dividing phenotype towards a quiescent like state.[255] Such a shift in cell cycle is a potential sign of a cell responding to stress or trauma.[256] Processing of cells in a microfluidic system over several hours subjects them to significant mechanical and oxidative stress. Fluid shear stress in epithelial cells subjected to microfluidic conditions has been shown to induce a DNA damage response and phenotype shift with increased EMT marker expression.[257] We observed a clear pattern of upregulation of genes known to be related to an oxidative stress response, but no significant change in expression of EMT related genes. Hypoxia-related gene signatures for both prognostication and treatment stratification of hypoxia-modifying therapy have shown promise in multiple cancer types.[258] In HNSCC, a hypoxia-immune signature has shown prognostic value and also potential for guiding targeted immunotherapies.[259] Therefore, any enrichment protocol that causes tumour cell hypoxia has potential for introducing

significant bias if such gene signatures are utilized for treatment stratification, further emphasising the argument for using fixative BCTs in CTC enrichment protocols.

We report that keratins 17 and 19 were significantly downregulated in unfixed microfluidic enriched epithelial cells. Cytokeratins have a key role in epithelial cell protection from mechanical and also non-mechanical stress (such as hypoxia), and in such conditions undergo reorganization and upregulation or downregulation of keratin proteins - which may explain the altered keratin expression seen in our cell line after microfluidic enrichment.[260] The clinical significance of such data could be profound, particularly if specific cytokeratins are being targeted for CTC identification or prognostic stratification. However, cytokeratins are abundant proteins in epithelial cells with high baseline expression, therefore a decrease in expression of certain cytokeratins (even if significant) may be acceptable for detection purposes when using antibodies targeting multiple cytokeratin proteins. Furthermore, our cell line samples were compared at baseline (0hrs time to processing) and it is unknown if prolonged storage of unfixed samples could cause further keratin dysregulation – a variable which should be internally validated in other enrichment protocols. Aside from keratin genes, expression of druggable targets such as EGFR or PD-L1 were not significantly altered in the enriched cell line, a key finding if unfixed CTCs are used as a source of potential predictive/therapeutic biomarkers.

Numerous protocols have been reported that detail complex lab-based methods to stabilise CTCs prior to (microfluidic) enrichment.[99] Our argument for using a commercially available fixative BCT is two-fold – firstly, the blood sample is fixed at point of collection, minimising CTC metabolic changes occurring during transport, and secondly it allows ease of collection

by clinical staff in a hospital environment and minimises laboratory time. Furthermore, the ability to reliably store blood samples while preserving CTC phenotypic integrity is paramount when CTC research is upscaled to larger cohort multi-site clinical trials. Previous studies have highlighted difficulties when attempting to RNA sequence CTCs that have been fixed/preserved.[253] Therefore, we wanted to provide proof-of-concept that RNA could be extracted from Transfix fixed cells and gene expression profiling was achievable. Epithelial cell transcripts from a Parsortix enriched population could be identified in Transfix preserved blood samples to confirm the presence of rare CTCs - using a 'bulk' sequencing method as described in previous studies.[130], [135] Considering the evidence above, we made the decision to switch exclusively to Transfix BCTs for further sample collection and CTC analysis.

8.3 Flow cytometry characterises epithelial and EMT CTC sub-groups

Previous reports have discussed the use of flow cytometry to detect and enumerate CTCs in multiple cancer types,[261]–[263] including HNSCC.[264] In all of these protocols the authors used an immunomagnetic epithelial cell enrichment step to reduce the number of cells needed to be analysed by flow cytometry. As discussed, enrichment for epithelial marker expression introduces potential bias against detecting epithelial low or mesenchymal CTCs. The utilisation here of Parsortix microfluidic 'size-dependent' pre-enrichment sought to remove this initial bias. More recently, Lopreseti *et al*. demonstrated the feasibility of removing the enrichment step entirely and performing flow cytometry on red cell lysed blood (albeit only 3ml of whole blood).[265] By prolonging the period of flow cytometry on a greater number of cells (up to 1hr compared to <5 minutes after pre-enrichment) the authors

reported a sensitive tool that actually reduced the processing time due to the removal of the, often lengthy, enrichment step. In theory, other than the capacity of the enrichment cassette, there is no limit on the volume of blood that Parsortix could enrich – up to 20ml as reported by the manufacturer. Therefore, if CTC counts are low to start with, 3ml may be inadequate to recover enough cells to perform meaningful multi-omic downstream analysis. In patient samples deemed to be CTC-positive we reported a range of 2-11.2 CTCs/ml whole blood – thus a reduction in starting blood volume would greatly impact the cell yield.

As discussed in the introduction, using CTCs as diagnostic biomarkers of disease burden and prognostic biomarkers for survival outcomes has been extensively researched by multiple groups. In our cohort of 20 patients, 13 (65%) demonstrated evidence of CTCs, which is comparable to previous data.[93] The presence of CTCs positively associated with advanced (III/IV) overall stage of disease ($p=0.0121$) and association with nodal metastases trended towards significance ($p=0.0850$). Previous reports have demonstrated an N stage of ≥ 2 is associated with CTC presence[264] – in our cohort it was noteworthy that 6 out of 7 patients with an N stage ≥ 2 were CTC positive. It would seem logical from a biological and mechanistic standpoint that a tumour with greater burden of disease and metastatic potential i.e. larger and with nodal metastases, would produce CTCs or increased numbers of CTCs. However, the published data is equivocal in this regard and meta-analyses in HNSCC have not been able to correlate TNM or overall stage of disease with CTCs.[107], [266] Such findings potentially argue for inherent tumour molecular characteristics i.e. increased EMT and mesenchymal expression, having a greater contribution to CTCs than purely just the TNM stage of the tumour.

While the majority of published data remains heterogenous and from small cohorts, in general, the presence of CTCs appears to correlate with increased rates of recurrence and decreased survival, although evidence is far from conclusive.[93], [107] While we observed a trend for decreased OS in epithelial CTC positive patients, we were unable to demonstrate significant associations between epithelial or EMT CTC presence and clinical outcomes. Possible explanations for the lack of association with clinicopathological and outcome variables, as observed in other studies, may be attributed to a small patient cohort. We accept this is a limitation, but unfortunately was unavoidable due to the COVID-19 global pandemic halting sample collection. Our cohort size was intended to provide proof-of-concept with regards validating a novel CTC enrichment technique in HNSCC. As the likely effect size was not known prior to my study a formal power calculation was not performed. Armed with this new information, an appropriately powered study to validate these preliminary findings can now be designed. The limited number of samples analysed longitudinally in the post-treatment phase provided a glimpse at the possible utility of CTCs to assess treatment response and detect recurrence in HNSCC, as demonstrated in other cancers.[63] Of particular interest would be the multi-parameter characterisation of CTCs at different timepoints in patients with recurrence, thus seeking to identify temporal ITH from liquid biopsy samples.

The expression of mesenchymal markers by CTCs and the EMT composition of the CTC population in our cohort was of particular interest and several noteworthy trends were observed. The presence of EMT CTCs was seen almost exclusively in advanced stage disease patients and we observed a strong trend between the ratio of EMT to epithelial CTCs and N-stage (Pearson R = 0.4125, P= 0.0642), i.e. patients with a higher N-stage exhibited an

increased mesenchymal profile within their CTC population. Such data suggests that the invasive and metastatic potential of a tumour may potentially be linked to the mesenchymal profile of the CTCs. Much of the biology of CTCs remains poorly understood and largely derived from in-vitro and/or in-vivo animal modelling, with several key questions still unanswered. Primarily - *how do CTCs get into the circulation and how do they change once they are there*? The mechanism of CTC intravasation is thought to be a combination of passive and active migration, with the latter occurring via an EMT to an invasive and motile mesenchymal phenotype.[70] Passive migration is thought to occur when a blood vessel (or lymphatic vessel) is compromised by an invading tumour.[70] The presence of CTCs with both epithelial only and/or EMT phenotypes (as seen in our data and previous studies) is an interesting phenomenon. It raises the question of whether this profile is representative of the tumour, is a product of EMT required for intravasation or is an adaptive function for CTC survival within the circulation. We found no significant relationship between tumour mesenchymal profile and CTC mesenchymal (EMT) profile in our cohort. However, we noted that as the CTC burden increased so did the proportion of EMT CTCs within the CTC population. In addition, EMT CTCs were only observed in the presence of epithelial CTCs. Such data may point towards an EMT being an adaptive function for tumour cells surviving within the circulation as it is increasingly seen in higher CTC burden patients.

EpCAM expression is observed to be downregulated in tumour cells during EMT,[267] including evidence for the same phenomenon in CTCs.[268] As will be discussed in the following chapter, EMT cannot be described based upon a single pathway only, and the interaction of so-called 'EMT master gene' transcription factors (Twist, Slug and Snail) and epithelial marker pathways (including EpCAM, cytokeratins and EGFR) is both complex and

dynamic.[269] If EMT CTCs hold particular promise regards prognostication and biomarker development, then from a practical viewpoint the enrichment and characterisation of this phenotypic sub-group poses several challenges. Strategies that use single marker identification (as is the current standard and used by ourselves in this chapter) may underestimate CTC burden or miss smaller more subtle sub-groups entirely. In this respect, single-cell multi-parameter characterisation is essential.

We found flow cytometry to be a viable tool to identify and enumerate CTCs. However, there are several shortcomings worth discussing. Flow cytometry is predominantly used to identify medium to large sized cell populations (i.e. those in the hundreds to thousands) and/or shifts in population marker expression. Data analysis and the 'clean-up of data' to remove debris and doublets is reliant upon manual gating which is subjective in nature with a potential margin of error. This risk is generally accepted as small when larger populations are analysed. However, this risk is substantial when seeking to identify rare populations of CTCs, in some cases as small as 10 or fewer, as demonstrated by the finding of low numbers of events meeting CTC criteria in healthy donors. This risk is further compounded by the lack of an absolute positive control for CTCs – as their precise makeup is seen to be heterogenous and variable between patients and different cancers, and issues such as downregulation of epithelial markers in EMT CTCs makes positive gating difficult. Given the limits of immunofluorescence and spectral overlap in flow cytometry, we sought to utilise mass cytometry to optimise a method of single-cell multi-parameter characterisation, as discussed below.

8.4 Mass cytometry identifies previously unseen CTC heterogeneity in HNSCC

We observed significant intra- and inter-patient heterogeneity of CTCs enriched from blood samples and characterised using mass cytometry – both in terms of EMT status, phenotype and intracellular signalling. With this in mind we will discuss our three key marker domains (EMT, phenotype, signalling) in turn, while also highlighting noteworthy correlations between these domains. We will demonstrate how, with the benefit of mass cytometry characterisation, existing knowledge of EMT activation and signalling pathways in CTCs can be compared and contrasted to data from existing HNSCC tumour analysis. Finally, we will discuss the comparison of mass cytometry and gene expression profiling of pooled CTCs and patient matched tumours, to highlight the benefit of single-cell mass cytometry analysis over existing methods of CTC characterisation.

8.4.1 CTC EMT plasticity

While current evidence demonstrates that CTCs can express various mesenchymal markers indicative of an EMT, no report has provided in-depth single-cell characterisation to identify the relationship between epithelial and mesenchymal markers that we observed. The concept of 'epithelial-mesenchymal plasticity' in tumour cells has recently been coined, to explain observations that an EMT is not simply a binary two stage process but a sequential acquisition of some, but potentially not all mesenchymal traits (and also a reversible process).[270] Furthermore, these mesenchymal transitioned tumour cells may retain specific epithelial cell features, which Bhatia *et al*. describe as '*dynamic and intermediate transitional states*'.[270]

This so-called 'hybrid EMT' state of epithelial *and* mesenchymal marker expression has been linked to increased tumour progression and metastasis.[271] Our findings indicate this concept also applies to CTCs. We observed cell groups that maintained epithelial marker expression, despite also expressing mesenchymal markers. While we categorised cells as predominantly epithelial and early/advanced EMT sub-groups, there was some degree of epithelial and mesenchymal marker plasticity among all sub-groups. For example, the specific pattern of high Twist (but not Snail) expression in our epithelial sub-groups. The current model of EMT would not expect CTCs with high expression of three epithelial markers (pan-cytokeratin, EpCAM and E-cadherin) to be expressing a mesenchymal transcription factor.

Twist has been observed to be highly expressed in HNSCC and is associated with tumour progression, local and distant metastasis and recurrence.[272] In addition, experimental evidence has demonstrated mechanisms whereby Twist promotes cancer cell proliferation.[273], [274] Therefore, the positive correlation of Twist and Ki-67 expression in proliferative CTC sub-groups (both epithelial and early/advanced EMT) in our data was interesting. Particularly in epithelial sub-groups, Twist expression may represent a pathway separate to EMT activation that is involved primarily to induce a CTC proliferative phenotype. This may explain heterogeneous expression of Twist within the advanced EMT sub-group (a group which appears already programmed in a mesenchymal-like state) and its association with Ki-67 expression to create proliferative EMT class I/II sub-groups. Alternatively, such epithelial CTCs with high Twist expression may already be in the aforementioned dynamic transitional state, being 'primed' to enter a mesenchymal phenotype.

We observed Snail expression to be increased in both early and advanced EMT sub-groups. Snail protein has been shown to bind the transcriptional start site of the E-cadherin gene and repress expression.[275] Loss of E-cadherin, as observed in the advanced EMT sub-groups, is a hallmark of EMT. An unexpected finding from our data was that Twist and Snail expression was heterogenous within the EMT sub-groups. On the whole, Snail was more highly expressed than Twist in both early and advanced EMT sub-groups (in contrast to epithelial CTC groups). However, there were also specific EMT phenotypic subgroups within patient samples where both Snail and Twist expression was decreased, or Twist expression was increased compared to Snail. Such heterogeneity of EMT transcription factors within mesenchymal-transitioning CTCs is a novel finding and likely represents the aforementioned dynamic transitional state that occurs within the circulation in response to multiple EMT inducing stimuli.

Comparing our dataset to existing evidence of EMT expression in CTCs is difficult. As has been highlighted in the introduction, much of the work to-date has used bulk RNA sequencing to identify upregulated genes from a pool of enriched CTCs (with contaminating leukocytes). Expression of both epithelial and mesenchymal CTC transcripts has been observed in patient samples, with evidence in HNSCC that this signature is associated with treatment failure and recurrence.[208] Crucially, however, such studies could not infer EMT expression patterns down to the single-cell level – the advantage of our approach. Of those studies that have achieved single-cell characterisation (the majority by immunofluorescent direct cell visualisation), several have reported that the presence of a hybrid (or partial) EMT CTC state is a potential prognostic marker of recurrence and reduced progression-free survival.[206] Again, these studies are lacking depth of insight due to small marker panels stemming from technological limitations.

8.4.2 Phenotypic diversity of CTC sub-groups

Aside from discussing EMT status, the findings of phenotypically diverse CTC sub-groups based upon proliferation and stemness markers are of particular interest when investigating markers underlying the metastatic behaviour of CTCs. CD44, a marker of stemness and also tumour metastatic potential, is expressed in up to 60% of HNSCC tumours and is associated with lymph node metastases and decreased survival.[249] CD44 expression by CTCs has been noted in several cancers, with one study reporting that 54% (15/28) of CTCs in HNSCC were positive for CD44. We specifically observed a strong correlation between CD44 expression and EMT sub-groups, in particular the advanced EMT sub-group. The idea that CD44 contributes to the metastatic potential of CTCs has recently been highlighted by Loreth *et al*. who demonstrated that in patients with brain metastases (from breast, lung and melanoma), almost all CTCs expressed CD44.[212] Furthermore, in the same study CTC expression of CD44 did not correlate with metastatic tissue. The authors argued that CTC CD44 expression was an adaptive response of CTC survival in the circulation. Thus, our findings of CD44 expressed almost exclusively on EMT CTCs adds to current knowledge, demonstrating potential mechanisms of metastatic potential for these mesenchymal-like CTCs.

The immune-regulatory marker CD47 has been demonstrated as highly expressed on CTCs in HNSCC, with two recent studies observing CD47 expression in 57% and 79% of CTCs.[112], [130] Therefore, of particular interest was a clear pattern of CD47 co-expression with CD44 – forming the basis for class II and IV phenotypic sub-group categorisation i.e. highly metastatic CTCs also express an immune evasion marker. While the apparent relationship between CD44 and CD47 in CTCs is highly novel, we were able to find reports of this metastatic and immune

evasive cancer phenotype in solid tumours. For example, in colorectal cancer[276] and ovarian cancer,[277] CD44 and CD47 co-expression has been correlated with metastasis and resistance to therapy. Furthermore, although a small cohort, in 3 patients with colorectal metastatic lesions treated with the PD-1 inhibitor nivolumab , CD44 and CD47 expression was also noted to be high.[276] Such evidence is noteworthy given we observed low immune-checkpoint expression in class II and IV CTCs (with high CD44 and CD47) potentially representing a CTC escape mechanism of therapeutic resistance (if alternative pathways of T-cell evasion are upregulated).

While the expression of CD44 and CD47, as described above, was prominent in EMT groups, particularly the advanced sub-group - the expression of stemness markers CD133 and CD24 appeared separate to these, creating the class I and III sub-groups, both of which were predominantly epithelial CTC groups. Evidence continues to grow that EMT and stemness phenotypes are linked to cancer progression in solid tumours,[270] however the above findings highlight a contrasting pattern in CTCs. Both of the class I and III stemness marker sub-groups demonstrated high immune-checkpoint expression, suggesting the potential for therapeutic targeting; in particular the proliferative class I sub-group. Although, without clinical follow-up data their prognostic significance is unknown.

8.4.3 Signalling pathway activation in CTC sub-groups

The ability of mass cytometry to assess post-transcriptional modification of signalling proteins, in the form of phosphorylation, is potentially a major advantage over conventional RNA sequencing based gene expression studies. Notwithstanding patterns in individual

phospho-protein expression, there was a general observation of 'activated' CTC sub-groups with high expression of multiple phospho-proteins. More commonly seen in the proliferative class I and II sub-groups, and the epithelial class III sub-group. However, pathway activation in these sub-groups was highly variable between samples – potentially attributable to the complex and heterogenous landscape of stimuli and surveillance present in individual patients.

While many questions regarding the complex multi-omic pathways regulating EMT remain unanswered,[278] we observed notable patterns of phosphorylated signalling protein expression in these groups. ERK is a known mediator of EMT - binding to the promotor region of EpCAM and silencing gene expression, while also inducing EMT transcription factors such as Twist and Snail expression. [279] The observation of high pERK expression in early and advanced EMT sub-groups with decreased EpCAM expression provides evidence to support this pathway of activation in CTCs. In addition, the ERK pathway has shown promise as a druggable target in cancer therapy.[230] Thus, the finding that pERK was highly expressed in CTC sub-groups with low immune checkpoint expression provides a method of targeting CTCs that would otherwise escape conventional immune checkpoint inhibitor therapy. The same can be said of STAT1, another potentially druggable target,[221] which was highly expressed in the advanced EMT class IV sub-group. STAT1 is thought to be a tumour suppressor via IFN signalling, but can also act as a tumour promoter via immune suppression mechanisms.[223] Evidence across multiple cancers is equivocal as to whether it is a marker of better or worse clinical outcomes.[221], [223] Thus, correlating clinical outcome to presence of the pSTAT1 high CTC sub-group would be particularly interesting in future studies.

Perhaps the most striking demonstration of signalling protein activation in our dataset was that of pCREB – being exclusively and highly expressed in both early and advanced EMT sub-groups, in stark contrast to epithelial sub-group expression. CREB is a transcription factor implicated in tumour cell proliferation, survival and metastasis; with experimental models having demonstrated multiple agents that could act as CREB inhibitors.[234] Thus, if hybrid EMT and mesenchymal CTCs are thought to be a significantly poor prognostic indicator, there is profound potential for clinical translation of this finding.

8.4.4 Immune-checkpoint expression

As discussed in chapter 2, CTC expression of numerous immune-related and checkpoint markers have been investigated in HNSCC, with no clear correlation to prognosis. We have highlighted above that CTC clusters in our dataset were separated into 'immune high' and 'immune low' sub-groups. We were unable to discern significant patterns of expression between checkpoint markers, with the exception that PD-L2 was on occasion independently elevated in class IV sub-groups. However, the predominance of EMT (and particularly advanced EMT) groups in the immune low sub-groups is both notable and concerning. If validated in larger studies, then the CTCs with most proliferative and metastatic potential appear to be refractory to currently available immune-checkpoint inhibitor therapy, thus alternative pathways would need to be targeted. Analysis of TCGA tumour gene expression datasets have demonstrated that mesenchymal-like tumours express significantly higher levels of checkpoints PD-L1 and CTLA4,[75] thus we would expect to see similar patterns in CTCs as they transition to a mesenchymal state or if they originate from a mesenchymal-like tumour. Therefore, the observation of EMT CTCs with decreased immune checkpoint

expression would further strengthen the argument of the plasticity and adaptation, for whatever reason, of CTCs within the circulation.

8.4.5 Comparison of mass cytometry to gene expression profiling of CTCs and tumour samples

Conventional methods of bulk RNAseq of enriched blood samples are unable to discern information relating to cellular heterogeneity, with the impact of contaminating leukocytes creating the potential for false positives in differential expression analysis.[280] While single-cell RNAseq is technically challenging and very costly. If the goal of an assay is merely to confirm the presence of CTCs, then bulk expression profiling appears sufficient, as evidenced in our data by a high concordance rate between epithelial cell transcript and mass cytometry CTC presence. However, we demonstrated that mass cytometry outperforms bulk CTC gene expression profiling, both to detect EMT and phenotype marker expression. For example, in our dataset, bulk expression profiling failed to identify EMT CTCs in over 60% of patients. In this respect, the benefits of single-CTC characterisation are somewhat obvious, in particular to identify low frequency CTC sub-groups that may have a clinical impact - such as contributing to treatment resistance.[281]

Few studies have compared CTC expression profiles to matched tumour tissue, on the whole reporting no correlation, for example PD-L1 expression in HNSCC.[112] In a similar fashion, we report patterns of CTC expression that appear distinct to those found in the tumour. The invasive tumour margin is observed to harbour a mesenchymal profile in a number of

epithelial cancers,[282], [283] including HNSCC.[35] Therefore, we were surprised to see that tumour margin samples in our data did not correlate with a mesenchymal gene expression profile. Instead, tumour mesenchymal profile was patient specific, i.e. not linked to the intra-tumoural sub-site. It is difficult to comment on trends in such a pilot cohort, but one reason may be sampling variability, which in our study is reliant upon the operating surgeon.

As we observed in our flow cytometry dataset in chapter 6, we were unable to correlate tumour and CTC EMT status. However, a notable finding in this study was the link between tumour mesenchymal profile and higher proportions of CTC sub-groups with proliferative and immune checkpoint marker expression. Such a finding asks the question: *if CTC characteristics such as EMT status are transient within the circulation (i.e. not correlated to tumour status), then are perhaps other characteristics such as phenotype/immune marker expression 'inherited' or linked to the primary tumour profile (or metastatic lesion)?* It provides further insight into the possible mechanistic underpinning of CTC biology. The hallmark of a transition to a mesenchymal phenotype is increased motility and invasive potential and it may appear that CTC retain this characteristic despite EMT plasticity. In solid tumours, EMT profiles associate with immune checkpoint (including PD-L1) expression[284], a correlation also observed in HNSCC.[75] Therefore, the observation of a tumour mesenchymal profile associated with CTC immune checkpoint expression is noteworthy as another druggable pathway of metastasis. Mesenchymal profile tumours demonstrate worse survival outcomes and the link with proliferative and immune suppressive CTCs may be one contributing factor to this.

As a final comment, it should be noted that the relationship of non-malignant tumour cells (immune and stromal cells) and the tumour micro-environment (such as hypoxia), to resulting CTC phenotypes remains poorly understood. Our tumour gene expression profiling is from bulk tumour cells, including malignant and non-malignant cells, of which the latter (being more prevalent) may impact data analysis. Single-cell sequencing can provide a depth of previously unobtainable information relating to the tumour micro-environment, including the relationship of tumour EMT and infiltrating immune cells to metastasis and clinical outcomes.[35]

8.5 Proteomics versus transcriptomics for CTC clinical translation?

The goal of this thesis has been the development and technical optimisation of a protocol to enrich and characterise CTCs in HNSCC - that would allow subsequent upscaling to a larger cohort clinical trial and eventual clinical translation. We emphasise the requirement to be able to collect, transport and store blood samples for up 72hrs or beyond, while maintaining stable cell protein expression. Such a protocol is crucial to allow multi-centre (or even international) trials to be conducted with reliable and reproducible data outputs – a key finding from this thesis. However, visualising a line of site to translate CTC biomarkers into routine clinical care remains challenging. As has been highlighted, we require a standardised, reproducible and universally accepted method of CTC enrichment. Furthermore, evidence is needed to demonstrate which of the available multi-omic CTC datasets hold most potential for clinical translation. Data remains equivocal as to the pros and cons of single-CTC proteomics versus genomics/transcriptomics, with no clear consensus and significant difficulty when comparing the data output of such studies.[281]

While this thesis focused on proteomics , considerable success has been achieved with single-CTC transcriptomic (and genomic) characterisation in multiple cancer types (as discussed in chapter 1).[280], [281] For example, recent evidence in gastric cancer correlated single-CTC gene expression patterns related to platelet adhesion with EMT and chemoresistance.[285] Such insights are only achievable through the resolution of data afforded by whole-transcriptome sequencing and removing the *a priori* selection bias of proteomic panels.

However, the cost of performing CTC transcriptomics versus proteomics should be highlighted. Single-CTC RNAseq costs in the region of thousands of pounds (dependent on platform and number of cells sequenced). In our experience, we estimate that processing a sample using mass cytometry costs around £110 (antibody and cytometer run time), which is comparable to mRNA quantitative gene expression profiling of bulk CTCs (approximately £100 per sample). While RNAseq of enriched CTCs in a 'bulk' fashion has shown success in previous studies to identify CTC specific transcripts - we have demonstrated the necessity of single-CTC analysis (proteomic or transcriptomic) to define CTC heterogeneity.

As discussed, transcript and protein expression patterns are not 100% correlated. Thus, we argue that the proteome provides the most accurate assessment of a cell's functional phenotype. Our test case of utilising mass spectrometry for pan-proteome CTC characterisation was a single 'mock CTC' sample only. However, it represents proof-of-concept that mass spectrometry may be applicable when only bulk CTC analysis is clinically required i.e. diagnostic purposes. In this regard we provide a glimpse at a potential future direction of our research into CTC proteomics. Although mass spectrometry cannot provide

single-CTC data, it is a cost-effective and sensitive assay for the unbiased characterisation of hundreds-thousands of potential CTC protein biomarkers. Similar to studies using mass spectrometry to define serum protein biomarkers for early detection in HNSCC[286] – mass spectrometry of CTC enriched blood could provide a sensitive means of longitudinal monitoring of disease burden and early detection of R/M via serial sampling.

We would argue that neither proteomics nor transcriptomics can be considered superior to the other, nor should they. Both methods of analysis can be aligned and utilised in a pathway of CTC liquid biopsy clinical translation. From a discovery science perspective, interrogating the whole transcriptome is invaluable. However, we would also argue that CTC proteomics is potentially a more viable solution to translate single-CTC analysis to the clinic, particularly considering the above cost implications - *i.e. once we have used transcriptomics to identify gene/protein expression signatures of merit, targeted single-CTC proteomics (such as mass cytometry) may allow a robust and cost-effective means of clinical upscaling.* Clearly this will be the focus of future research.

8.6 CTCs or circulating tumour DNA (ctDNA) in a liquid biopsy?

While this project has focused on CTCs, if one is discussing the merits of a liquid biopsy and how it translates to clinical practice, it is impossible not to highlight the role of ctDNA. While it is outside the scope of this thesis to discuss ctDNA in-depth (we have undertaken this review separately[287]), we wish to highlight several key points. Several papers have attempted to compare ctDNA and CTC biomarkers, however fundamentally they are providing different

viewpoints of tumour progression and the metastatic process through differing availability of multi-omic datasets. Genotyping of ctDNA provides an assessment of the oncogenic mutations driving tumour growth and progression, with evidence slowly accumulating for the advantage of ctDNA to detect intra-tumoural spatial heterogeneity,[288] not afforded by single site tumour biopsies.[33] In contrast, CTCs offer the ability to investigate the phenotype of whole tumour cells that have a direct causative link to metastasis. While several studies have compared tumour and CTC expression profiles, our findings and those of others demonstrate that once a tumour cell enters the circulation it changes significantly. CTCs are adapting to a hostile environment, with phenotypic plasticity markedly different to that found in the primary tumour. Therefore, serving as independent prognostic/predictive biomarkers.

Direct comparison between CTCs and ctDNA is difficult and dependent upon the output required from the liquid biopsy test, i.e. is the primary goal diagnostic screening or therapeutic targeting in recurrent/metastatic patients? For diagnostic purposes, the utility of CTCs may be questionable. Our data from flow and mass cytometry reports the number of CTCs (in a skewed cohort of advanced stage patients) to be in the region of 1-20 cells per ml of whole blood – such a low number of cells creates potential for a high margin of error in any CTC cancer screening test (this may be obviated by the sensitivity of mass spectrometry CTC proteomics). Using regression modelling of CTC enumeration data from 836 cancer patients, Coumans *et al.* demonstrated that 99% of cancer patients will have at least one epithelial positive CTC, but that up to 5 litres of blood would need to be analysed to reach this level of sensitivity.[289] Few studies have directly compared the sensitivity of CTCs and ctDNA in early cancer cohorts – one study reported a sensitivity of 67% vs 90% to detect CTCs and ctDNA, albeit in metastatic cancer patients. Recently, studies using targeted ctDNA panels have

demonstrated sensitivities >90%, across multiple cancer types, to detect early stage (T1 and T2) cancers.[290], [291] In addition, ctDNA as a source of predictive/therapeutic biomarkers has gained a lot more traction than CTCs and this is not entirely surprising. In contrast to CTCs, ctDNA extraction and processing is an optimised and reproducible assay. While evidence is not conclusive, ctDNA likely represents a more accurate picture of a tumours genomic landscape.[292] Therefore, obtaining therapeutic genomic biomarkers from CTCs, such as tumour mutational burden, is significantly challenging. To this end, several ctDNA companion diagnostics are now FDA approved[293] and recommended by NICE to detect mutations associated with therapeutic agents, for example the FoundationOne Liquid CDx[294] and EGFR mutation assessment in lung cancer.[295]

However, we believe that over the next decade research into CTCs will undergo a paradigm shift to highlight the importance of this liquid biopsy compartment. Standardised and reproducible enrichment protocols, combined with routine single-CTC characterisation and long-term follow-up data in large cohort studies will allow researchers to define novel CTC-specific predictive and/or therapeutic biomarkers – beyond that which is available from ctDNA. On the other hand, perhaps an ideal liquid biopsy should analyse both ctDNA and CTCs. Several studies have reported protocols for simultaneous ctDNA and CTC assessment from single samples.[296], [297] For example, in a recent study of 99 lung cancer patients, the sensitivity of ctDNA and CTCs was 73% and 66% respectively – however this rose to 95% sensitivity when the these were combined, outperforming conventional tumour markers.[296] A separate study of 16 lung and 21 breast cancer patients identified significant genomic heterogeneity between tumour tissue, CTC and ctDNA.[297] CTCs and ctDNA contained mutations not present in the tumour, but heterogeneity was also evident between

CTCs and ctDNA i.e. both were beneficial to assess ITH. Therefore, assessing both CTCs and ctDNA may provide complimentary biomarkers, while also allowing differing multi-omic biomarkers of tumour progression and metastasis to be assessed. While it was not investigated in this thesis, other groups have optimised protocols to successfully analyse both ctDNA and CTCs from a single blood sample using the Parsortix platform.[298] This combined ctDNA and CTC liquid biopsy is another area of future research for our group.

8.7 Challenges and lessons learnt

Notwithstanding the task of undertaking a PhD during a prolonged global pandemic, there were several challenges and lessons learnt during this project. First and foremost, the task of optimising a novel piece of equipment in a field of discovery science was a challenge on multiple fronts – far beyond initial estimations. The recurring variable we encountered was the inability to control for the presence (or absence) of CTCs in a patient sample when testing novel methods of characterisation. This was further compounded by an evolving criteria for defining a CTC as the sensitivity of characterisation protocols improve. The obvious example of this is capturing CTCs which are in a mesenchymal-transitional state – these are the cells thought to have the highest metastatic potential. However, downregulated epithelial markers (and heterogeneity in EMT markers as seen in our data) results in significant challenges with regard CTC identification and phenotyping. For this reason, validating a novel device *and* optimising each stage in our sample processing protocol was particularly difficult.

The Italian proverb *"perfect is the enemy of good"* (Voltaire, 1770) is never truer than during a PhD project. Our intention to add further CTC markers to the mass cytometry panel was tempered by the ever-increasing time spent optimising antibodies and the need to make progress with patient samples having had unforeseen project delays. In addition, we were unable to perform mass cytometry analysis of tumour tissue, which would have been of particular interest as a direct comparison to the novel CTC sub-groups we defined – this represents an avenue to pursue in future research. Much like a tumour cell entering the circulation, the greatest lesson learnt has been one of adaptation, while keeping all avenues of possible research open. Modifying initial project aims and objectives was key to allow a meaningful endpoint and data output to be achieved. For example, one of our initial aims was to achieve single-CTC RNAseq. Having established that Transfix tubes were the optimum method of CTC preservation we trialled the 10x genomics single-cell platform on a patient sample, with no success. The protocol and library preparation was not compatible with fixed cells. However, this led us to mass cytometry for single-CTC characterisation – a technique we think is eminently more translatable than current single-cell RNAseq methods.

Multiple potential side-projects have been pursued, both in CTC and ctDNA research, for example the CTC proteomics data presented above. These projects have been hypothesis generating, providing proof-of-concept and pilot data for potential future funding applications. All the chapters in this thesis necessitated the development of different laboratory and bioinformatics skillsets - from cell culture, to cytometry and high-dimensional data analysis, all of which have strengthened my research portfolio.

8.8 Conclusion

To conclude, we have optimised and validated a protocol using immediate sample fixation at point of blood collection and microfluidic enrichment to capture CTCs in HNSCC. We demonstrate that mass cytometry can be successfully used to provide multi-parameter single-CTC characterisation of extracellular, intracellular and activated signalling proteins. The ability of mass cytometry to interrogate the expression of >40 markers allows a previously unseen view of the dynamic state of CTC epithelial and mesenchymal marker expression. Even in a small cohort pilot study we highlight an interesting trend that primary tumours with mesenchymal gene expression profiles exhibit increased levels of proliferative and immune-checkpoint high CTCs. Furthermore, the heterogeneity of EMT and phenotypic sub-groups and their relationship to immune-checkpoint markers is notable – providing both an avenue for therapeutic targeting, but also a means of therapeutic escape in 'immune low' CTCs. As a feasibility study to provide proof-of-concept, we believe our project has been a success. A low processing cost, minimally invasive nature of a simple blood test and ability to conduct serial measurements longitudinally, provides real potential that a CTC-based liquid biopsy can have significant clinical impact. HNSCC is a cancer that presents late and has a high rate of recurrence, therefore these characteristics of a liquid biopsy could profoundly benefit clinicians in monitoring disease burden i.e. early cancer detection and detecting residual disease/recurrence in the post-treatment phase. In addition, our data has provided an insight into the potential for biomarker development (prognostic and predictive) by stratifying patients based upon novel CTC sub-groups using single-CTC proteome characterisation.

Future avenues of research include the exploration of pan-proteome CTC assessment using mass spectrometry, complimentary analysis of ctDNA and CTCs and further investigating the impact of the tumour profile and micro-environment on CTC expression patterns. However, it remains to be seen if such multi-omic liquid biopsy assessment can provide information that is *greater than the sum of its parts* for biomarker-led patient stratification, thus justifying the increased burden of cost to attain these datasets. The pertinent research question arising from our findings is *which CTC sub-group(s) will hold prognostic and/or predictive value?* Such inference will require subsequent validation studies leading to larger clinical trials. Only then do we have the knowledge to specifically target CTCs

Nevertheless, we have laid a platform on which to build and inform the design of such future studies.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[2] Cancer Research UK, "Head and neck cancers statistics | Cancer Research UK." https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers#heading-Zero (accessed Mar. 19, 2020).

[3] A. K. Chaturvedi *et al.*, "Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States," *J. Clin. Oncol.*, vol. 29, no. 32, pp. 4294–4301, Nov. 2011, doi: 10.1200/JCO.2011.36.4596.

[4] K. K. Ang *et al.*, "Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer," *N. Engl. J. Med.*, vol. 363, no. 1, pp. 24–35, Jul. 2010, doi: 10.1056/NEJMoa0912217.

[5] E. A. Van Dyne, S. J. Henley, M. Saraiya, C. C. Thomas, L. E. Markowitz, and V. B. Benard, "Trends in Human Papillomavirus–Associated Cancers — United States, 1999–2015," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 67, no. 33, pp. 918–924, Aug. 2018, doi: 10.15585/mmwr.mm6733a2.

[6] W. Lydiatt *et al.*, "Head and Neck Cancers-Major changes in the American Joint Committe on Cancer Eighth Edition Cancer Staging Manual," *CA Cancer J Clin*, vol. 67, no. 2, pp. 122–137, 2017, doi: 10.3322/caac.21389.

[7] A. G. Sacco and E. E. Cohen, "Current treatment options for recurrent or metastatic head and neck squamous cell carcinoma," *J. Clin. Oncol.*, vol. 33, no. 29, pp. 3305–3315, 2015, doi: 10.1200/JCO.2015.62.0963.

[8] K. Kim, M. M. Amonkar, D. Högberg, and F. Kasteng, "Economic burden of resected squamous cell carcinoma of the head and neck in an incident cohort of patients in the UK," *Head Neck Oncol.*, vol. 3, no. 1, p. 47, 2011, doi: 10.1186/1758-3284-3-47.

[9] H. R. Vermorken JB, Mesia R, Rivera F, Remenar E, Kawecki A, Rottey S, Erfan J, Zabolotnyy D, Kienzer HR, Cupissol D, Peyrade F, Benasso M, Vynnychenko I, De Raucourt D, Bokemeyer C, Schueler A, Amellal N, "Platinum-Based Chemotherapy plus Cetuximab in Head and Neck Cancer," *N. Engl. J. Med.*, vol. 359, no. 11, pp. 1116–1127, 2008, [Online]. Available: %5C%5Cgbkenpfsphg0001%5CResolute%5CDATA%5CPurpleHat%5Cclients.ph%5CMerckSerono%5CAdmin%5CReferences%5CSCCHN%5CVermorken_N Engl J Med_2008_EXTREME.pdf.

[10] R. L. Ferris *et al.*, "Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck," *N. Engl. J. Med.*, vol. 375, no. 19, pp. 1856–1867, 2016, doi: 10.1056/NEJMoa1602252.

[11] D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation," *Cell*, vol. 144, pp. 646–674, 2011, doi: 10.1016/j.cell.2011.02.013.

[12] G. P. Pfeifer, M. F. Denissenko, M. Olivier, N. Tretyakova, S. S. Hecht, and P. Hainaut, "Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers," *Oncogene*, vol. 21, no. 48, pp. 7435–7451, Oct. 2002, doi: 10.1038/sj.onc.1205803.

[13] T. Strachan and A. Read, *Human Molecular Genetics*, 3rd ed. Garland Science, 2004.

[14] X. Shao *et al.*, "Copy number variation is highly correlated with differential gene expression: a pan-cancer study," *BMC Med. Genet.*, vol. 20, no. 1, p. 175, Dec. 2019, doi: 10.1186/s12881-019-0909-5.

[15] R. Beroukhim *et al.*, "The landscape of somatic copy-number alteration across human cancers.," *Nature*, vol. 463, no. 7283, pp. 899–905, Feb. 2010, doi: 10.1038/nature08822.

[16] J. C. Smith and J. M. Sheltzer, "Systematic identification of mutations and copy number alterations associated with cancer patient prognosis," *Elife*, vol. 7, pp. 1–26, 2018, doi: 10.7554/eLife.39217.

[17] D. A. Lawson, K. Kessenbrock, R. T. Davis, N. Pervolarakis, and Z. Werb, "Tumour heterogeneity and metastasis at single-cell resolution," *Nat. Cell Biol.*, vol. 20, pp. 1349–1360,

2018, doi: 10.1038/s41556-018-0236-7.
[18] L. B. Alexandrov *et al.*, “Signatures of mutational processes in human cancer.,” *Nature*, vol. 500, no. 7463, pp. 415–21, Aug. 2013, doi: 10.1038/nature12477.
[19] M. Greaves and C. C. Maley, “Clonal evolution in cancer.,” *Nature*, vol. 481, no. 7381, pp. 306–13, Jan. 2012, doi: 10.1038/nature10762.
[20] A. A. Alizadeh *et al.*, “Toward understanding and exploiting tumor heterogeneity.,” *Nat. Med.*, vol. 21, no. 8, pp. 846–53, Aug. 2015, doi: 10.1038/nm.3915.
[21] N. McGranahan and C. Swanton, “Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future,” *Cell*, vol. 168, no. 4, pp. 613–628, 2017, doi: 10.1016/j.cell.2017.01.018.
[22] Y. Jiang, Y. Qiu, A. J. Minn, and N. R. Zhang, “Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 37, pp. E5528-37, Sep. 2016, doi: 10.1073/pnas.1522203113.
[23] M. Krzywinski, “Visualizing Clonal Evolution in Cancer,” *Mol. Cell*, vol. 62, pp. 652–656, 2016, doi: 10.1016/j.molcel.2016.05.025.
[24] A. Kreso and J. E. Dick, “Evolution of the Cancer Stem Cell Model,” *Cell Stem Cell*, vol. 14, no. 3, pp. 275–291, Mar. 2014, doi: 10.1016/J.STEM.2014.02.006.
[25] D. Hanahan and L. M. Coussens, “Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment,” *Cancer Cell*, vol. 21, no. 3, pp. 309–322, Mar. 2012, doi: 10.1016/J.CCR.2012.02.022.
[26] A. Anchal Sharma, E. Merritt, X. Hu, J. Malhotra, G. M. Riedlinger, and S. De Correspondence, “Non-Genetic Intra-Tumor Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary Dynamics in Lung Tumors In Brief,” *CellReports*, vol. 29, pp. 2164-2174.e5, 2019, doi: 10.1016/j.celrep.2019.10.045.
[27] N. Stransky *et al.*, “The Mutational Landscape of Head and Neck Squamous Cell Carcinoma,” *Science*, vol. 333, no. 6046, p. 1157, 2011, doi: 10.1126/SCIENCE.1208130.
[28] N. Agrawal *et al.*, “Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1.,” *Science*, vol. 333, no. 6046, pp. 1154–7, Aug. 2011, doi: 10.1126/science.1206923.
[29] R. Carlsen *et al.*, “Comprehensive genomic characterization of head and neck squamous cell carcinomas,” *Nature*, vol. 517, no. 7536, pp. 576–582, 2015, doi: 10.1038/nature14129.
[30] T. Y. Seiwert *et al.*, “Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas.,” *Clin. Cancer Res.*, vol. 21, no. 3, pp. 632–41, Feb. 2015, doi: 10.1158/1078-0432.CCR-13-3310.
[31] C. H. Chung *et al.*, “Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression,” *Cancer Cell*, vol. 5, no. 5, pp. 489–500, May 2004, doi: 10.1016/S1535-6108(04)00112-6.
[32] V. Walter *et al.*, “Molecular Subtypes in Head and Neck Cancer Exhibit Distinct Patterns of Chromosomal Gain and Loss of Canonical Cancer Genes,” *PLoS One*, vol. 8, no. 2, p. e56823, Feb. 2013, doi: 10.1371/journal.pone.0056823.
[33] M. Gerlinger *et al.*, “Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing,” *NEJM*, vol. 366, no. 10, pp. 883–892, 2012, doi: 10.1056/NEJMoa1113205.Intratumor.
[34] C. Gawad, W. Koh, and S. R. Quake, “Single-cell genome sequencing: current state of the science,” *Nat. Rev. Genet.*, vol. 17, no. 3, pp. 175–188, Mar. 2016, doi: 10.1038/nrg.2015.16.
[35] S. V Puram, I. Tirosh, A. S. Parikh, D. T. Lin, A. Regev, and B. E. Bernstein, “Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer,” *Cell*, vol. 171, no. 7, pp. 1611–1624, 2017, doi: 10.1016/j.cell.2017.10.044.
[36] N. Navin *et al.*, “Tumour evolution inferred by single-cell sequencing.,” *Nature*, vol. 472, no. 7341, pp. 90–4, Apr. 2011, doi: 10.1038/nature09807.
[37] N. E. Navin, “The first five years of single-cell cancer genomics and beyond,” *Genome Res.*,

vol. 25, no. 10, pp. 1499–1507, Oct. 2015, doi: 10.1101/GR.191098.115.
[38] T. M. Van and C. U. Blank, “A user’s perspective on GeoMxTM digital spatial profiling,” *Immuno-Oncology Technol.*, vol. 1, pp. 11–18, Jul. 2019, doi: 10.1016/J.IOTECH.2019.05.001.
[39] H. R. Ali *et al.*, “Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer,” *Nat. Cancer*, vol. 1, no. 2, pp. 163–175, Feb. 2020, doi: 10.1038/s43018-020-0026-6.
[40] E. A. Mroz, A. M. Tward, R. J. Hammon, Y. Ren, and J. W. Rocco, “Intra-tumor Genetic Heterogeneity and Mortality in Head and Neck Cancer: Analysis of Data from The Cancer Genome Atlas,” *PLoS Med.*, vol. 12, no. 2, pp. 1–27, 2015, doi: 10.1371/journal.pmed.1001786.
[41] X. C. Zhang *et al.*, “Tumor Evolution and Intratumor Heterogeneity of an Oropharyngeal Squamous Cell Carcinoma Revealed by Whole-Genome Sequencing 1,2,” *Neoplasia*, vol. 15, p. 1371, 2013, doi: 10.1593/neo.131400.
[42] M. Canning *et al.*, “Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on Immunotherapy.,” *Front. cell Dev. Biol.*, vol. 7, p. 52, 2019, doi: 10.3389/fcell.2019.00052.
[43] B. Peltanova, M. Raudenska, and M. Masarik, “Effect of tumor microenvironment on pathogenesis of the head and neck squamous cell carcinoma: a systematic review,” *Mol. Cancer*, vol. 18, no. 1, p. 63, Dec. 2019, doi: 10.1186/s12943-019-0983-5.
[44] E. A. Mroz and J. W. Rocco, “MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma.,” *Oral Oncol.*, vol. 49, no. 3, pp. 211–5, Mar. 2013, doi: 10.1016/j.oraloncology.2012.09.007.
[45] E. A. Mroz, A. D. Tward, C. R. Pickering, J. N. Myers, R. L. Ferris, and J. W. Rocco, “High intra-tumor genetic heterogeneity is related to worse outcome in head and neck squamous cell carcinoma,” *Cancer*, vol. 119, no. 16, pp. 3034–3042, 2013, doi: 10.1002/cncr.28150.
[46] A. Rajput, T. Bocklage, A. Greenbaum, J.-H. Lee, and S. A. Ness, “Mutant-Allele Tumor Heterogeneity (MATH) Scores Correlate with Risk of Metastases in Colon Cancer: MATH-Scores Correlate with Colon Ca Metastasis,” *Clin. Colorectal Cancer*, vol. 16, no. 3, p. e165, 2017, doi: 10.1016/J.CLCC.2016.11.004.
[47] M. Yarchoan, A. Hopkins, and E. M. Jaffee, “Tumor Mutational Burden and Response Rate to PD-1 Inhibition.,” *N. Engl. J. Med.*, vol. 377, no. 25, pp. 2500–2501, 2017, doi: 10.1056/NEJMc1713444.
[48] G. J. Hanna *et al.*, “Frameshift events predict anti-PD-1/L1 response in head and neck cancer.,” *JCI insight*, vol. 3, no. 4, 2018, doi: 10.1172/jci.insight.98811.
[49] T. N. Schumacher and R. D. Schreiber, “Neoantigens in cancer immunotherapy.,” *Science*, vol. 348, no. 6230, pp. 69–74, Apr. 2015, doi: 10.1126/science.aaa4971.
[50] A. Forschner *et al.*, “Tumor mutation burden and circulating tumor DNA in combined CTLA-4 and PD-1 antibody therapy in metastatic melanoma – results of a prospective biomarker study,” *J. Immunother. Cancer*, vol. 7, no. 1, p. 180, Dec. 2019, doi: 10.1186/s40425-019-0659-0.
[51] Y. K. Chae *et al.*, “Clinical Implications of Circulating Tumor DNA Tumor Mutational Burden (ctDNA TMB) in Non-Small Cell Lung Cancer.,” *Oncologist*, vol. 24, no. 6, pp. 820–828, 2019, doi: 10.1634/theoncologist.2018-0433.
[52] N. McGranahan *et al.*, “Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade.,” *Science*, vol. 351, no. 6280, pp. 1463–9, Mar. 2016, doi: 10.1126/science.aaf1490.
[53] J. J. Havel, D. Chowell, and T. A. Chan, “The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy.,” *Nat. Rev. Cancer*, vol. 19, no. 3, pp. 133–150, 2019, doi: 10.1038/s41568-019-0116-x.
[54] L. Fancello, S. Gandini, P. G. Pelicci, and L. Mazzarella, “Tumor mutational burden quantification from targeted gene panels: major advancements and challenges,” *J.*

Immunother. Cancer, vol. 7, no. 1, p. 183, Dec. 2019, doi: 10.1186/s40425-019-0647-4.
[55] R. M. Samstein *et al.*, "Tumor mutational load predicts survival after immunotherapy across multiple cancer types.," *Nat. Genet.*, vol. 51, no. 2, pp. 202–206, 2019, doi: 10.1038/s41588-018-0312-8.
[56] T. R. Ashworth, "A Case of Cancer in Which Cells Similar to Those in the Tumours Were Seen in the Blood after Death.," *Med. J. Aust.*, vol. 14, pp. 146–147, 1869.
[57] S. A. Leon, B. Shapiro, D. M. Sklaroff, and M. J. Yaros, "Free DNA in the serum of cancer patients and the effect of therapy.," *Cancer Res.*, vol. 37, no. 3, pp. 646–50, Mar. 1977, Accessed: Jun. 03, 2020. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/837366.
[58] M. Stroun, P. Anker, P. Maurice, J. Lyautey, C. Lederrey, and M. Beljanski, "Neoplastic characteristics of the DNA found in the plasma of cancer patients.," *Oncology*, vol. 46, no. 5, pp. 318–22, 1989, doi: 10.1159/000226740.
[59] F. Diehl *et al.*, "Circulating mutant DNA to assess tumor dynamics," *Nat. Med.*, vol. 14, no. 9, pp. 985–990, 2008, doi: 10.1038/nm.1789.Circulating.
[60] M. Murtaza *et al.*, "Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA," *Nature*, vol. 497, no. 7447, pp. 108–112, May 2013, doi: 10.1038/nature12065.
[61] S.-J. Dawson *et al.*, "Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer," *N. Engl. J. Med.*, vol. 368, no. 13, pp. 1199–1209, 2013, doi: 10.1056/NEJMoa1213261.
[62] S. Maheswaran *et al.*, "Detection of mutations in EGFR in circulating lung-cancer cells," *N. Engl. J. Med.*, vol. 359, pp. 366–377, 2008.
[63] M. Cristofanilli *et al.*, "Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer," *N. Engl. J. Med.*, vol. 351, no. 8, pp. 781–791, 2004.
[64] K. Payne *et al.*, "Circulating tumor DNA as a biomarker and liquid biopsy in head and neck squamous cell carcinoma," *Head Neck*, vol. 40, no. 7, pp. 1598–1604, 2018, doi: 10.1002/hed.25140.
[65] K. Payne *et al.*, "Circulating Tumour Cell Biomarkers in Head and Neck Cancer: Current Progress and Future Prospects," *Cancers (Basel).*, vol. 11, no. 8, p. 1115, Aug. 2019, doi: 10.3390/cancers11081115.
[66] E. Kilgour, D. G. Rothwell, G. Brady, and C. Dive, "Liquid Biopsy-Based Biomarkers of Treatment Response and Resistance," *Cancer Cell*, vol. 37, no. 4, 2020, doi: 10.1016/J.CCELL.2020.03.012.
[67] J. J. Chabon *et al.*, "Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients," *Nat. Commun.*, vol. 7, no. May, pp. 1–14, 2016, doi: 10.1038/ncomms11815.
[68] Kloten, Lampignano, Krahn, and Schlange, "Circulating Tumor Cell PD-L1 Expression as Biomarker for Therapeutic Efficacy of Immune Checkpoint Inhibition in NSCLC," *Cells*, vol. 8, no. 8, p. 809, 2019, doi: 10.3390/cells8080809.
[69] A. R. Parikh *et al.*, "Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers.," *Nat. Med.*, vol. 25, no. 9, pp. 1415–1421, 2019, doi: 10.1038/s41591-019-0561-9.
[70] D. S. Micalizzi, S. Maheswaran, and D. A. Haber, "A conduit to metastasis: Circulating tumor cell biology," *Genes Dev.*, vol. 31, no. 18, pp. 1827–1840, 2017, doi: 10.1101/gad.305805.117.
[71] T. M. Gorges and K. Pantel, "Circulating tumor cells as therapy-related biomarkers in cancer patients," *Cancer Immunol. Immunother.*, vol. 62, no. 5, pp. 931–939, May 2013, doi: 10.1007/s00262-012-1387-1.
[72] A. Chambers, A. Groon, and I. MacDonald, "Dissemination and growth of cancer cells in metastatic sites," *Nat. Rev.*, vol. 2, pp. 563–572, 2002, doi: 10.1038/nrc865.
[73] R. Kalluri and R. Weinberg, "The basics of epithelial-mesenchymal transition," *J. Clin. Invest.*, vol. 119, pp. 1420–1428, 2009, doi: 10.1172/JCI39104.1420.
[74] C. J. Gröger, M. Grubinger, T. Waldhör, K. Vierlinger, and W. Mikulits, "Meta-Analysis of Gene

Expression Signatures Defining the Epithelial to Mesenchymal Transition during Cancer Progression," *PLoS One*, vol. 7, no. 12, p. e51136, Dec. 2012, doi: 10.1371/journal.pone.0051136.

[75] A. R. Jung, C.-H. Jung, J. K. Noh, Y. C. Lee, and Y.-G. Eun, "Epithelial-mesenchymal transition gene signature is associated with prognosis and tumor microenvironment in head and neck squamous cell carcinoma," *Sci. Rep.*, vol. 10, no. 1, p. 3652, Dec. 2020, doi: 10.1038/s41598-020-60707-x.

[76] C.-L. Chen *et al.*, "Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer," *Prostate*, vol. 73, no. 8, pp. 813–826, Jun. 2013, doi: 10.1002/pros.22625.

[77] L. KJ *et al.*, "Multistep Nature of Metastatic Inefficiency: Dormancy of Solitary Cells After Successful Extravasation and Limited Survival of Early Micrometastases," *Am. J. Pathol.*, vol. 153, no. 3, 1998, doi: 10.1016/S0002-9440(10)65628-3.

[78] A. Brouwer *et al.*, "Evaluation and consequences of heterogeneity in the circulating tumor cell compartment," *Oncotarget*, vol. 7, no. 30, pp. 48625–48643, 2016, doi: 10.18632/oncotarget.8015.

[79] Y. Dong *et al.*, "Microfluidics and circulating tumor cells," *J. Mol. Diagnostics*, vol. 15, no. 2, pp. 149–157, 2013, doi: 10.1016/j.jmoldx.2012.09.004.

[80] S. Paget, "THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST.," *Lancet*, vol. 133, no. 3421, pp. 571–573, Mar. 1889, doi: 10.1016/S0140-6736(00)49915-0.

[81] L. Zhang *et al.*, "The identification and characterization of breast cancer CTCs competent for brain metastasis," *Sci Transl Med*, vol. 5, no. 180, 2013, doi: 10.1126/scitranslmed.3005109.The.

[82] D. Boral *et al.*, "Molecular characterization of breast cancer CTCs associated with brain metastasis," *Nat. Commun.*, vol. 8, no. 1, p. 196, Dec. 2017, doi: 10.1038/s41467-017-00196-1.

[83] G. T. Gibney, L. M. Weiner, and M. B. Atkins, "Predictive biomarkers for checkpoint inhibitor-based immunotherapy," *Lancet Oncol.*, vol. 17, no. 12, pp. e542–e551, 2016, doi: 10.1016/S1470-2045(16)30406-5.

[84] A. Strati *et al.*, "Prognostic significance of PD-L1 expression on circulating tumor cells in patients with head and neck squamous cell carcinoma," *Ann. Oncol.*, vol. 28, no. 8, pp. 1923–1933, 2017, doi: 10.1093/annonc/mdx206.

[85] A. Kulasinghe, C. Perry, L. Kenny, M. E. Warkiani, C. Nelson, and C. Punyadeera, "PD-L1 expressing circulating tumour cells in head and neck cancers," *BMC Cancer*, vol. 17, no. 1, p. 333, Dec. 2017, doi: 10.1186/s12885-017-3316-3.

[86] B. M. Szczerba *et al.*, "Neutrophils escort circulating tumour cells to enable cell cycle progression," *Nature*, vol. 566, no. 7745, pp. 553–557, Feb. 2019, doi: 10.1038/s41586-019-0915-y.

[87] S. Jansson, P.-O. Bendahl, A.-M. Larsson, K. E. Aaltonen, and L. Rydén, "Prognostic impact of circulating tumor cell apoptosis and clusters in serial blood samples from patients with metastatic breast cancer in a prospective observational cohort," *BMC Cancer*, vol. 16, no. 1, p. 433, Dec. 2016, doi: 10.1186/s12885-016-2406-y.

[88] A. Kulasinghe *et al.*, "A Collective Route to Head and Neck Cancer Metastasis," *Sci. Rep.*, vol. 8, no. 1, p. 746, Dec. 2018, doi: 10.1038/s41598-017-19117-9.

[89] A. Kulasinghe, J. Zhou, L. Kenny, I. Papautsky, and C. Punyadeera, "Capture of Circulating Tumour Cell Clusters Using Straight Microfluidic Chips," *Cancers (Basel).*, vol. 11, no. 1, 2019, doi: 10.3390/CANCERS11010089.

[90] A. Fabisiewicz and E. Grzybowska, "CTC clusters in cancer progression and metastasis," *Med. Oncol. 2016 341*, vol. 34, no. 1, pp. 1–10, Dec. 2016, doi: 10.1007/S12032-016-0875-0.

[91] N. Aceto *et al.*, "Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis," *Cell*, vol. 158, no. 5, pp. 1110–1122, 2014, doi: 10.1016/j.cell.2014.07.013.

[92] S. Gkountela *et al.*, “Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding.,” *Cell*, vol. 176, no. 1–2, pp. 98-112.e14, Jan. 2019, doi: 10.1016/j.cell.2018.11.046.

[93] K. P. Mcmullen, J. J. Chalmers, J. C. Lang, P. Kumar, and K. R. Jatana, “Circulating tumor cells in head and neck cancer : A review,” *World J. Otorhinolaryngol. Neck Surg.*, vol. 2, no. 2, pp. 109–116, 2016, doi: 10.1016/j.wjorl.2016.05.003.

[94] P. K. Grover, A. G. Cummins, T. J. Price, I. C. Roberts-Thomson, and J. E. Hardingham, “Circulating tumour cells: the evolving concept and the inadequacy of their enrichment by EpCAM-based methodology for basic and clinical cancer research,” *Ann. Oncol.*, vol. 25, no. 8, pp. 1506–1516, Aug. 2014, doi: 10.1093/annonc/mdu018.

[95] D. Grölz *et al.*, “Liquid Biopsy Preservation Solutions for Standardized Pre-Analytical Workflows—Venous Whole Blood and Plasma,” *Curr. Pathobiol. Rep.*, vol. 6, no. 4, pp. 275–286, Dec. 2018, doi: 10.1007/s40139-018-0180-z.

[96] J. Qin, J. R. Alt, B. A. Hunsley, T. L. Williams, and M. R. Fernando, “Stabilization of circulating tumor cells in blood using a collection device with a preservative reagent.,” *Cancer Cell Int.*, vol. 14, no. 1, p. 23, Mar. 2014, doi: 10.1186/1475-2867-14-23.

[97] D. L. Evers, C. B. Fowler, B. R. Cunningham, J. T. Mason, and T. J. O’Leary, “The effect of formaldehyde fixation on RNA: optimization of formaldehyde adduct removal.,” *J. Mol. Diagn.*, vol. 13, no. 3, pp. 282–8, May 2011, doi: 10.1016/j.jmoldx.2011.01.010.

[98] S. S. Yee *et al.*, “A novel approach for next-generation sequencing of circulating tumor cells,” *Mol. Genet. Genomic Med.*, vol. 4, no. 4, p. 395, 2016, doi: 10.1002/MGG3.210.

[99] K. H. K. Wong *et al.*, “Whole blood stabilization for the microfluidic isolation and molecular characterization of circulating tumor cells,” *Nat. Commun.*, vol. 8, 2017, doi: 10.1038/S41467-017-01705-Y.

[100] P. Apostolou, D.-A. Ntanovasilis, and I. Papasotiriou, “Evaluation of a simple method for storage of blood samples that enables isolation of circulating tumor cells 96 h after sample collection,” *J. Biol. Res.*, vol. 24, 2017, doi: 10.1186/S40709-017-0068-9.

[101] A. A. Powell *et al.*, “Single Cell Profiling of Circulating Tumor Cells : Transcriptional Heterogeneity and Diversity from Breast Cancer Cell Lines,” *PLoS One*, vol. 7, no. 5, p. e33788, 2012, doi: 10.1371/journal.pone.0033788.

[102] Y.-F. Sun *et al.*, “Circulating Tumor Cells from Different Vascular Sites Exhibit Spatial Heterogeneity in Epithelial and Mesenchymal Composition and Distinct Clinical Significance in Hepatocellular Carcinoma.,” *Clin. Cancer Res.*, vol. 24, no. 3, pp. 547–559, 2018, doi: 10.1158/1078-0432.CCR-17-1063.

[103] A. Kulasinghe, C. Perry, L. Jovanovic, C. Nelson, and C. Punyadeera, “Circulating Tumour Cells in Metastatic Head and Neck Cancers.,” *Int. J. cancer*, vol. 61, no. 0, pp. 1–26, 2014, doi: 10.1002/ijc.29108.

[104] M. Buglione *et al.*, “Circulating tumour cells in locally advanced head and neck cancer: Preliminary report about their possible role in predicting response to non-surgical treatment and survival,” *Eur. J. Cancer*, vol. 48, no. 16, pp. 3019–3026, 2012, doi: 10.1016/j.ejca.2012.05.007.

[105] T. Kawada *et al.*, “Circulating tumor cells in patients with head and neck squamous cell carcinoma: Feasibility of detection and quantitation,” *Head Neck*, vol. 39, no. 11, pp. 2180–2186, 2017, doi: 10.1002/hed.24893.

[106] P. Weller, I. Nel, P. Hassenkamp, and T. Gauler, “Detection of Circulating Tumor Cell Subpopulations in Patients with Head and Neck Squamous Cell Carcinoma (HNSCC),” *PLoS One*, vol. 9, no. 12, p. e113706, 2014, doi: 10.1371/journal.pone.0113706.

[107] X. L. Wu, Q. Tu, G. Faure, P. Gallet, C. Kohler, and M. D. C. Bittencourt, “Diagnostic and Prognostic Value of Circulating Tumor Cells in Head and Neck Squamous Cell Carcinoma: A systematic review and meta-analysis,” *Sci. Rep.*, vol. 6, p. 2010, 2016, doi: 10.1038/srep20210.

[108] E. Heitzer *et al.*, "Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing," *Cancer Res.*, vol. 73, no. 10, pp. 2965–2975, 2013, doi: 10.1158/0008-5472.CAN-12-4140.
[109] T. M. Gorges *et al.*, "Accession of Tumor Heterogeneity by Multiplex Transcriptome Profiling of Single Circulating Tumor Cells.," *Clin. Chem.*, vol. 62, no. 11, pp. 1504–1515, Nov. 2016, doi: 10.1373/clinchem.2016.260299.
[110] T. Fehm *et al.*, "HER2 status of circulating tumor cells in patients with metastatic breast cancer: a prospective, multicenter trial.," *Breast Cancer Res. Treat.*, vol. 124, no. 2, pp. 403–12, Nov. 2010, doi: 10.1007/s10549-010-1163-x.
[111] M. DT *et al.*, "Androgen Receptor Signaling in Circulating Tumor Cells as a Marker of Hormonally Responsive Prostate Cancer," *Cancer Discov.*, vol. 2, no. 11, 2012, doi: 10.1158/2159-8290.CD-12-0222.
[112] K. Chikamatsu *et al.*, "Expression of immune-regulatory molecules in circulating tumor cells derived from patients with head and neck squamous cell carcinoma," *Oral Oncol.*, vol. 89, pp. 34–39, Feb. 2019, doi: 10.1016/J.ORALONCOLOGY.2018.12.002.
[113] D. T. Miyamoto *et al.*, "RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance.," *Science*, vol. 349, no. 6254, pp. 1351–6, Sep. 2015, doi: 10.1126/science.aab0917.
[114] A. Kalikaki *et al.*, "KRAS Genotypic Changes of Circulating Tumor Cells during Treatment of Patients with Metastatic Colorectal Cancer," *PLoS One*, vol. 9, no. 8, p. e104902, Aug. 2014, doi: 10.1371/journal.pone.0104902.
[115] S. C. Winter *et al.*, "Long term survival following the detection of circulating tumour cells in head and neck squamous cell carcinoma," *BMC Cancer*, vol. 9, pp. 1–15, 2009, doi: 10.1186/1471-2407-9-424.
[116] P. Balasubramanian *et al.*, "Confocal images of circulating tumor cells obtained using a methodology and technology that removes normal cells.," *Mol. Pharm.*, vol. 6, no. 5, pp. 1402–8, 2009, doi: 10.1021/mp9000519.
[117] P. Balasubramanian *et al.*, "Multiparameter analysis, including EMT markers, on negatively enriched blood samples from patients with squamous cell carcinoma of the head and neck," *PLoS One*, vol. 7, no. 7, 2012, doi: 10.1371/journal.pone.0042048.
[118] J. C. Hsieh *et al.*, "Prognostic value of circulating tumor cells with podoplanin expression in patients with locally advanced or metastatic head and neck squamous cell carcinoma," *Head Neck*, vol. 37, no. 10, pp. 1448–1455, 2015, doi: 10.1002/HED.
[119] I. Tinhofer *et al.*, "Detection of circulating tumor cells for prediction of recurrence after adjuvant chemoradiation in locally advanced squamous cell carcinoma of the head and neck," *Ann. Oncol.*, vol. 25, pp. 2042–2047, 2014, doi: 10.1093/annonc/mdu271.
[120] K. Jatana *et al.*, "Significance of Circulating Tumor Cells in Patients With Squamous Cell Carcinoma of the Head and Neck: Initial Results," *Arch Otolaryngol Head Neck Surg*, vol. 136, no. 12, pp. 1274–1279, 2010, doi: 10.1001/archoto.2010.223.Significance.
[121] S. Grisanti *et al.*, "Circulating tumor cells in patients with recurrent or metastatic head and neck carcinoma: Prognostic and predictive significance," *PLoS One*, vol. 9, no. 8, pp. 1–9, 2014, doi: 10.1371/journal.pone.0103918.
[122] K. Liu, N. Chen, J. Wei, L. Ma, S. Yang, and X. Zhang, "Clinical significance of circulating tumor cells in patients with locally advanced head and neck squamous cell carcinoma.," *Oncol. Rep.*, vol. 43, no. 5, pp. 1525–1535, May 2020, doi: 10.3892/or.2020.7536.
[123] R. Garrel *et al.*, "Circulating Tumor Cells as a Prognostic Factor in Recurrent or Metastatic Head and Neck Squamous Cell Carcinoma: The CIRCUTEC Prospective Study," *Clin. Chem.*, vol. 65, no. 10, pp. 1267–1275, Oct. 2019, doi: 10.1373/clinchem.2019.305904.
[124] J. Inhestern *et al.*, "Prognostic Role of Circulating Tumor Cells during Induction Chemotherapy Followed by Curative Surgery Combined with Postoperative Radiotherapy in Patients with Locally Advanced Oral and Oropharyngeal Squamous Cell Cancer," *PLoS One*, vol. 10, no. 7, p.

e0132901, 2015, doi: 10.1371/journal.pone.0132901.
[125] M. Schwaederle *et al.*, "Impact of Precision Medicine in Diverse Cancers: A Meta-Analysis of Phase II Clinical Trials," *J. Clin. Oncol.*, vol. 33, no. 32, 2015, doi: 10.1200/JCO.2015.61.5997.
[126] M. Schwaederle *et al.*, "Association of Biomarker-Based Treatment Strategies With Response Rates and Progression-Free Survival in Refractory Malignant Neoplasms," *JAMA Oncol.*, vol. 2, no. 11, p. 1452, Nov. 2016, doi: 10.1001/jamaoncol.2016.2129.
[127] T. Y. Seiwert *et al.*, "Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial," *Lancet Oncol.*, vol. 17, no. 7, pp. 956–965, 2016, doi: 10.1016/S1470-2045(16)30066-3.
[128] L. Q. M. Chow *et al.*, "Antitumor Activity of Pembrolizumab in Biomarker-Unselected Patients With Recurrent and/or Metastatic Head and Neck Squamous Cell Carcinoma: Results From the Phase Ib KEYNOTE-012 Expansion Cohort.," *J. Clin. Oncol.*, vol. 34, no. 32, pp. 3838–3845, 2016, doi: 10.1200/JCO.2016.68.1478.
[129] C. Nicolazzo *et al.*, "Monitoring PD-L1 positive circulating tumor cells in non-small cell lung cancer patients treated with the PD-1 inhibitor Nivolumab," *Sci. Rep.*, vol. 6, no. 1, p. 31726, Aug. 2016, doi: 10.1038/srep31726.
[130] H. Tada, H. Takahashi, Y. Kuwabara-Yokobori, M. Shino, and K. Chikamatsu, "Molecular profiling of circulating tumor cells predicts clinical outcome in head and neck squamous cell carcinoma," *Oral Oncol.*, vol. 102, p. 104558, Mar. 2020, doi: 10.1016/j.oraloncology.2019.104558.
[131] M. Tellez-Gabriel, D. Cochonneau, M. Cadé, C. Jubellin, M.-F. Heymann, and D. Heymann, "Circulating Tumor Cell-Derived Pre-Clinical Models for Personalized Medicine.," *Cancers (Basel).*, vol. 11, no. 1, Dec. 2018, doi: 10.3390/cancers11010019.
[132] Y. Zhang *et al.*, "Microfluidic chip for isolation of viable circulating tumor cells of hepatocellular carcinoma for their culture and drug sensitivity assay," *Cancer Biol. Ther.*, vol. 17, no. 11, pp. 1177–1187, Nov. 2016, doi: 10.1080/15384047.2016.1235665.
[133] C. L. Hodgkinson *et al.*, "Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer.," *Nat. Med.*, vol. 20, no. 8, pp. 897–903, Aug. 2014, doi: 10.1038/nm.3600.
[134] B. J. Drapkin *et al.*, "Genomic and Functional Fidelity of Small Cell Lung Cancer Patient-Derived Xenografts," *Cancer Discov.*, vol. 8, no. 5, pp. 600–615, May 2018, doi: 10.1158/2159-8290.CD-17-0935.
[135] C.-J. Liao, C.-H. Hsieh, F.-C. Hung, H.-M. Wang, W.-P. Chou, and M.-H. Wu, "The Integration of a Three-Dimensional Spheroid Cell Culture Operation in a Circulating Tumor Cell (CTC) Isolation and Purification Process: A Preliminary Study of the Clinical Significance and Prognostic Role of the CTCs Isolated from the Blood Samples o," doi: 10.3390/cancers11060783.
[136] M. D. Forster and M.-J. Devlin, "Immune Checkpoint Inhibition in Head and Neck Cancer," *Front. Oncol.*, vol. 8, no. August, pp. 1–9, 2018, doi: 10.3389/fonc.2018.00310.
[137] E. E. W. Cohen *et al.*, "The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of squamous cell carcinoma of the head and neck (HNSCC)," *J. Immunother. Cancer*, vol. 7, no. 1, p. 184, Dec. 2019, doi: 10.1186/s40425-019-0662-5.
[138] M. Ayers *et al.*, "IFN- g – related mRNA profile predicts clinical response to PD-1 blockade Find the latest version : IFN- γ – related mRNA profile predicts clinical response to PD-1 blockade," *J. Clin. Invest.*, vol. 127, no. 8, pp. 2930–2940, 2017, doi: 10.1172/JCI91190.
[139] V. Kloten, R. Lampignano, T. Krahn, and T. Schlange, "Circulating Tumor Cell PD-L1 Expression as Biomarker for Therapeutic Efficacy of Immune Checkpoint Inhibition in NSCLC.," *Cells*, vol. 8, no. 8, 2019, doi: 10.3390/cells8080809.
[140] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred Reporting Items for Systematic

Reviews and Meta-Analyses: The PRISMA Statement," *PLoS Med.*, vol. 6, no. 7, p. e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.
[141] L. M. McShane *et al.*, "REporting recommendations for tumour MARKer prognostic studies (REMARK).," *Br. J. Cancer*, vol. 93, no. 4, pp. 387–91, Aug. 2005, doi: 10.1038/sj.bjc.6602678.
[142] P. M. S. Gurung *et al.*, "Prognostic DNA Methylation Biomarkers in High-risk Non-muscle-invasive Bladder Cancer: A Systematic Review to Identify Loci for Prospective Validation.," *Eur. Urol. Focus*, vol. 6, no. 4, pp. 683–697, 2020, doi: 10.1016/j.euf.2019.02.012.
[143] P. Economopoulou *et al.*, "Surrogates of immunologic cell death (ICD) and chemoradiotherapy outcomes in head and neck squamous cell carcinoma (HNSCC)," *Oral Oncol.*, vol. 94, no. May, pp. 93–100, 2019, doi: 10.1016/j.oraloncology.2019.05.019.
[144] P. Economopoulou *et al.*, "Prognostic impact of indoleamine 2,3-dioxygenase 1 (IDO1) mRNA expression on circulating tumour cells of patients with head and neck squamous cell carcinoma," *ESMO Open*, vol. 5, p. e000646, 2020, doi: 10.1136/esmoopen-2019-000646.
[145] A. Kulasinghe *et al.*, "The prognostic significance of circulating tumor cells in head and neck and non-small-cell lung cancer," *Cancer Med.*, vol. 7, pp. 5910–5919, 2018, doi: 10.1002/cam4.1832.
[146] J. P. Oliveira-Costa *et al.*, "Gene expression patterns through oral squamous cell carcinoma development: PD-L1 expression in primary tumor and circulating tumor cells," *Oncotarget*, vol. 6, no. 25, pp. 20902–20920, Aug. 2015, doi: 10.18632/oncotarget.3939.
[147] W. Yang, M. C. M. Wong, P. J. Thomson, K.-Y. Li, and Y. Su, "The prognostic role of PD-L1 expression for survival in head and neck squamous cell carcinoma: A systematic review and meta-analysis," *Oral Oncol.*, vol. 86, pp. 81–90, Nov. 2018, doi: 10.1016/J.ORALONCOLOGY.2018.09.016.
[148] M. A. Khattak *et al.*, "PD-L1 Expression on Circulating Tumor Cells May Be Predictive of Response to Pembrolizumab in Advanced Melanoma: Results from a Pilot Study," *Oncologist*, vol. 25, no. 3, pp. e520–e527, Mar. 2020, doi: 10.1634/theoncologist.2019-0557.
[149] L. Zhang *et al.*, "PD-L1+ aneuploid circulating tumor endothelial cells (CTECs) exhibit resistance to the checkpoint blockade immunotherapy in advanced NSCLC patients," *Cancer Lett.*, vol. 469, no. July 2019, pp. 355–366, 2020, doi: 10.1016/j.canlet.2019.10.041.
[150] L. Zhai *et al.*, "IDO1 in cancer: a Gemini of immune checkpoints," *Cell. Mol. Immunol.*, vol. 15, no. 5, pp. 447–457, May 2018, doi: 10.1038/cmi.2017.143.
[151] X. Wang, Q. Sun, Q. Liu, C. Wang, R. Yao, and Y. Wang, "CTC immune escape mediated by PD-L1.," *Med. Hypotheses*, vol. 93, pp. 138–9, Aug. 2016, doi: 10.1016/j.mehy.2016.05.022.
[152] C. Yue *et al.*, "Dynamic change of PD-L1 expression on circulating tumor cells in advanced solid tumor patients undergoing PD-1 blockade therapy," *Oncoimmunology*, vol. 7, no. 7, 2018, doi: 10.1080/2162402X.2018.1438111.
[153] N. Guibert *et al.*, "PD-L1 Expression in Circulating Tumor Cells of Advanced Non-Small Cell Lung Cancer Patients Treated With Nivolumab," *Lung Cancer*, vol. 120, 2018, doi: 10.1016/J.LUNGCAN.2018.04.001.
[154] M. Janning *et al.*, "Determination of PD-L1 Expression in Circulating Tumor Cells of NSCLC Patients and Correlation with Response to PD-1/PD-L1 Inhibitors.," *Cancers (Basel).*, vol. 11, no. 6, Jun. 2019, doi: 10.3390/cancers11060835.
[155] L. Q. M. Chow *et al.*, "Antitumor Activity of Pembrolizumab in Biomarker-Unselected Patients With Recurrent and/or Metastatic Head and Neck Squamous Cell Carcinoma: Results From the Phase Ib KEYNOTE-012 Expansion Cohort.," *J. Clin. Oncol.*, vol. 34, no. 32, pp. 3838–3845, 2016, doi: 10.1200/JCO.2016.68.1478.
[156] S. Gettinger *et al.*, "Five-Year Follow-Up of Nivolumab in Previously Treated Advanced Non-Small-Cell Lung Cancer: Results From the CA209-003 Study," *J. Clin. Oncol.*, vol. 36, no. 17, 2018, doi: 10.1200/JCO.2017.77.0412.
[157] F. R. Hirsch *et al.*, "PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project.," *J. Thorac. Oncol.*, vol. 12, no. 2, pp.

208–222, 2017, doi: 10.1016/j.jtho.2016.11.2228.
[158] J. McLaughlin *et al.*, "Quantitative Assessment of the Heterogeneity of PD-L1 Expression in Non-Small-Cell Lung Cancer.," *JAMA Oncol.*, vol. 2, no. 1, pp. 46–54, Jan. 2016, doi: 10.1001/jamaoncol.2015.3638.
[159] X. Wang, G. Zhang, and Q. Sun, "PD-L1 expression on circulating tumor cells and prognosis of breast cancer patients.," *J. Clin. Oncol.*, vol. 37, no. 15_suppl, pp. e14028–e14028, May 2019, doi: 10.1200/JCO.2019.37.15_suppl.e14028.
[160] J. Bryant *et al.*, "Repurposed quinacrine synergizes with cisplatin, reducing the effective dose required for treatment of head and neck squamous cell carcinoma," *Oncotarget*, vol. 10, no. 50, pp. 5229–5244, Aug. 2019, doi: 10.18632/oncotarget.27156.
[161] "Resources - ANGLE PLC." https://angleplc.com/library/resources/#guides (accessed Oct. 11, 2021).
[162] Miltenyi Biotec, "CD45 MicroBeads, human - Tumor cells - MicroBeads and Isolation Kits - Cell separation reagents - MACS Cell Separation - Products - Miltenyi Biotec - Great Britain." https://www.miltenyibiotec.com/GB-en/products/cd45-microbeads-human.html#gref (accessed May 26, 2020).
[163] L. Han *et al.*, "Single-cell mass cytometry reveals intracellular survival/proliferative signaling in FLT3-ITD-mutated AML stem/progenitor cells.," *Cytometry. A*, vol. 87, no. 4, pp. 346–56, Apr. 2015, doi: 10.1002/cyto.a.22628.
[164] A. Dobin *et al.*, "STAR: ultrafast universal RNA-seq aligner.," *Bioinformatics*, vol. 29, no. 1, pp. 15–21, Jan. 2013, doi: 10.1093/bioinformatics/bts635.
[165] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.," *Genome Biol.*, vol. 15, no. 12, p. 550, 2014, doi: 10.1186/s13059-014-0550-8.
[166] W. Luo, M. S. Friedman, K. Shedden, K. D. Hankenson, and P. J. Woolf, "GAGE: generally applicable gene set enrichment for pathway analysis.," *BMC Bioinformatics*, vol. 10, p. 161, May 2009, doi: 10.1186/1471-2105-10-161.
[167] V. Kelly, A. Al-Rawi, D. Lewis, G. Kustatscher, and T. Ly, "Low cell number proteomic analysis using in-cell protease digests reveals a robust signature for cell cycle state classification.," *Mol. Cell. Proteomics*, vol. 0, no. 0, p. 100169, Nov. 2021, doi: 10.1016/j.mcpro.2021.100169.
[168] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nat. Biotechnol.*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008, doi: 10.1038/nbt.1511.
[169] Angle Plc, "How It Works - ANGLE PLC." https://angleplc.com/parsortix-technology/how-it-works/?gclid=Cj0KCQjw-O35BRDVARIsAJU5mQW5ny2n6N1cLfEJEl9YDHivWdwwNC2U37I2Rlaiejgtkufr0FvmTwoaAri5EALw_wcB (accessed Aug. 18, 2020).
[170] G. E. Hvichia *et al.*, "A novel microfluidic platform for size and deformability based separation and the subsequent molecular characterization of viable circulating tumor cells," *Int. J. Cancer*, vol. 138, no. 12, pp. 2894–2904, 2016, doi: 10.1002/ijc.30007.
[171] J. Chudziak *et al.*, "Clinical evaluation of a novel microfluidic device for epitope-independent enrichment of circulating tumour cells in patients with small cell lung cancer," *Analyst*, vol. 141, no. 2, pp. 669–678, 2016, doi: 10.1039/c5an02156a.
[172] Y. Maertens *et al.*, "Comparison of isolation platforms for detection of circulating renal cell carcinoma cells," *Oncotarget*, vol. 8, no. 50, pp. 87710–87717, 2017, doi: 10.18632/oncotarget.21197.
[173] R. Lampignano *et al.*, "A novel workflow to enrich and isolate patient-matched EpCAMhighand EpCAMlow/negativeCTCs enables the comparative characterization of the PIK3CA status in metastatic breast cancer," *Int. J. Mol. Sci.*, vol. 18, no. 9, 2017, doi: 10.3390/ijms18091885.
[174] E. Obermayr *et al.*, "Molecular Characterization of Circulating Tumor Cells Enriched by A

Microfluidic Platform in Patients with Small-Cell Lung Cancer," *Cells*, vol. 8, no. 8, p. 880, Aug. 2019, doi: 10.3390/cells8080880.
[175] L. Xu *et al.*, "Noninvasive Detection of Clinically Significant Prostate Cancer Using Circulating Tumor Cells," *J. Urol.*, p. 101097JU0000000000000475, Aug. 2019, doi: 10.1097/JU.0000000000000475.
[176] K. Gorges *et al.*, "Intra-Patient Heterogeneity of Circulating Tumor Cells and Circulating Tumor DNA in Blood of Melanoma Patients," *Cancers (Basel).*, vol. 11, no. 11, p. 1685, Oct. 2019, doi: 10.3390/cancers11111685.
[177] M. Zavridou *et al.*, "Direct comparison of size-dependent versus EpCAM-dependent CTC enrichment at the gene expression and DNA methylation level in head and neck squamous cell carcinoma.," *Sci. Rep.*, vol. 10, no. 1, p. 6551, Apr. 2020, doi: 10.1038/s41598-020-63055-y.
[178] A. J. Hobro and N. I. Smith, "An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging," *Vib. Spectrosc.*, vol. 91, pp. 31–45, Jul. 2017, doi: 10.1016/J.VIBSPEC.2016.10.012.
[179] L. Rochette, M. Zeller, Y. Cottin, and C. Vergely, "Insights Into Mechanisms of GDF15 and Receptor GFRAL: Therapeutic Targets.," *Trends Endocrinol. Metab.*, vol. 31, no. 12, pp. 939–951, Dec. 2020, doi: 10.1016/j.tem.2020.10.004.
[180] I. Hollerer *et al.*, "The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates mammalian mRNA expression in a quantitative and qualitative fashion.," *RNA*, vol. 22, no. 9, pp. 1441–53, 2016, doi: 10.1261/rna.055657.115.
[181] W. Li *et al.*, "MALAT1 Promotes Cell Apoptosis and Suppresses Cell Proliferation in Testicular Ischemia-Reperfusion Injury by Sponging MiR-214 to Modulate TRPV4 Expression.," *Cell. Physiol. Biochem.*, vol. 46, no. 2, pp. 802–814, 2018, doi: 10.1159/000488738.
[182] J. Li, Z. Yue, W. Xiong, P. Sun, K. You, and J. Wang, "TXNIP overexpression suppresses proliferation and induces apoptosis in SMMC7221 cells through ROS generation and MAPK pathway activation," *Oncol. Rep.*, vol. 37, no. 6, pp. 3369–3376, Jun. 2017, doi: 10.3892/or.2017.5577.
[183] C.-P. Kung, S. Khaku, M. Jennis, Y. Zhou, and M. E. Murphy, "Identification of TRIML2, a novel p53 target, that enhances p53 SUMOylation and regulates the transactivation of proapoptotic genes.," *Mol. Cancer Res.*, vol. 13, no. 2, pp. 250–62, Feb. 2015, doi: 10.1158/1541-7786.MCR-14-0385.
[184] S. Gencer, A. Cebeci, and M. B. Irmak-Yazicioglu, "Matrix metalloproteinase gene expressions might be oxidative stress targets in gastric cancer cell lines.," *Chin. J. Cancer Res.*, vol. 25, no. 3, pp. 322–33, Jun. 2013, doi: 10.3978/j.issn.1000-9604.2013.06.05.
[185] J. Liu *et al.*, "Global Gene Expression Profiling Reveals Functional Importance of Sirt2 in Endothelial Cells under Oxidative Stress.," *Int. J. Mol. Sci.*, vol. 14, no. 3, pp. 5633–49, Mar. 2013, doi: 10.3390/ijms14035633.
[186] M. J. VandeKopple, J. Wu, E. N. Auer, A. J. Giaccia, N. C. Denko, and I. Papandreou, "HILPDA Regulates Lipid Metabolism, Lipid Droplet Abundance, and Response to Microenvironmental Stress in Solid Tumors.," *Mol. Cancer Res.*, vol. 17, no. 10, pp. 2089–2101, 2019, doi: 10.1158/1541-7786.MCR-18-1343.
[187] E. Association for Research in Vision and Ophthalmology., M. Hiraoka, L. R. Lin, Y. S. Ho, and V. N. Reddy, *Investigative ophthalmology & visual science.*, vol. 45, no. 13. C.V. Mosby Co, 1977.
[188] K. Sinha, J. Das, P. B. Pal, and P. C. Sil, "Oxidative stress: the mitochondria-dependent and mitochondria-independent pathways of apoptosis," *Arch. Toxicol.*, vol. 87, no. 7, pp. 1157–1180, Jul. 2013, doi: 10.1007/s00204-013-1034-4.
[189] M. Haque, J. K. Kendal, R. M. MacIsaac, and D. J. Demetrick, "WSB1: from homeostasis to hypoxia.," *J. Biomed. Sci.*, vol. 23, no. 1, p. 61, Aug. 2016, doi: 10.1186/s12929-016-0270-3.

[190] Qiagen, “RNeasy micro kit.” https://www.qiagen.com/ie/products/top-sellers/rneasy-micro-kit/#orderinginformation (accessed Sep. 09, 2020).
[191] ThermoFisher, “PicoPure™ RNA Isolation Kit.” https://www.thermofisher.com/order/catalog/product/KIT0204#/KIT0204 (accessed Aug. 05, 2021).
[192] P. Moll, M. Ante, A. Seitz, and T. Reda, “QuantSeq 3′ mRNA sequencing for RNA quantification,” *Nat. Methods*, vol. 11, no. 12, pp. i–iii, Dec. 2014, doi: 10.1038/nmeth.f.376.
[193] A. Kulasinghe *et al.*, “Epithelial-mesenchymal axis in head and neck cancer cell lines,” *J. Solid Tumors*, vol. 6, no. 1, p. 28, Dec. 2015, doi: 10.5430/jst.v6n1p28.
[194] “The Human Protein Atlas.” https://www.proteinatlas.org/ENSG00000119888-EPCAM/blood (accessed Sep. 13, 2021).
[195] G. Barriere, P. Fici, G. Gallerani, F. Fabbri, W. Zoli, and M. Rigaud, “Circulating tumor cells and epithelial, mesenchymal and stemness markers: characterization of cell subpopulations.,” *Ann. Transl. Med.*, vol. 2, no. 11, p. 109, Nov. 2014, doi: 10.3978/j.issn.2305-5839.2014.10.04.
[196] M. P. Mak *et al.*, “A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition,” *Clin Cancer Res*, vol. 22, no. 3, 2016, doi: 10.1158/1078-0432.CCR-15-0876.
[197] N. Leelatian, K. Diggins, and J. Irish, “Characterizing Phenotypes and Signaling Networks of Single Human Cells by Mass Cytometry,” *Curr. Opin. Biotechnol.*, vol. 23, no. 1, pp. 83–88, 2012, doi: 10.1016/j.copbio.2011.11.023.
[198] N. Leelatian *et al.*, “Single cell analysis of human tissues and solid tumors with mass cytometry,” *Cytometry B. Clin. Cytom.*, vol. 92, no. 1, p. 68, 2017, doi: 10.1002/CYTO.B.21481.
[199] I. Poláková *et al.*, “Implementation of Mass Cytometry for Immunoprofiling of Patients with Solid Tumors.,” *J. Immunol. Res.*, vol. 2019, p. 6705949, 2019, doi: 10.1155/2019/6705949.
[200] A. F. Jarnuczak *et al.*, “An integrated landscape of protein expression in human cancer,” *Sci. Data*, vol. 8, no. 1, p. 115, Dec. 2021, doi: 10.1038/s41597-021-00890-2.
[201] P. Akan *et al.*, “Comprehensive analysis of the genome transcriptome and proteome landscapes of three tumor cell lines,” *Genome Med.*, vol. 4, no. 11, p. 86, Nov. 2012, doi: 10.1186/gm387.
[202] D. Wang *et al.*, “A deep proteome and transcriptome abundance atlas of 29 healthy human tissues.,” *Mol. Syst. Biol.*, vol. 15, no. 2, p. e8503, 2019, doi: 10.15252/msb.20188503.
[203] F. Ardito, M. Giuliani, D. Perrone, G. Troiano, and L. Lo Muzio, “The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review),” *Int. J. Mol. Med.*, vol. 40, no. 2, pp. 271–280, 2017, doi: 10.3892/ijmm.2017.3036.
[204] N. Leelatian, K. E. Diggins, and J. M. Irish, “Characterizing Phenotypes and Signaling Networks of Single Human Cells by Mass Cytometry.,” *Methods Mol. Biol.*, vol. 1346, pp. 99–113, 2015, doi: 10.1007/978-1-4939-2987-0_8.
[205] G. B *et al.*, “Clinical recovery from surgery correlates with single-cell immune signatures,” *Sci. Transl. Med.*, vol. 6, no. 255, 2014, doi: 10.1126/SCITRANSLMED.3009701.
[206] A. Semaan *et al.*, “Characterisation of circulating tumour cell phenotypes identifies a partial-EMT sub-population for clinical stratification of pancreatic cancer,” *Br. J. Cancer*, vol. 124, no. 12, pp. 1970–1977, Jun. 2021, doi: 10.1038/s41416-021-01350-9.
[207] J. Kitz, D. Goodale, C. Postenka, L. E. Lowes, and A. L. Allan, “EMT-independent detection of circulating tumor cells in human blood samples and pre-clinical mouse models of metastasis,” *Clin. Exp. Metastasis*, vol. 38, no. 1, pp. 97–108, Feb. 2021, doi: 10.1007/s10585-020-10070-y.
[208] H. TADA, H. TAKAHASHI, S. IDA, Y. NAGATA, and K. CHIKAMATSU, “Epithelial–Mesenchymal Transition Status of Circulating Tumor Cells Is Associated With Tumor Relapse in Head and Neck Squamous Cell Carcinoma,” *Anticancer Res.*, vol. 40, no. 6, pp. 3559–3564, 2020, doi: 10.21873/anticanres.14345.
[209] A. Smith, T. N. Teknos, and Q. Pan, “Epithelial to Mesenchymal Transition in Head and Neck Squamous Cell Carcinoma,” *Oral Oncol.*, vol. 49, no. 4, p. 287, 2013, doi:

10.1016/J.ORALONCOLOGY.2012.10.009.
[210] S. Kaufhold and B. Bonavida, “Central role of Snail1 in the regulation of EMT and resistance in cancer: a target for therapeutic intervention,” *J. Exp. Clin. Cancer Res.*, vol. 33, no. 1, p. 62, Dec. 2014, doi: 10.1186/s13046-014-0062-0.
[211] V. Modur, P. Joshi, D. Nie, K. T. Robbins, A. U. Khan, and K. Rao, “CD24 Expression May Play a Role as a Predictive Indicator and a Modulator of Cisplatin Treatment Response in Head and Neck Squamous Cellular Carcinoma.,” *PLoS One*, vol. 11, no. 6, p. e0156651, 2016, doi: 10.1371/journal.pone.0156651.
[212] D. Loreth *et al.*, “CD74 and CD44 Expression on CTCs in Cancer Patients with Brain Metastasis.,” *Int. J. Mol. Sci.*, vol. 22, no. 13, Jun. 2021, doi: 10.3390/ijms22136993.
[213] H. Xu, M. Niu, X. Yuan, K. Wu, and A. Liu, “CD44 as a tumor biomarker and therapeutic target,” *Exp. Hematol. Oncol.*, vol. 9, no. 1, p. 36, Dec. 2020, doi: 10.1186/s40164-020-00192-0.
[214] D. Wangsa *et al.*, “Ki-67 expression predicts locoregional recurrence in stage I oral tongue carcinoma,” *Br. J. Cancer*, vol. 99, no. 7, pp. 1121–1128, Oct. 2008, doi: 10.1038/sj.bjc.6604633.
[215] H. Peng *et al.*, “Clinical significance of Ki67 and circulating tumor cells with an epithelial-mesenchymal transition phenotype in non-small cell lung cancer.,” *Am. J. Transl. Res.*, vol. 12, no. 6, pp. 2916–2928, 2020, Accessed: Dec. 08, 2021. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32655819.
[216] A. Kulasinghe *et al.*, “Enrichment of circulating head and neck tumour cells using spiral microfluidic technology,” *Sci. Rep.*, vol. 7, no. 1, p. 42517, Mar. 2017, doi: 10.1038/srep42517.
[217] A. Strati *et al.*, “Prognostic significance of PD-L1 expression on circulating tumor cells in patients with head and neck squamous cell carcinoma,” *Ann. Oncol.*, vol. 28, no. 8, pp. 1923–1933, 2017, doi: 10.1093/annonc/mdx206.
[218] K. Koike *et al.*, “Prognostic value of FoxP3 and CTLA-4 expression in patients with oral squamous cell carcinoma,” *PLoS One*, vol. 15, no. 8, p. e0237465, Aug. 2020, doi: 10.1371/journal.pone.0237465.
[219] A. K. Moesta, X.-Y. Li, and M. J. Smyth, “Targeting CD39 in cancer,” *Nat. Rev. Immunol.*, vol. 20, no. 12, pp. 739–755, Dec. 2020, doi: 10.1038/s41577-020-0376-4.
[220] S. Chen *et al.*, “CD73: an emerging checkpoint for cancer immunotherapy.,” *Immunotherapy*, vol. 11, no. 11, pp. 983–997, 2019, doi: 10.2217/imt-2018-0200.
[221] Y. Verhoeven *et al.*, “The potential and controversy of targeting STAT family members in cancer,” *Semin. Cancer Biol.*, vol. 60, pp. 41–56, Feb. 2020, doi: 10.1016/J.SEMCANCER.2019.10.002.
[222] S. J. Thomas, J. A. Snowden, M. P. Zeidler, and S. J. Danson, “The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours,” *Br. J. Cancer*, vol. 113, no. 3, pp. 365–371, Jul. 2015, doi: 10.1038/bjc.2015.233.
[223] K. Meissl, S. Macho-Maschler, M. Müller, and B. Strobl, “The good and the bad faces of STAT1 in solid tumours,” *Cytokine*, vol. 89, pp. 12–20, Jan. 2017, doi: 10.1016/J.CYTO.2015.11.011.
[224] J. L. Geiger, J. R. Grandis, and J. E. Bauman, “The STAT3 pathway as a therapeutic target in head and neck cancer: Barriers and innovations.,” *Oral Oncol.*, vol. 56, pp. 84–92, May 2016, doi: 10.1016/j.oraloncology.2015.11.022.
[225] H. Stegeman *et al.*, “Combining radiotherapy with MEK1/2, STAT5 or STAT6 inhibition reduces survival of head and neck cancer lines,” *Mol. Cancer*, vol. 12, no. 1, p. 133, Nov. 2013, doi: 10.1186/1476-4598-12-133.
[226] K. Do and A. P. Chen, “Molecular pathways: targeting PARP in cancer treatment.,” *Clin. Cancer Res.*, vol. 19, no. 5, pp. 977–84, Mar. 2013, doi: 10.1158/1078-0432.CCR-12-0163.
[227] M. Moutafi, P. Economopoulou, D. Rimm, and A. Psyrri, “PARP inhibitors in head and neck cancer: Molecular mechanisms, preclinical and clinical data.,” *Oral Oncol.*, vol. 117, p.

105292, 2021, doi: 10.1016/j.oraloncology.2021.105292.
[228] N. Jiang, Q. Dai, X. Su, J. Fu, X. Feng, and J. Peng, “Role of PI3K/AKT pathway in cancer: the framework of malignant behavior.,” *Mol. Biol. Rep.*, vol. 47, no. 6, pp. 4587–4629, Jun. 2020, doi: 10.1007/s11033-020-05435-1.
[229] H. Wang *et al.*, “Downregulation of FAP suppresses cell proliferation and metastasis through PTEN/PI3K/AKT and Ras-ERK signaling in oral squamous cell carcinoma,” *Cell Death Dis.*, vol. 5, no. 4, pp. e1155–e1155, Apr. 2014, doi: 10.1038/cddis.2014.122.
[230] A. A. Samatar and P. I. Poulikakos, “Targeting RAS–ERK signalling in cancer: promises and challenges,” *Nat. Rev. Drug Discov.*, vol. 13, no. 12, pp. 928–942, Dec. 2014, doi: 10.1038/nrd4281.
[231] Y. Guo, W. Pan, S. Liu, Z. Shen, Y. Xu, and L. Hu, “ERK/MAPK signalling pathway and tumorigenesis (Review),” *Exp. Ther. Med.*, vol. 19, no. 3, pp. 1997–2007, Jan. 2020, doi: 10.3892/etm.2020.8454.
[232] C. Rong *et al.*, “Adaptive ERK signalling activation in response to therapy and in silico prognostic evaluation of EGFR-MAPK in HNSCC,” *Br. J. Cancer*, vol. 123, no. 2, pp. 288–297, Jul. 2020, doi: 10.1038/s41416-020-0892-9.
[233] K. M. Sakamoto and D. A. Frank, “CREB in the pathophysiology of cancer: implications for targeting transcription factors for cancer therapy.,” *Clin. Cancer Res.*, vol. 15, no. 8, pp. 2583–7, Apr. 2009, doi: 10.1158/1078-0432.CCR-08-1137.
[234] L. Sapio, A. Salzillo, A. Ragone, M. Illiano, A. Spina, and S. Naviglio, “Targeting CREB in Cancer Therapy: A Key Candidate or One of Many? An Update.,” *Cancers (Basel).*, vol. 12, no. 11, Oct. 2020, doi: 10.3390/cancers12113166.
[235] A. Martínez-Limón, M. Joaquin, M. Caballero, F. Posas, and E. de Nadal, “The p38 Pathway: From Biology to Cancer Therapy.,” *Int. J. Mol. Sci.*, vol. 21, no. 6, Mar. 2020, doi: 10.3390/ijms21061913.
[236] R. Chaisuparat, S. Rojanawatsirivej, and S. Yodsanga, “Ribosomal protein S6 phosphorylation is associated with epithelial dysplasia and squamous cell carcinoma of the oral cavity.,” *Pathol. Oncol. Res.*, vol. 19, no. 2, pp. 189–93, Apr. 2013, doi: 10.1007/s12253-012-9568-y.
[237] S. Kang *et al.*, “p90 ribosomal S6 kinase 2 promotes invasion and metastasis of human head and neck squamous cell carcinoma cells.,” *J. Clin. Invest.*, vol. 120, no. 4, pp. 1165–77, Apr. 2010, doi: 10.1172/JCI40582.
[238] Cytomark Plc, “TransFix Compatible Antibodies – Cytomark.” https://www.cytomark.co.uk/2019/10/23/transfix-compatible-antibodies/ (accessed Oct. 21, 2021).
[239] J. W. Mandell, “Phosphorylation state-specific antibodies: applications in investigative and diagnostic pathology.,” *Am. J. Pathol.*, vol. 163, no. 5, pp. 1687–98, Nov. 2003, doi: 10.1016/S0002-9440(10)63525-0.
[240] abcam, “KD value: a quantitative measurement of antibody affinity | Abcam.” https://www.abcam.com/primary-antibodies/kd-value-a-quantitive-measurement-of-antibody-affinity (accessed May 08, 2022).
[241] C. A. W. Geuijen *et al.*, “Affinity ranking of antibodies using flow cytometry: application in antibody phage display-based target discovery.,” *J. Immunol. Methods*, vol. 302, no. 1–2, pp. 68–77, Jul. 2005, doi: 10.1016/j.jim.2005.04.016.
[242] C. B. Bagwell *et al.*, “Automated Data Cleanup for Mass Cytometry,” *Cytom. Part A*, vol. 97, no. 2, pp. 184–198, Feb. 2020, doi: 10.1002/cyto.a.23926.
[243] Fluidigm, “Impact of Cryopreservation on Performance of the Maxpar Direct Immune Profiling System.” Accessed: May 05, 2022. [Online]. Available: https://assets.fluidigm.com/asset/260.
[244] A. R. Schulz, S. Baumgart, J. Schulze, M. Urbicht, A. Grützkau, and H. E. Mei, “Stabilizing Antibody Cocktails for Mass Cytometry,” *Cytom. Part A*, vol. 95, no. 8, p. cyto.a.23781, May 2019, doi: 10.1002/cyto.a.23781.

[245] H. R. Sumatoh, K. W. W. Teng, Y. Cheng, and E. W. Newell, “Optimization of mass cytometry sample cryopreservation after staining,” *Cytom. Part A*, vol. 91, no. 1, pp. 48–61, Jan. 2017, doi: 10.1002/cyto.a.23014.
[246] K. Quintelier, A. Couckuyt, A. Emmaneel, J. Aerts, Y. Saeys, and S. Van Gassen, “Analyzing high-dimensional cytometry data using FlowSOM,” *Nat. Protoc.*, vol. 16, no. 8, pp. 3775–3801, Aug. 2021, doi: 10.1038/s41596-021-00550-0.
[247] A. K. Kimball, L. M. Oko, B. L. Bullock, R. A. Nemenoff, L. F. van Dyk, and E. T. Clambey, “A Beginner’s Guide to Analyzing and Visualizing Mass Cytometry Data,” *J. Immunol.*, vol. 200, no. 1, pp. 3–22, 2017, doi: 10.4049/jimmunol.1701494.
[248] X. Liu *et al.*, “A comparison framework and guideline of clustering methods for mass cytometry data,” doi: 10.1186/s13059-019-1917-7.
[249] J. Chen, J. Zhou, J. Lu, H. Xiong, X. Shi, and L. Gong, “Significance of CD44 expression in head and neck cancer: a systemic review and meta-analysis,” *BMC Cancer*, vol. 14, p. 15, 2014, doi: 10.3109/00016489.2014.913311.
[250] S. Li *et al.*, “An Integrated Platform for Isolation, Processing, and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood.,” *Mol. Cell. Proteomics*, vol. 14, no. 6, pp. 1672–83, Jun. 2015, doi: 10.1074/mcp.M114.045724.
[251] Y. Zhu *et al.*, “Proteome Profiling of 1 to 5 Spiked Circulating Tumor Cells Isolated from Whole Blood Using Immunodensity Enrichment, Laser Capture Microdissection, Nanodroplet Sample Processing, and Ultrasensitive nanoLC–MS,” *Anal. Chem.*, vol. 90, no. 20, pp. 11756–11759, Oct. 2018, doi: 10.1021/acs.analchem.8b03268.
[252] T. Ly *et al.*, “Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS,” 2017, doi: 10.7554/eLife.27574.001.
[253] A. Luk *et al.*, “CTC-mRNA (AR-V7) Analysis from Blood Samples—Impact of Blood Collection Tube and Storage Time,” *Int. J. Mol. Sci.*, vol. 18, no. 5, p. 1047, May 2017, doi: 10.3390/ijms18051047.
[254] C. Koch *et al.*, “Pre-Analytical and Analytical Variables of Label-Independent Enrichment and Automated Detection of Circulating Tumor Cells in Cancer Patients,” *Cancers (Basel).*, vol. 12, no. 2, p. 442, Feb. 2020, doi: 10.3390/cancers12020442.
[255] A. B. Alber and D. M. Suter, “Dynamics of protein synthesis and degradation through the cell cycle.,” *Cell Cycle*, vol. 18, no. 8, pp. 784–794, 2019, doi: 10.1080/15384101.2019.1598725.
[256] C. Bertoli, J. M. Skotheim, and R. A. M. de Bruin, “Control of cell cycle transcription during G1 and S phases.,” *Nat. Rev. Mol. Cell Biol.*, vol. 14, no. 8, pp. 518–28, Aug. 2013, doi: 10.1038/nrm3629.
[257] H. A. Cognart, J.-L. Viovy, and C. Villard, “Fluid shear stress coupled with narrow constrictions induce cell type-dependent morphological and molecular changes in SK-BR-3 and MDA-MB-231 cells,” *Sci. Rep.*, vol. 10, no. 1, p. 6386, Dec. 2020, doi: 10.1038/s41598-020-63316-w.
[258] B. H. L. Harris, A. Barberis, C. M. L. West, and F. M. Buffa, “Gene Expression Signatures as Biomarkers of Tumour Hypoxia.,” *Clin. Oncol. (R. Coll. Radiol).*, vol. 27, no. 10, pp. 547–60, Oct. 2015, doi: 10.1016/j.clon.2015.07.004.
[259] J. M. Brooks *et al.*, “Development and Validation of a Combined Hypoxia and Immune Prognostic Classifier for Head and Neck Cancer,” *Clin. Cancer Res.*, vol. 25, no. 17, pp. 5315–5328, Sep. 2019, doi: 10.1158/1078-0432.CCR-18-3314.
[260] V. Karantza, “Keratins in health and cancer: more than mere epithelial cell markers,” *Oncogene*, vol. 30, no. 2, pp. 127–138, Jan. 2011, doi: 10.1038/onc.2010.456.
[261] M. J. M. Magbanua and J. W. Park, “Isolation of circulating tumor cells by immunomagnetic enrichment and fluorescence-activated cell sorting (IE/FACS) for molecular profiling.,” *Methods*, vol. 64, no. 2, pp. 114–8, Dec. 2013, doi: 10.1016/j.ymeth.2013.07.029.
[262] A. Ring *et al.*, “EpCAM based capture detects and recovers circulating tumor cells from all subtypes of breast cancer except claudin-low.,” *Oncotarget*, vol. 6, no. 42, pp. 44623–34, Dec. 2015, doi: 10.18632/oncotarget.5977.

[263] Y. Li, X. Cheng, Z. Chen, Y. Liu, Z. Liu, and S. Xu, “Circulating tumor cells in peripheral and pulmonary venous blood predict poor long-term survival in resected non-small cell lung cancer patients.,” *Sci. Rep.*, vol. 7, no. 1, p. 4971, 2017, doi: 10.1038/s41598-017-05154-x.
[264] T. Hristozova *et al.*, “The presence of circulating tumor cells (CTCs) correlates with lymph node metastasis in nonresectable squamous cell carcinoma of the head and neck region,” *Ann. Oncol.*, vol. 22, pp. 1878–1885, 2011, doi: 10.1093/annonc/mdr130.
[265] A. Lopresti *et al.*, “Sensitive and easy screening for circulating tumor cells by flow cytometry.,” *JCI insight*, vol. 5, no. 14, 2019, doi: 10.1172/jci.insight.128180.
[266] Z. Wang, K. Cui, Y. Xue, F. Tong, and S. Li, “Prognostic value of circulating tumor cells in patients with squamous cell carcinoma of the head and neck: a systematic review and meta-analysis,” *Med. Oncol.*, vol. 32, no. 5, 2015, doi: 10.1007/s12032-015-0579-x.
[267] T. C. Brown, N. V. Sankpal, and W. E. Gillanders, “Functional Implications of the Dynamic Regulation of EpCAM during Epithelial-to-Mesenchymal Transition,” *Biomolecules*, vol. 11, no. 7, p. 956, Jun. 2021, doi: 10.3390/biom11070956.
[268] K.-A. Hyun *et al.*, “Epithelial-to-mesenchymal transition leads to loss of EpCAM and different physical properties in circulating tumor cells from metastatic breast cancer.,” *Oncotarget*, vol. 7, no. 17, pp. 24677–87, Apr. 2016, doi: 10.18632/oncotarget.8250.
[269] P. Savagner, “The epithelial-mesenchymal transition (EMT) phenomenon.,” *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, vol. 21 Suppl 7, pp. vii89-92, Oct. 2010, doi: 10.1093/annonc/mdq292.
[270] S. Bhatia, P. Wang, A. Toh, and E. W. Thompson, “New Insights Into the Role of Phenotypic Plasticity and EMT in Driving Cancer Progression,” *Front. Mol. Biosci.*, vol. 7, p. 71, Apr. 2020, doi: 10.3389/fmolb.2020.00071.
[271] I. Pastushenko and C. Blanpain, “EMT Transition States during Tumor Progression and Metastasis.,” *Trends Cell Biol.*, vol. 29, no. 3, pp. 212–226, 2019, doi: 10.1016/j.tcb.2018.12.001.
[272] X. Zhuo, H. Luo, A. Chang, D. Li, H. Zhao, and Q. Zhou, “Is overexpression of TWIST, a transcriptional factor, a prognostic biomarker of head and neck carcinoma? Evidence from fifteen studies,” *Sci. Rep.*, vol. 5, no. 1, p. 18073, Nov. 2016, doi: 10.1038/srep18073.
[273] N. Matsuo *et al.*, “Twist expression promotes migration and invasion in hepatocellular carcinoma,” *BMC Cancer*, vol. 9, no. 1, p. 240, Dec. 2009, doi: 10.1186/1471-2407-9-240.
[274] J. Qian, Y. Luo, X. Gu, W. Zhan, and X. Wang, “Twist1 Promotes Gastric Cancer Cell Proliferation through Up-Regulation of FoxM1,” *PLoS One*, vol. 8, no. 10, p. e77625, Oct. 2013, doi: 10.1371/journal.pone.0077625.
[275] J. P. Thiery, “Epithelial–mesenchymal transitions in tumour progression,” *Nat. Rev. Cancer*, vol. 2, no. 6, pp. 442–454, Jun. 2002, doi: 10.1038/nrc822.
[276] R. Fujiwara-Tani *et al.*, “Concurrent Expression of CD47 and CD44 in Colorectal Cancer Promotes Malignancy,” *Pathobiology*, vol. 86, no. 4, pp. 182–189, 2019, doi: 10.1159/000496027.
[277] H. Wang *et al.*, “Expression and significance of CD44, CD47 and c-met in ovarian clear cell carcinoma.,” *Int. J. Mol. Sci.*, vol. 16, no. 2, pp. 3391–404, Feb. 2015, doi: 10.3390/ijms16023391.
[278] D. P. Lavin and V. K. Tiwari, “Unresolved Complexity in the Gene Regulatory Network Underlying EMT,” *Front. Oncol.*, vol. 10, p. 554, May 2020, doi: 10.3389/fonc.2020.00554.
[279] N. V Sankpal, T. P. Fleming, P. K. Sharma, H. J. Wiedner, and W. E. Gillanders, “A double-negative feedback loop between EpCAM and ERK contributes to the regulation of epithelial–mesenchymal transition in cancer,” *Oncogene*, vol. 36, pp. 3706–3717, 2017, doi: 10.1038/onc.2016.504.
[280] S. Bin Lim, W. Di Lee, J. Vasudevan, W.-T. Lim, and C. T. Lim, “Liquid biopsy: one cell at a time,” *npj Precis. Oncol.*, vol. 3, no. 1, p. 23, Dec. 2019, doi: 10.1038/s41698-019-0095-0.
[281] E. Rossi and R. Zamarchi, “Single-Cell Analysis of Circulating Tumor Cells: How Far Have We

Come in the -Omics Era?," *Front. Genet.*, vol. 10, p. 958, Oct. 2019, doi: 10.3389/fgene.2019.00958.
[282] T. Brabletz *et al.*, "Variable β-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 18, p. 10356, 2001, doi: 10.1073/PNAS.171610498.
[283] Z. M, B. L, E. S, and P. H, "Bulk tumour cell migration in lung carcinomas might be more common than epithelial-mesenchymal transition and be differently regulated," *BMC Cancer*, vol. 18, no. 1, 2018, doi: 10.1186/S12885-018-4640-Y.
[284] Y. Jiang and H. Zhan, "Communication between EMT and PD-L1 signaling: New insights into tumor immune evasion.," *Cancer Lett.*, vol. 468, pp. 72–81, 2020, doi: 10.1016/j.canlet.2019.10.013.
[285] R. Negishi *et al.*, "Transcriptomic profiling of single circulating tumor cells provides insight into human metastatic gastric cancer," *Commun. Biol.*, vol. 5, no. 1, p. 20, Dec. 2022, doi: 10.1038/s42003-021-02937-x.
[286] J. Y. Lee *et al.*, "Detection of Head and Neck Cancer Based on Longitudinal Changes in Serum Protein Abundance," *Cancer Epidemiol. Biomarkers Prev.*, vol. 29, no. 8, pp. 1665–1672, Aug. 2020, doi: 10.1158/1055-9965.EPI-20-0192.
[287] K. Payne *et al.*, "Circulating tumor DNA as a biomarker and liquid biopsy in head and neck squamous cell carcinoma," *Head Neck*, vol. 40, no. 7, pp. 1598–1604, 2018, doi: 10.1002/hed.25140.
[288] B. Pereira *et al.*, "Cell-free DNA captures tumor heterogeneity and driver alterations in rapid autopsies with pre-treated metastatic cancer," *Nat. Commun.*, vol. 12, no. 1, p. 3199, Dec. 2021, doi: 10.1038/s41467-021-23394-4.
[289] F. A. W. Coumans, S. T. Ligthart, J. W. Uhr, and L. W. M. M. Terstappen, "Challenges in the enumeration and phenotyping of CTC.," *Clin. Cancer Res.*, vol. 18, no. 20, pp. 5711–8, Oct. 2012, doi: 10.1158/1078-0432.CCR-12-1585.
[290] J. Phallen *et al.*, "Direct detection of early-stage cancers using circulating tumor DNA.," *Sci. Transl. Med.*, vol. 9, no. 403, Aug. 2017, doi: 10.1126/scitranslmed.aan2415.
[291] J. Lee *et al.*, "Multimodal circulating tumor DNA (ctDNA) colorectal neoplasia detection assay for asymptomatic and early-stage colorectal cancer (CRC).," *J. Clin. Oncol.*, vol. 39, no. 15_suppl, pp. 3536–3536, May 2021, doi: 10.1200/JCO.2021.39.15_suppl.3536.
[292] H. E. Liu *et al.*, "Detection of EGFR Mutations in cfDNA and CTCs, and Comparison to Tumor Tissue in Non-Small-Cell-Lung-Cancer (NSCLC) Patients," *Front. Oncol.*, vol. 10, p. 2030, Oct. 2020, doi: 10.3389/fonc.2020.572895.
[293] "List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools) | FDA." https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools (accessed Nov. 22, 2021).
[294] "FoundationOne Liquid CDx | Foundation Medicine." https://www.foundationmedicine.com/test/foundationone-liquid-cdx (accessed Nov. 23, 2021).
[295] "The technologies | Plasma EGFR mutation tests for adults with locally advanced or metastatic non-small-cell lung cancer | Advice | NICE," Accessed: Nov. 22, 2021. [Online]. Available: https://www.nice.org.uk/advice/mib137/chapter/the-technologies#current-nhs-pathway-or-current-care-pathway.
[296] S. M. Moon *et al.*, "Clinical Utility of Combined Circulating Tumor Cell and Circulating Tumor DNA Assays for Diagnosis of Primary Lung Cancer.," *Anticancer Res.*, vol. 40, no. 6, pp. 3435–3444, Jun. 2020, doi: 10.21873/anticanres.14329.
[297] S. L. Kong *et al.*, "Complementary Sequential Circulating Tumor Cell (CTC) and Cell-Free Tumor DNA (ctDNA) Profiling Reveals Metastatic Heterogeneity and Genomic Changes in Lung Cancer and Breast Cancer," *Front. Oncol.*, vol. 11, p. 2781, Jul. 2021, doi: 10.3389/fonc.2021.698551.

[298] E. Ortolan *et al.*, “Blood-based genomics of triple-negative breast cancer progression in patients treated with neoadjuvant chemotherapy 5,” 2021, doi: 10.1016/j.esmoop.2021.100086.

APPENDIX

The following appendix figures detail antibody optimisation experiments, mass cytometry panel iterations and antibody panel design/titration experiments:

TABLE OF CONTENTS

Figure A1.1: Flow cytometry antibody optimisation

Figure A1.2: Mass cytometry panel iterations

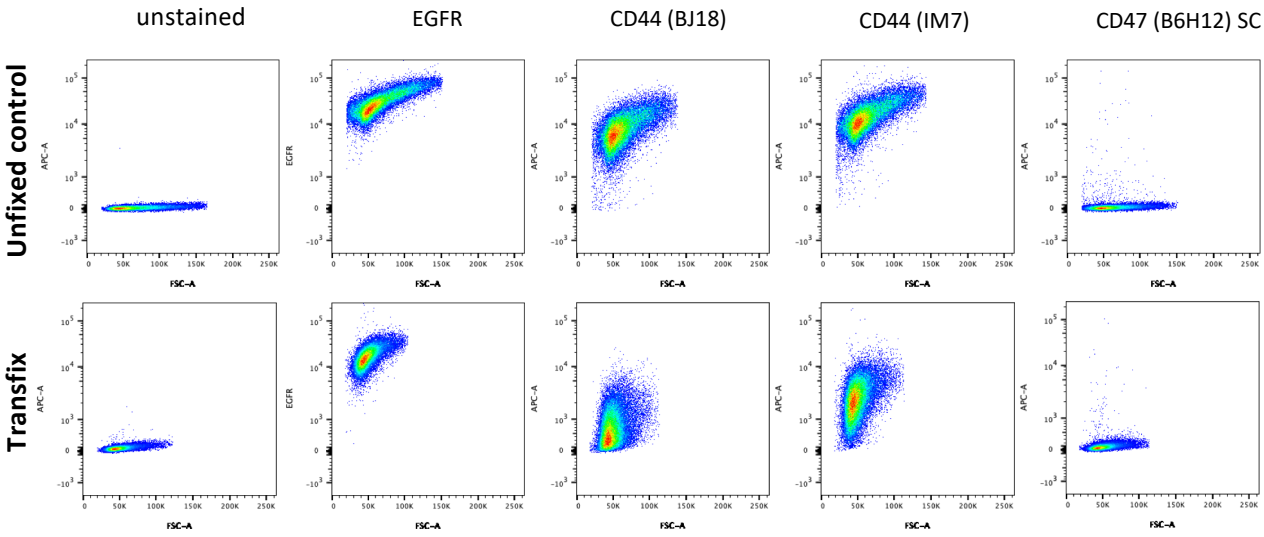
Figure A1.3: Mass cytometry marker optimization

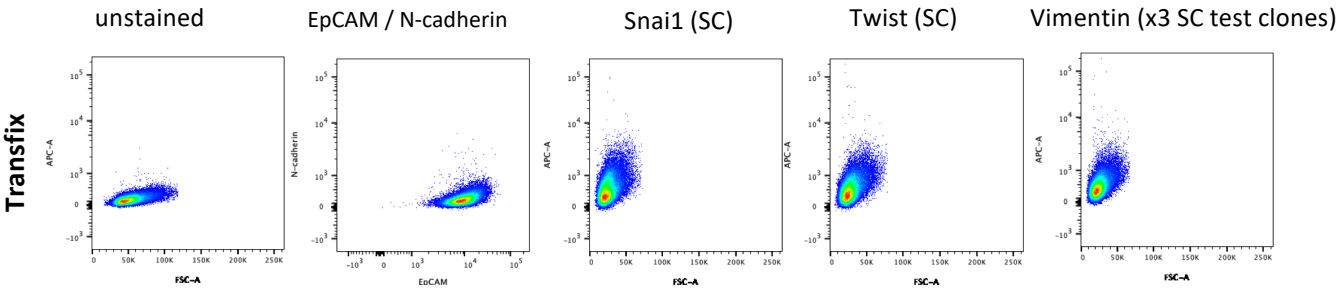
Figure A1.4: Extracellular and intracellular marker titration

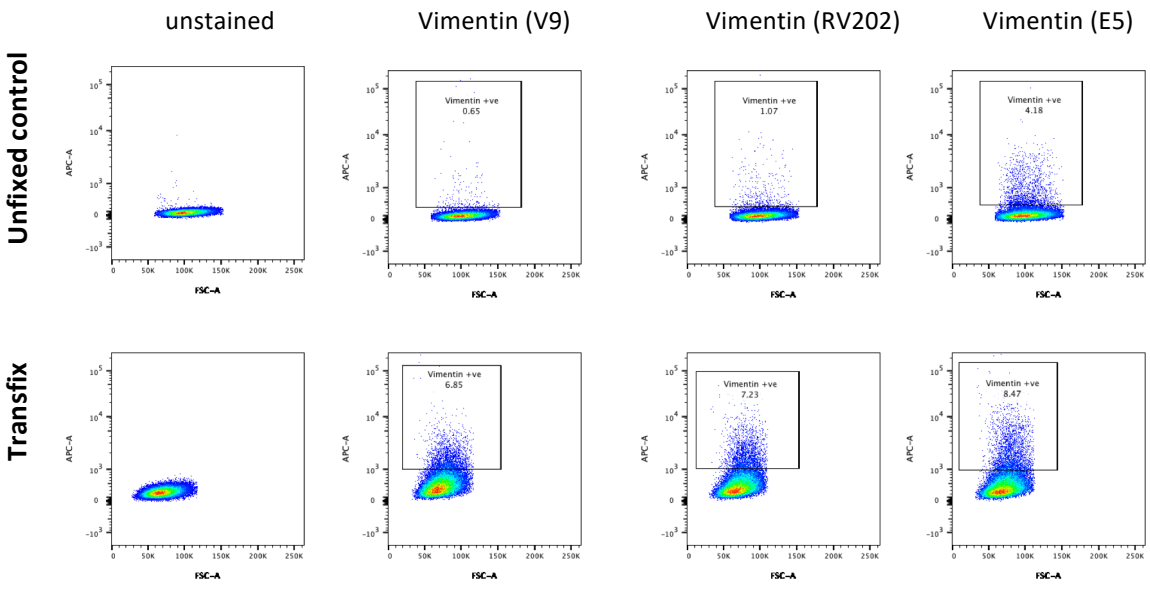
Figure A1.1: Flow cytometry antibody optimisation

Biaxial flow cytometry plots are displayed for multiple optimisation experiments testing antibody clones on Transfix fixed cells. Antibody signal is plotted on y-axis, assessed using fluorophore log scale intensity, with FSC as a common comparator on the x-axis. In each experiment, unfixed and fixed (Transfix) samples are compared. Marker signal intensity is compared to a negative (unstained) control. HNSCC cell lines and healthy donor PBMCs were used as positive controls.

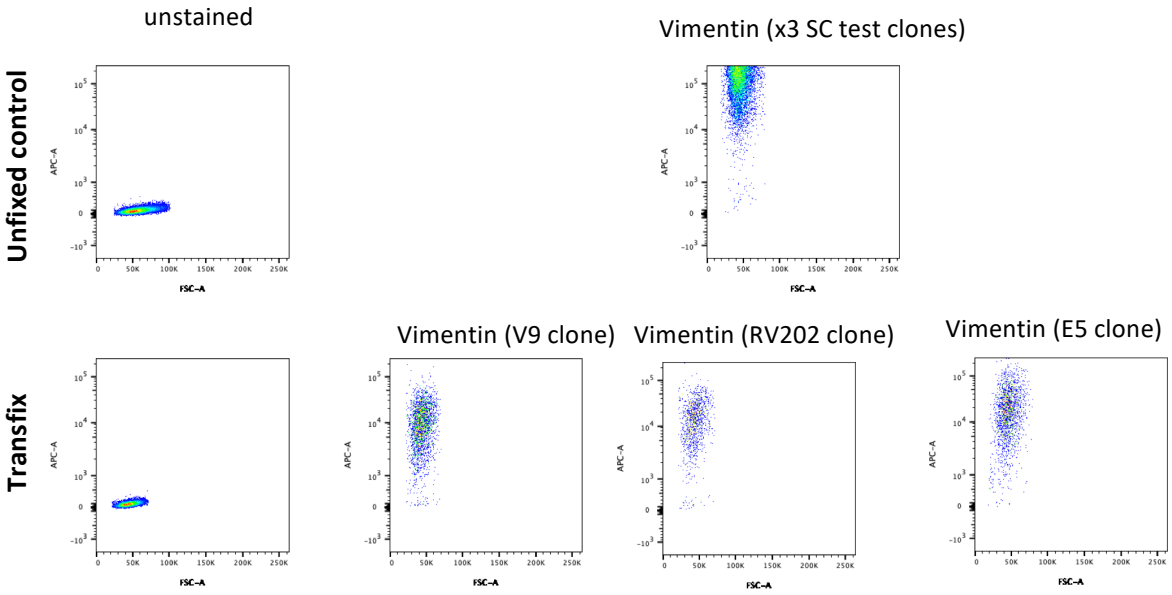
FaDu control cell line



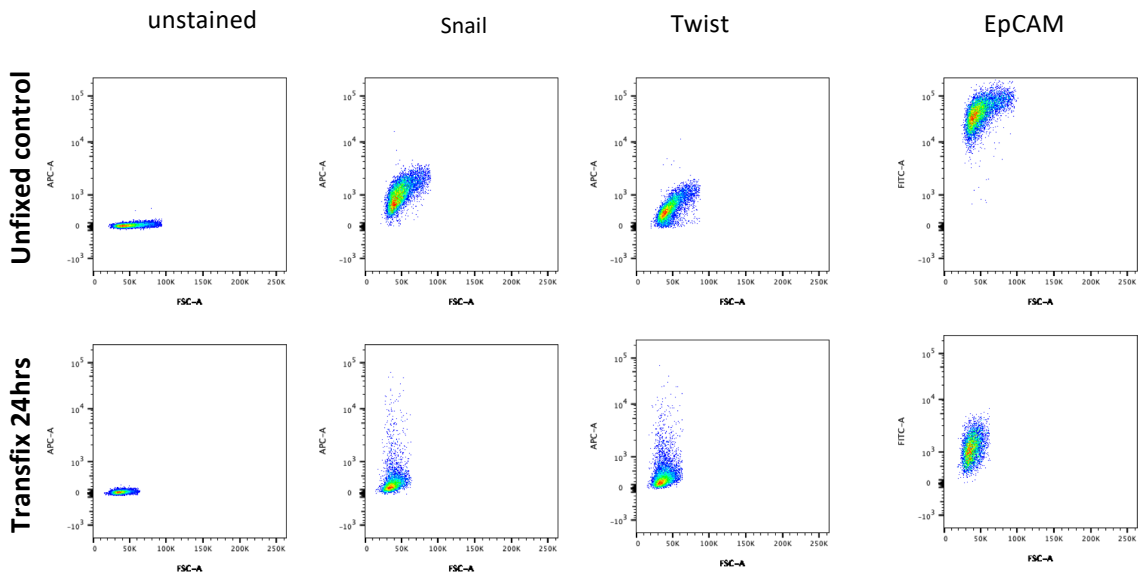




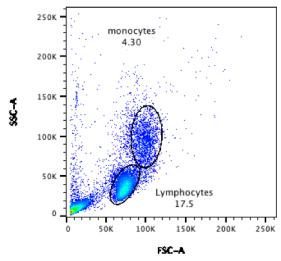
Fibroblast control cell line



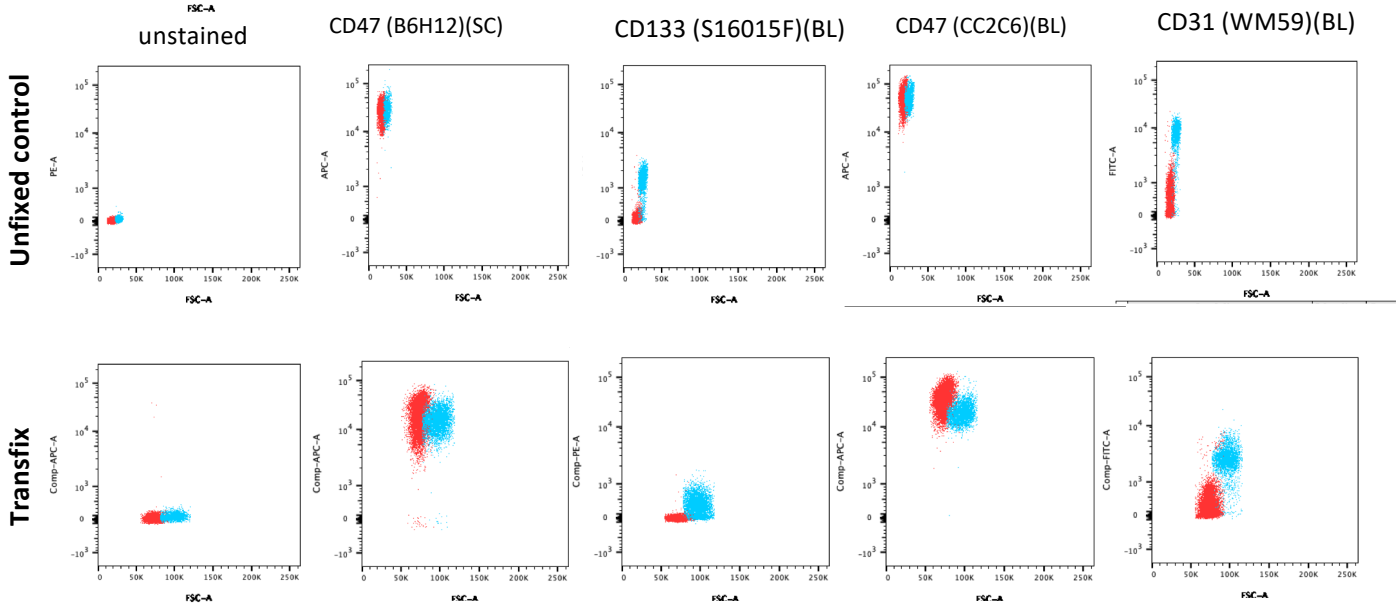
CAL27 control cell line

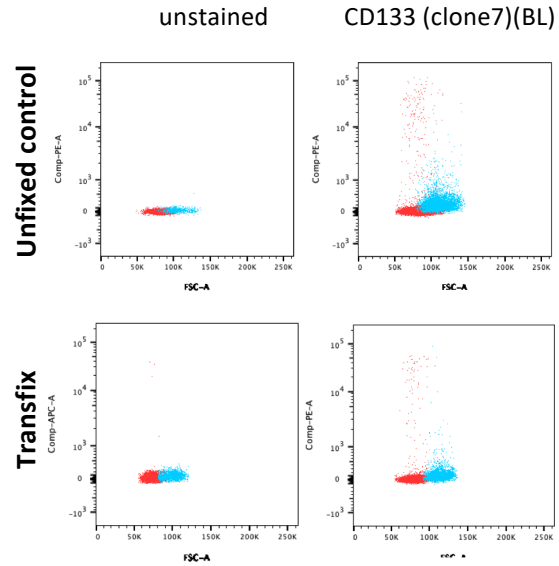


Healthy donor PBMC control

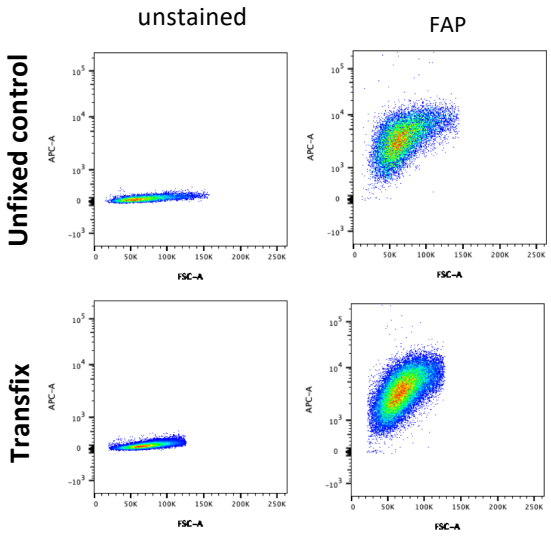


Lymophocyte and monocyte populations
gated on FSC vs SSC plots





Fibroblast control cell



Fibroblast and FaDu control cell mix

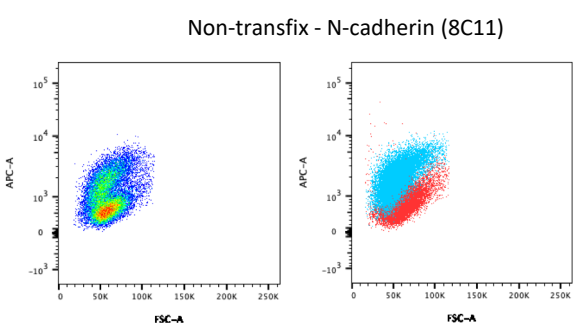


Figure A1.2: Mass cytometry panel iterations

Multiple test mass cytometry antibody panels are displayed below, as utilised in figure A1.3 and A1.4. An iterative process of panel design was utilised, assessing both metal signal intensity, spillover and metal oxidation. Initially, immediately available metal-tagged antibodies were used (from collaborators) to build a panel. Panel tests were subsequently conducted to refine position of markers.

Mass cytometry – Panel iterations

Test 1

Metal	Marker
110	Cd CD14 (NK)
114	Qdot CD45 (GT)
115	(EPCAM from NPC)
141	CD103
142	-------
143	CD38
144	CD8 (
145	CD4 (fluidigm)
146	-------
147	CXCR3 (OLIVIA)
148	CD16 (FLUIDIGM CLL)
149	CD127 (FLUIDIGM)
150	pSTAT5 (FLUIDIGM)
151	---------
152	pAKT
153	pSTAT1
154	TIGIT (CARRIE W CLONE)
155	PD1
156	p38
158	pSTAT5
159	-------
160	CD39
161	CTLA4
162	CD27 (FLUIDIGM)
163	CD56 (Fluidigm))
164	CD95
165	CD19
166	NFKB p65
167	CCR7 (GT)
168	CXCR5 (clone J25D4)
169	CD25 (Fluidigm)
170	LOX1
171	pERK1/2
172	Ki67
173	e-Cadherin (NPC)
174	HLA-DR (Fluidigm)
175	pS6
176	pCREB
194	HLA-ABC
198	--------
209	CD11b

Test 2

Metal	Marker
198	Pan-keratin
173	CD3
161	CD163
163	CD56
172 (158)	PD-L1
175	PD-L2
89	CD45
194	HLA-ABC
174	HLA-DR
160	CD39
154	CD73
144	CD14
159	CD86
115	Epcam
153	pSTAT1
158	pSTAT 3
150	pSTAT5
156	p38
175	pS6
176	pCREB
152	pAKT
172	Ki-67
166	CD19
143	PARP

Test 3

Metal	Marker
89	CD45
114	CD8
115	Epcam
142	
143	PARP
144	CD8
145	
149	CD25
150	pSTAT5
152	pAKT
153	pSTAT1
154	CD73
155	PD1
156	p38
158	pSTAT 3
159	CD86
160	CD39
161	CD163
163	CD56
165	CD19
166	CD24
170	
171	pERK
172	PD-L1
	e-cadherin and
173	CD3
174	HLA-DR
175	pS6
176	pCREB
194	HLA-ABC
198	Pan-keratin
209	CD11b

Test 4

Metal	marker
089Y	CD45
Qdot 114	
Indium 115	
141Pr	EpCAM
142Nd	
143Nd	PARP
144Nd	CD31
145Nd	
146Nd	
147Sm	
148Sm	
149Sm	CD25
150Nd	pSTAT5
151Eu	
152Sm	pAKT
153Eu	pSTAT1
154Sm	CD73
155Gd	PD1
156Gd	p38
158Eu	pSTAT 3
159Tb	CD86
160Gd	CD39
161Dy	CTLA4
162Dy	CD44
163Dy	CD56
164Dy	
165Ho	PD-L1
166Er	CD24
167Er	
168Er	Ki-67
169Tm	
170Er	EGFR
171Yb	pERK
172Yb	
173Yb	E-cadherin
174Yb	HLA-DR
175Lu	pS6
176Yb	pCREB
194	HLA-ABC
198P	pan-keratin
209Bi	CD47

Final Panel (table 7.3)

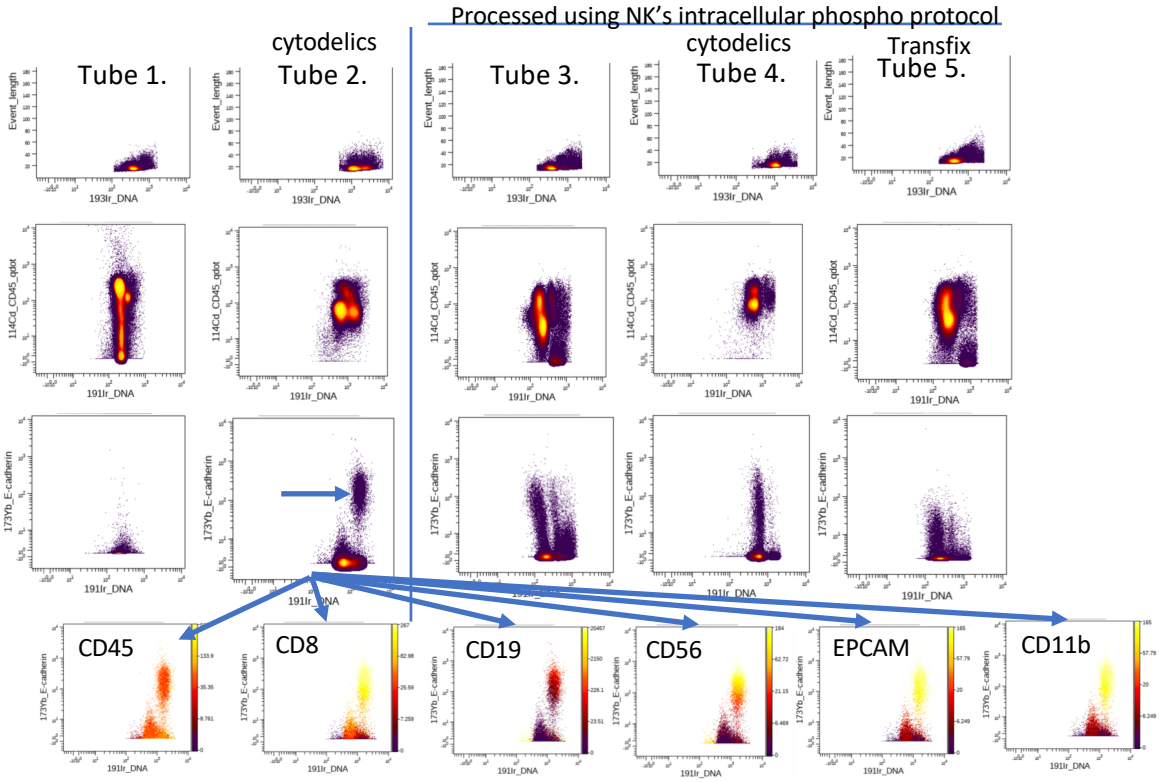
Metal	marker
089Y	CD45
110	CD14
113	CD19
114	CD44
115	CD3
116	CD4
139	CD27
141Pr	EpCAM
142Nd	CD73
143Nd	PARP
144Nd	CD31
145Nd	Vimentin
146Nd	CD86
147Sm	Twist
148Sm	PD-L2
149Sm	CD25
150Nd	pSTAT5
151Eu	
152Sm	pAKT
153Eu	pSTAT1
154Sm	E-cadherin
155Gd	PD1
156Gd	p38
158Eu	pSTAT 3
159Tb	CD133
160Gd	CD39
161Dy	CTLA4
162Dy	CD44
163Dy	CD56
164Dy	pan-keratin
165Ho	PD-L1
166Er	CD24
167Er	FAP
168Er	Ki-67
169Tm	N-cadherin
170Er	EGFR
171Yb	pERK
172Yb	CD163
173Yb	
174Yb	Snail 1
175Lu	pS6
176Yb	pCREB
194	
195	CD8
196	HLA-DR
198	HLA-ABC
209Bi	CD47

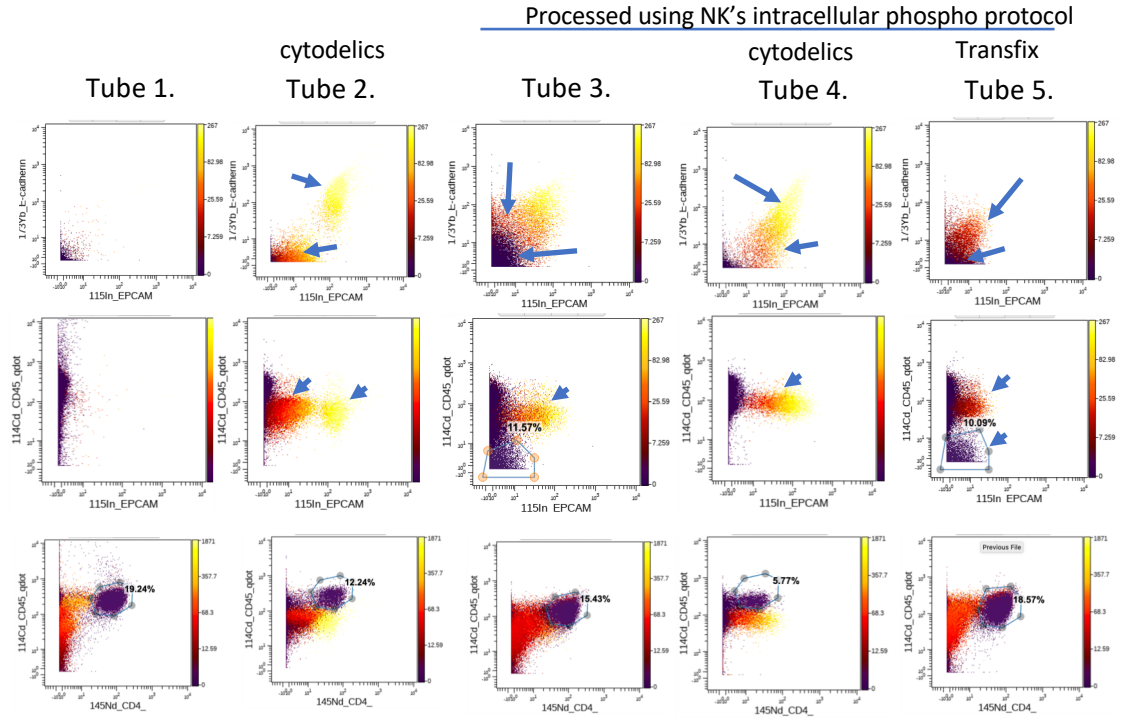
Figure A1.3: Mass cytometry marker optimisation

Bi-axial plots from multiple mass cytometry test experiments. In each experiment a cell mix of healthy donor PBMCs and positive control HNSCC cell line (FaDu) was used. The primary aim of these experiments was to test antibody clones on fixed cells that were not available for flow cytometry. In addition, we tested different protocols of sample storage (frozen vs unfrozen) and also tested the Cytodelics whole blood preservation protocol for sample fixation, all compared to our standard Transfix fixation protocol. As per previous data, freezing cells had no impact upon staining. Cytodelics impacted intracellular staining, which re-enforced our decision to adopt Transfix blood collection tubes (NK = Dr Naeem Khan).

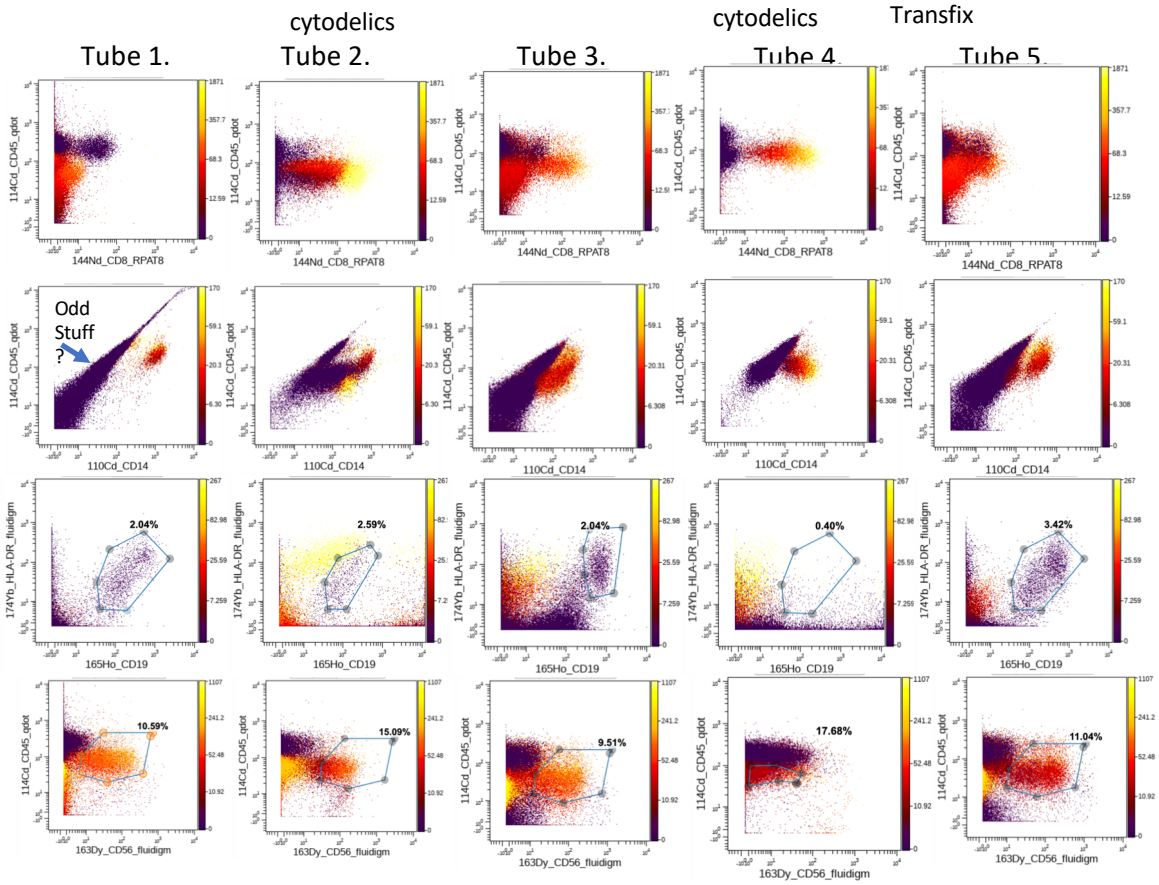
Testing of cytodelics, transfix and phospho-flow staining. 5 tubes:

Tube 1. whole blood, unprocessed.
Tube 2. whole blood + epithelial cells, cytodelics stabilized, frozen ,recovered.
Tube 3. whole blood + epithelial cells, NK's phospho-flow protocol
Tube 4. as 3 but cells preserved in cytodelics before phosphoflow protocol
Tube 5. as 3 but cells stored in transfix before phosphoflow

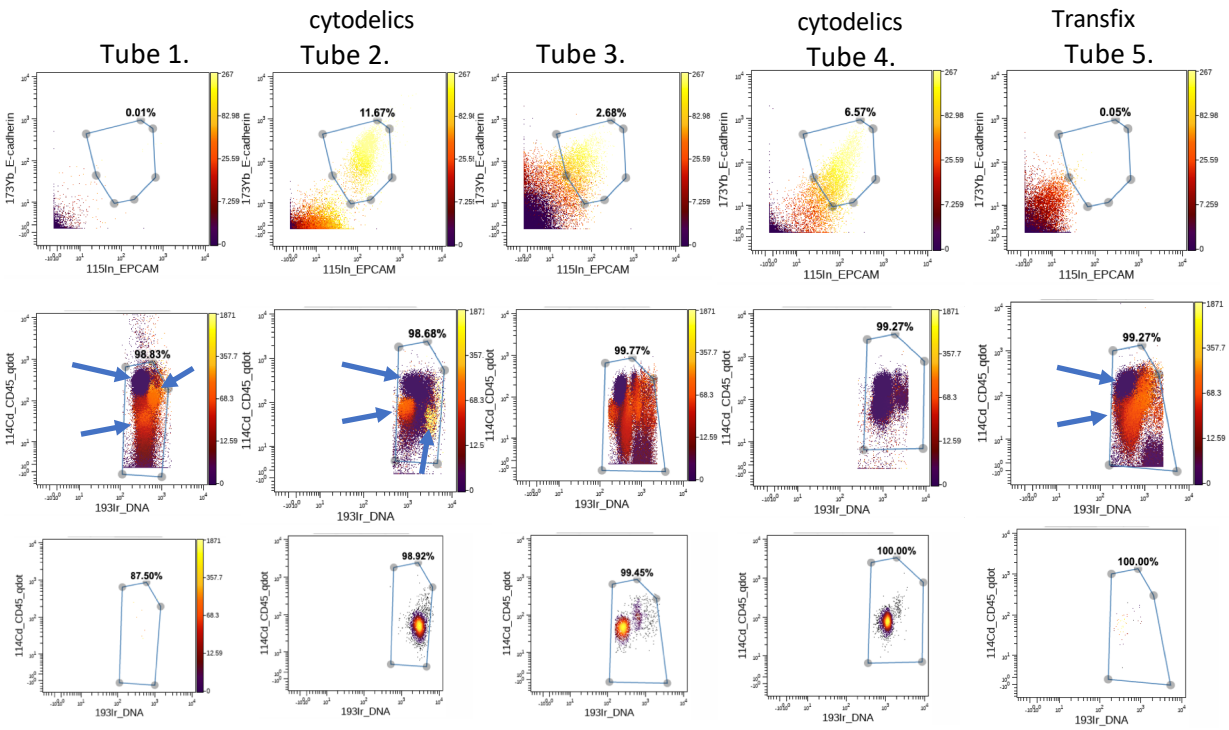




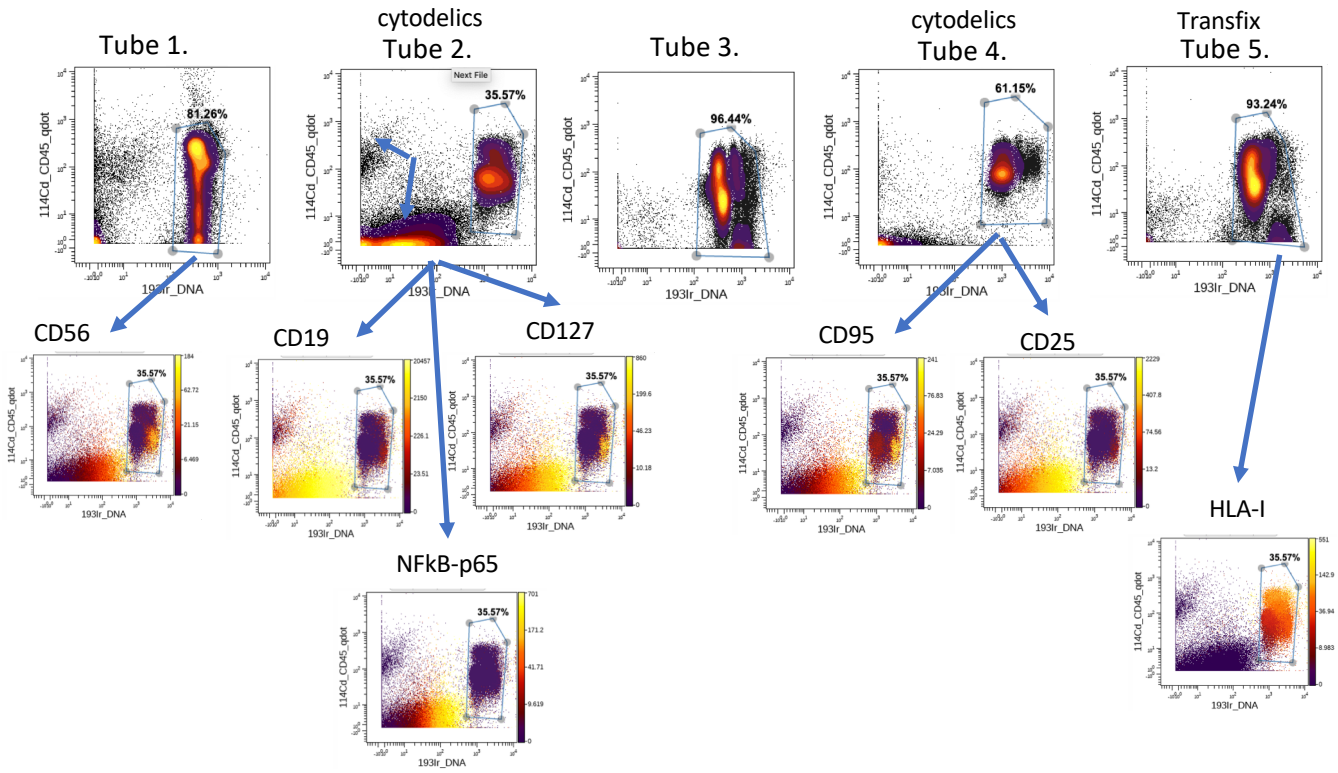
Processed using NK's intracellular phospho protocol



Processed using NK's intracellular phosph protocol

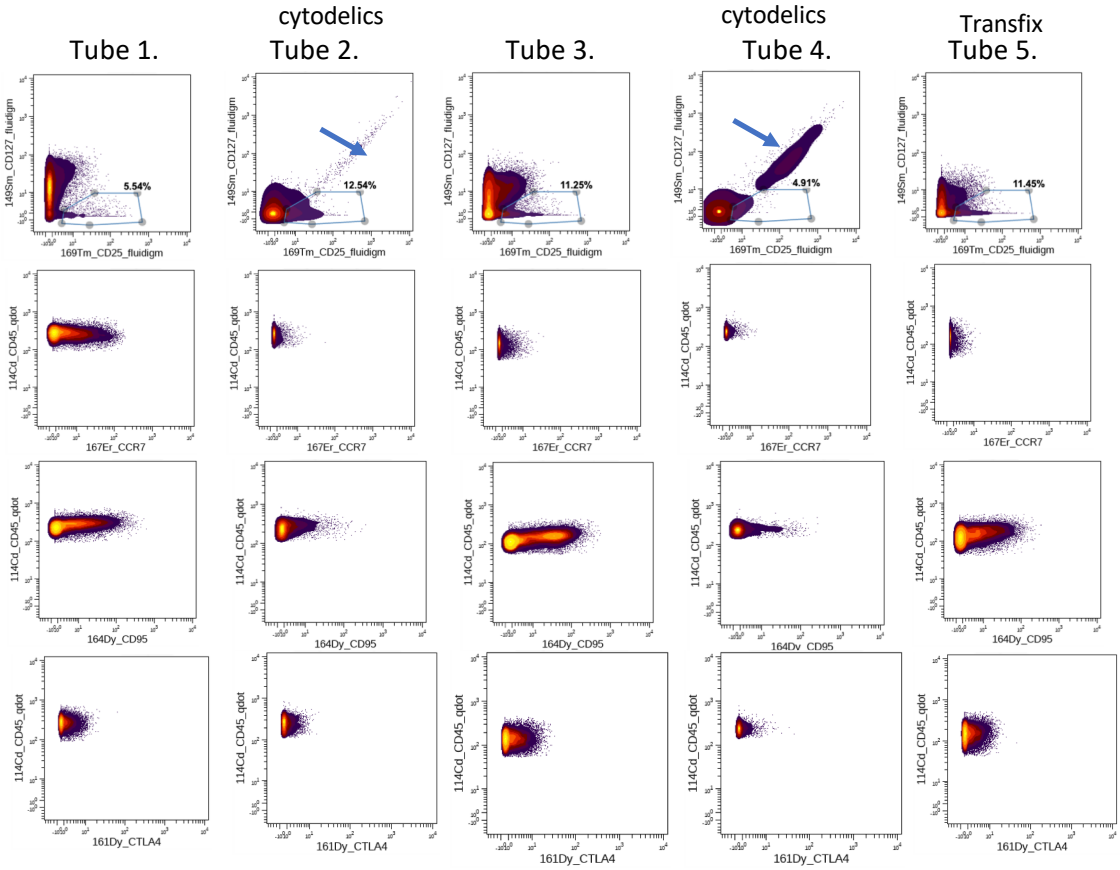


Processed using NK's intracellular phospho protocol



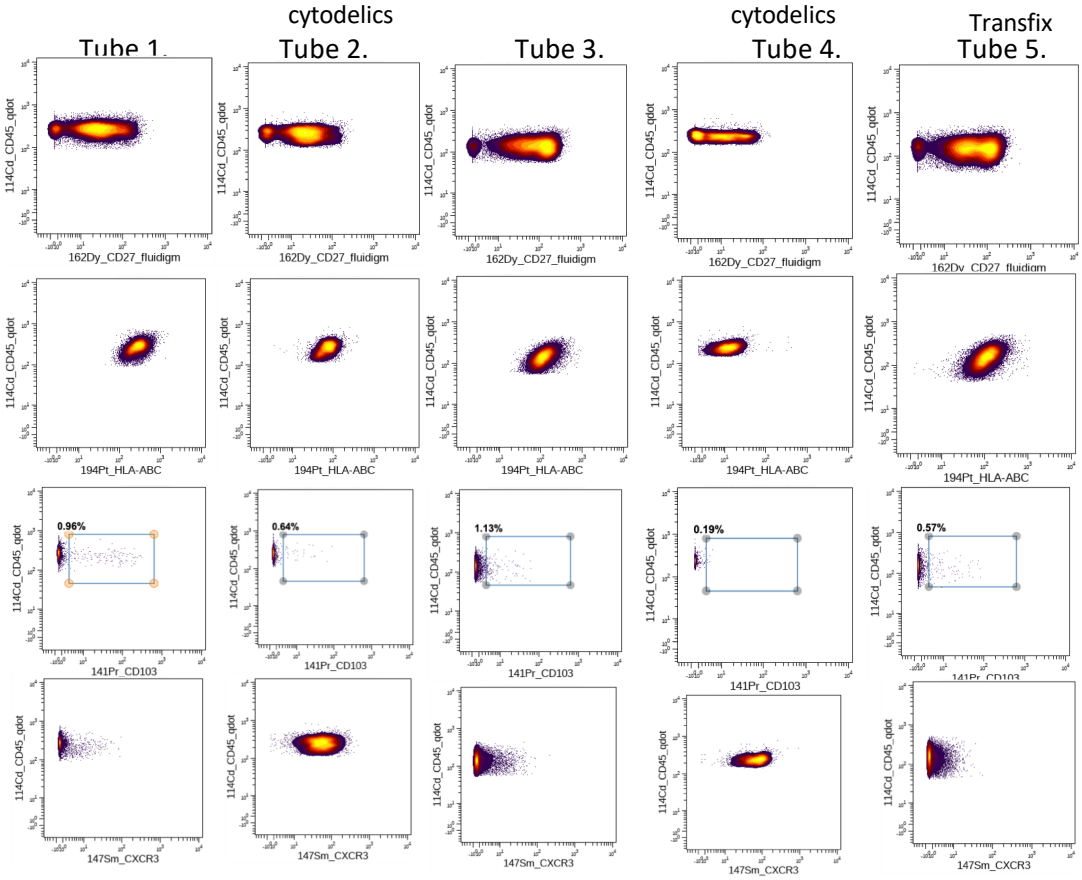
Gating=**CD4 T-cells**

Processed using NK's intracellular phosph protocol

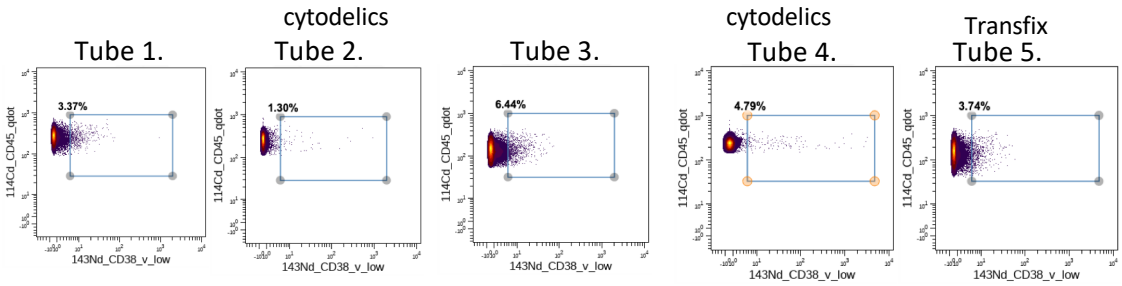


Gating=**CD4 T-cells**

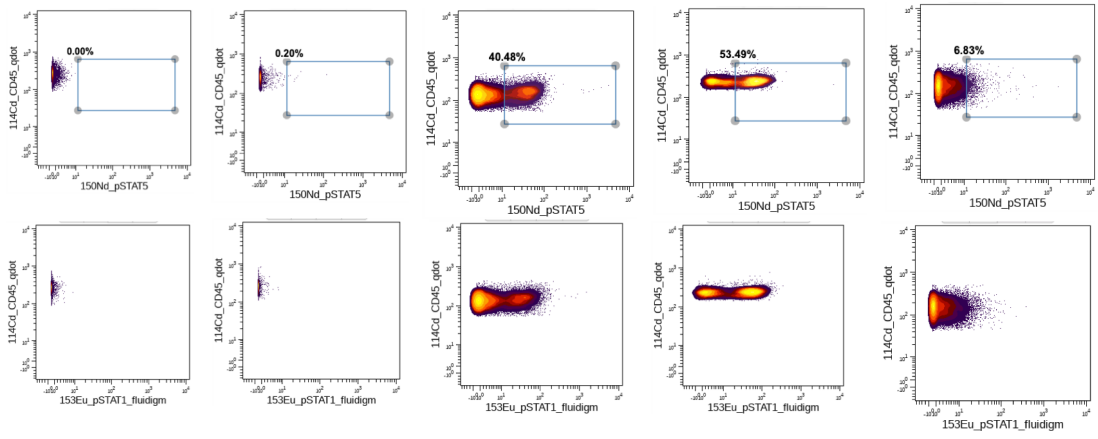
Processed using NK's intracellular phosph protocol



Processed using NK's intracellular phosph protocol

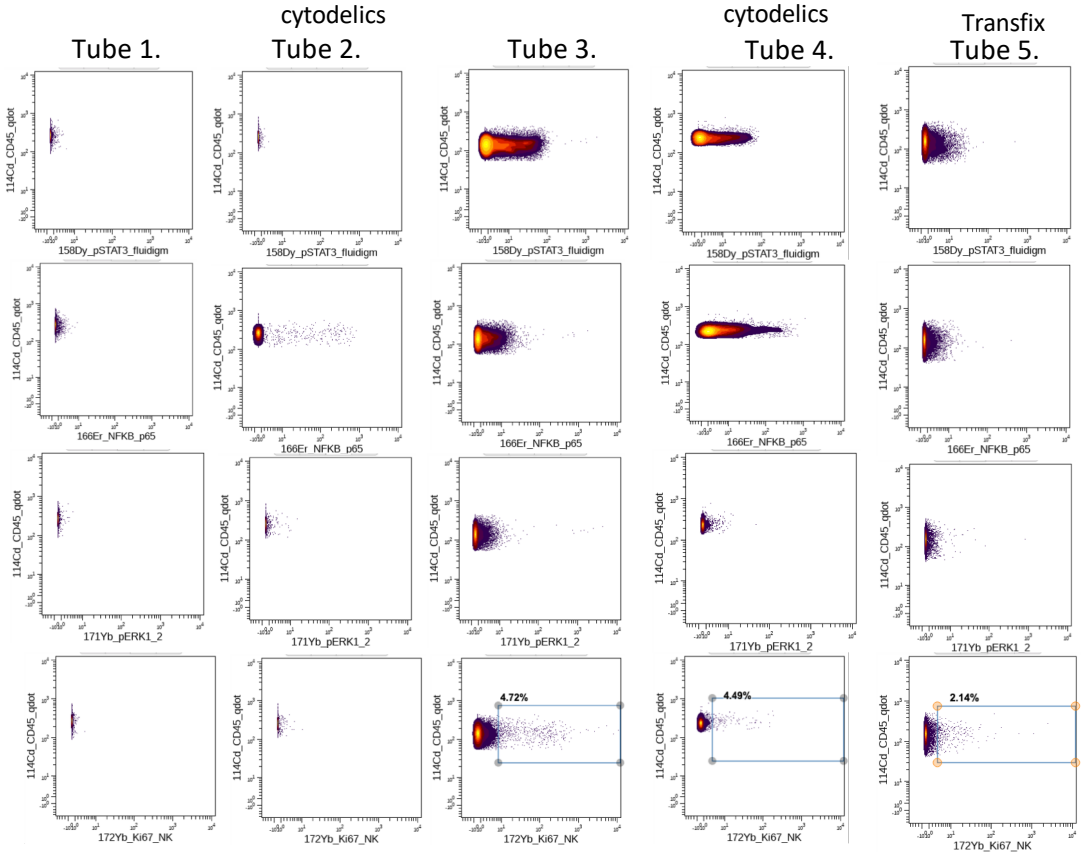


Test the phospho-specific antibodies

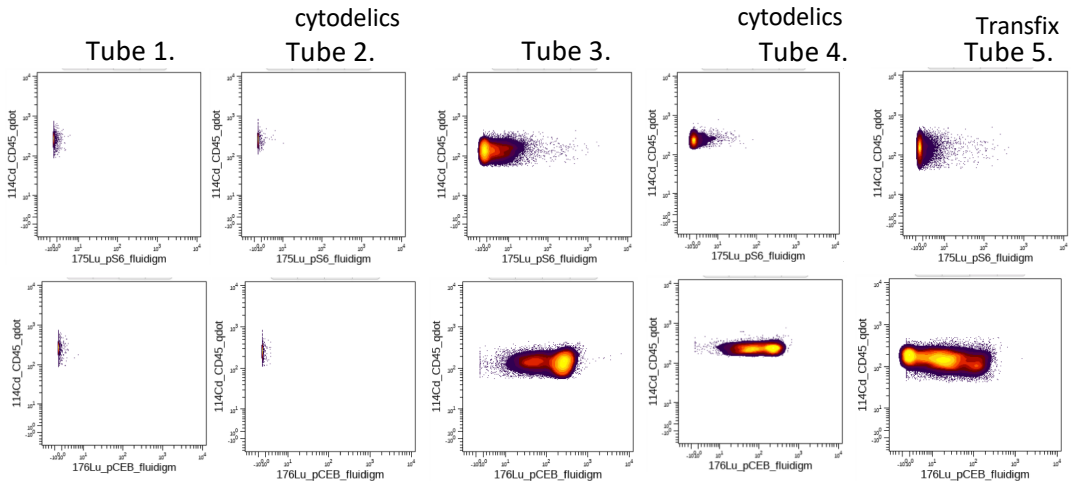


Gating=**CD4 T-cells**

Processed using NK's intracellular phosph protocol

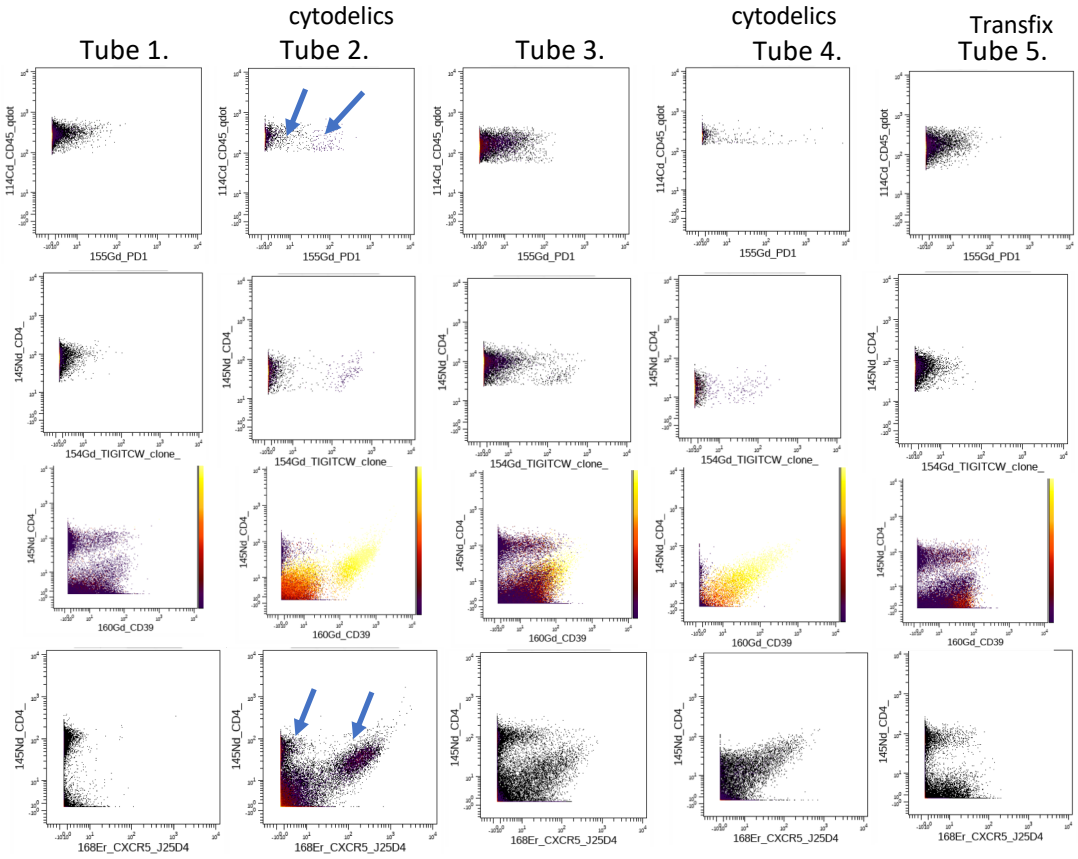


Processed using NK's intracellular phosph protocol

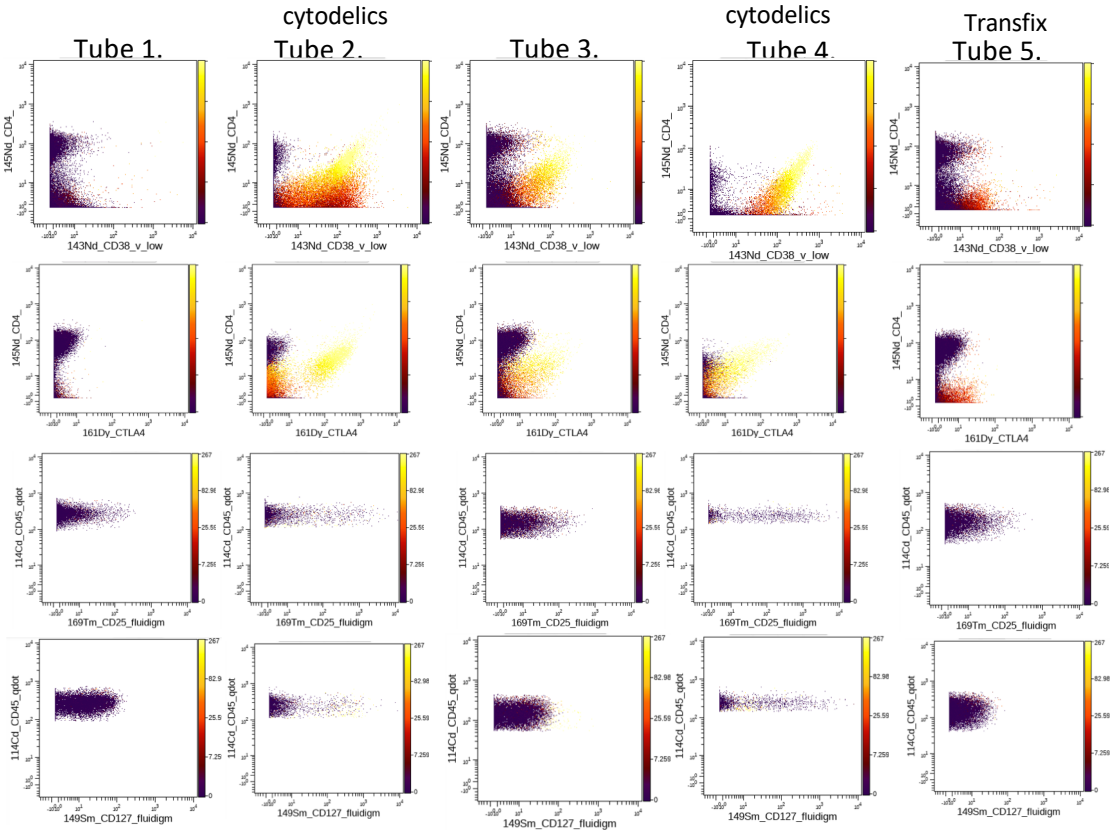


Gating=**CD4 T-cells**

Processed using NK's intracellular phosph protocol

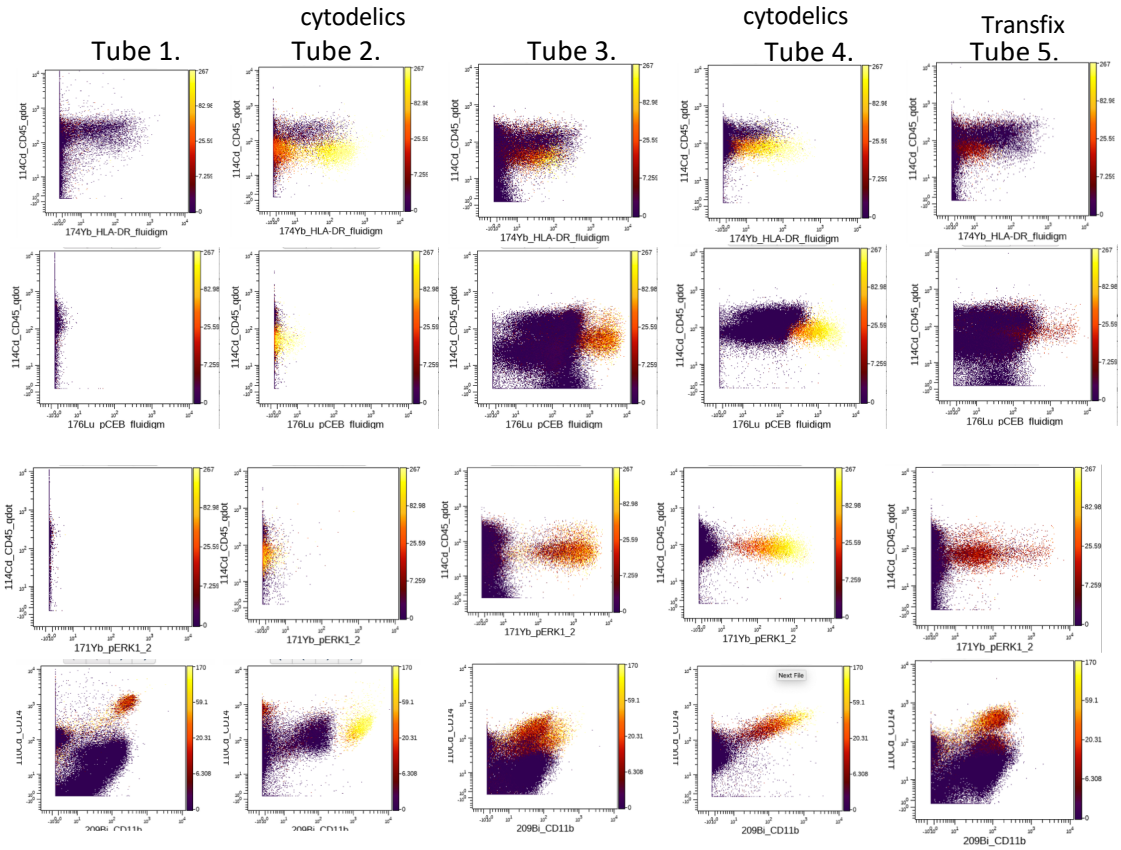


Processed using NK's intracellular phosph protocol

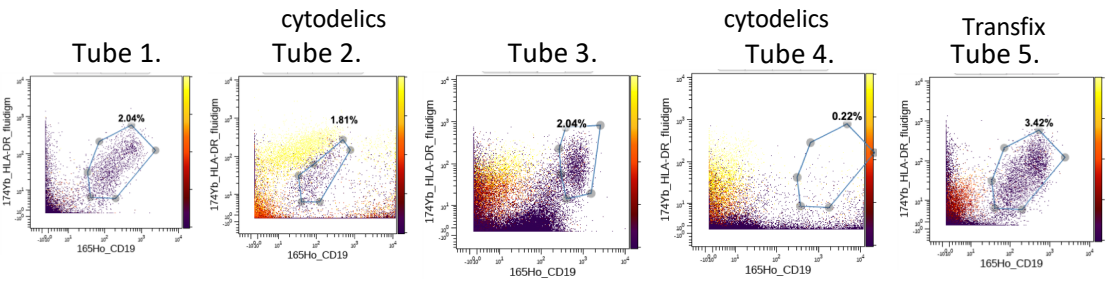


Gating=B-cells

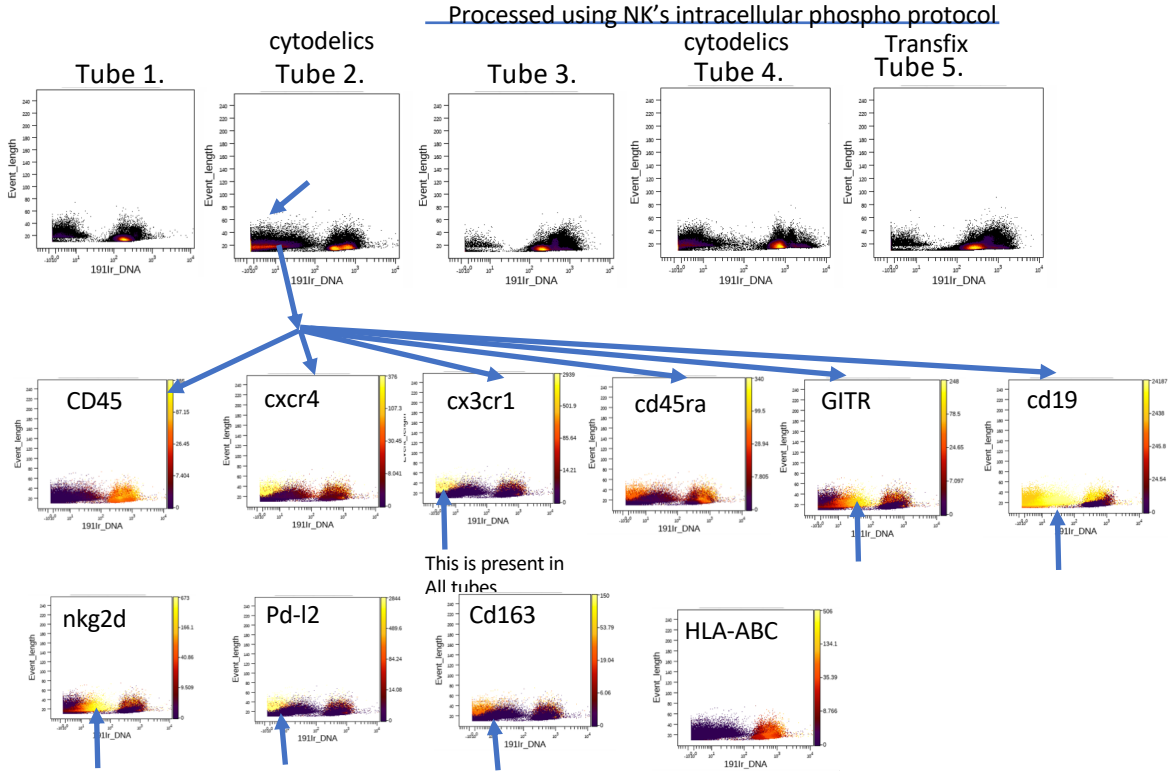
Processed using NK's intracellular phosph protocol

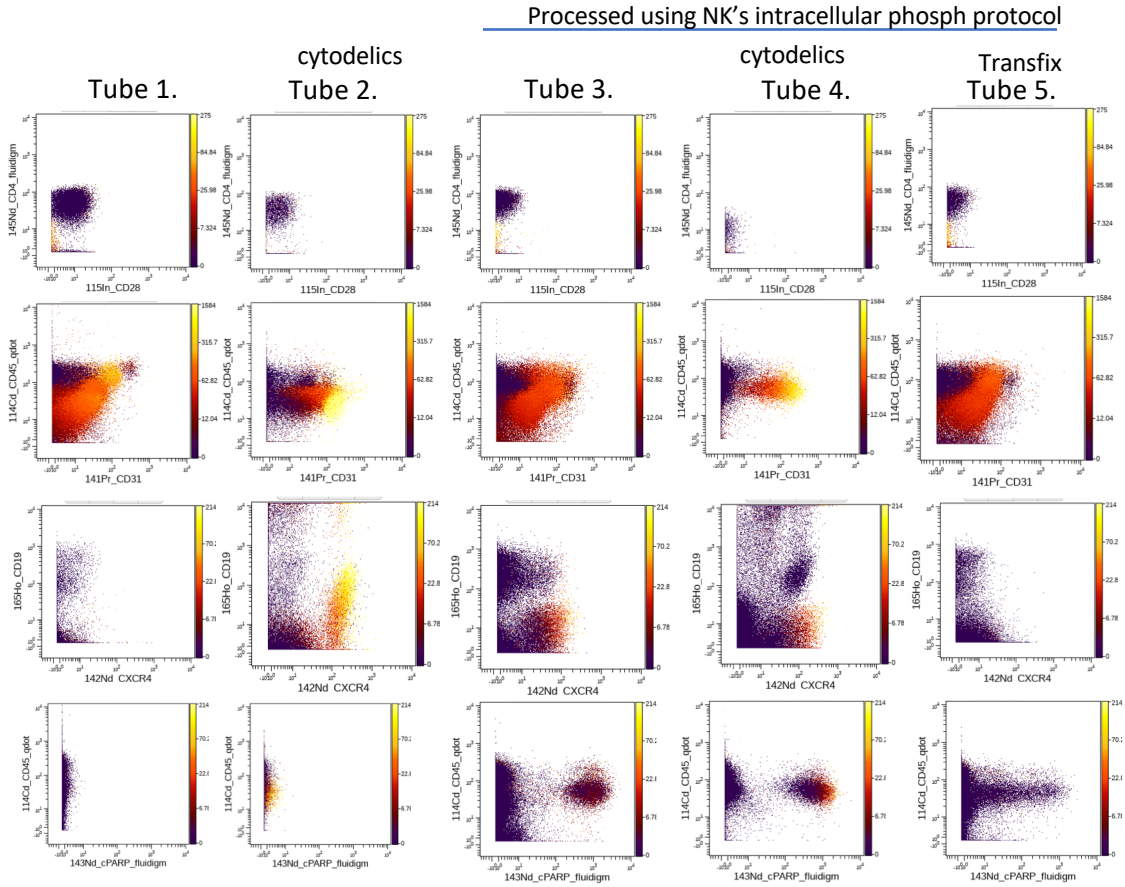


Processed using NK's intracellular phosph protocol



GATING ON ALL EVENTS (excluding beads)





Processed using NK's intracellular phosph protocol

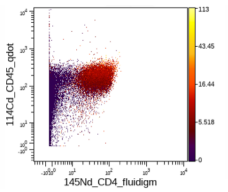
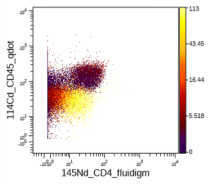
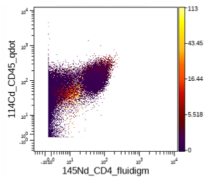
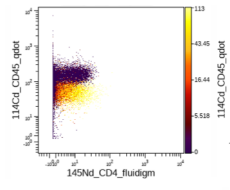
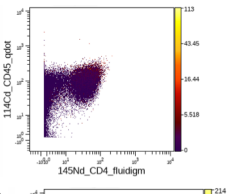
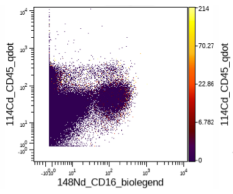
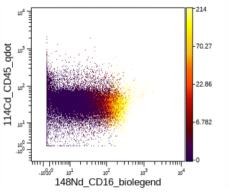
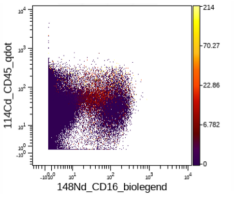
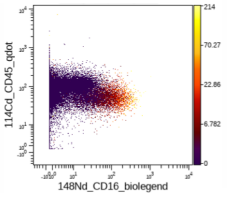
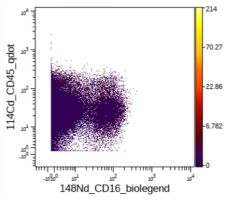
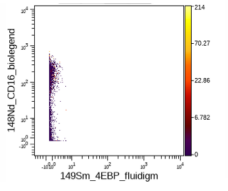
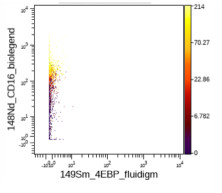
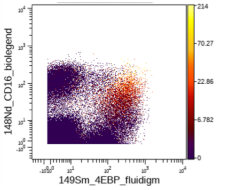
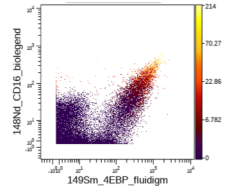
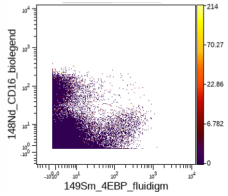
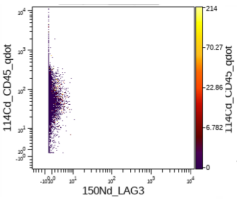
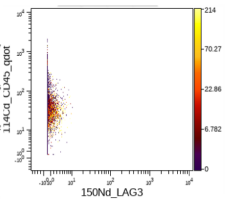
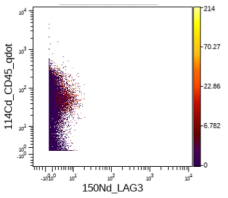
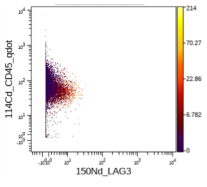
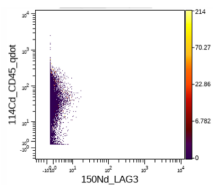
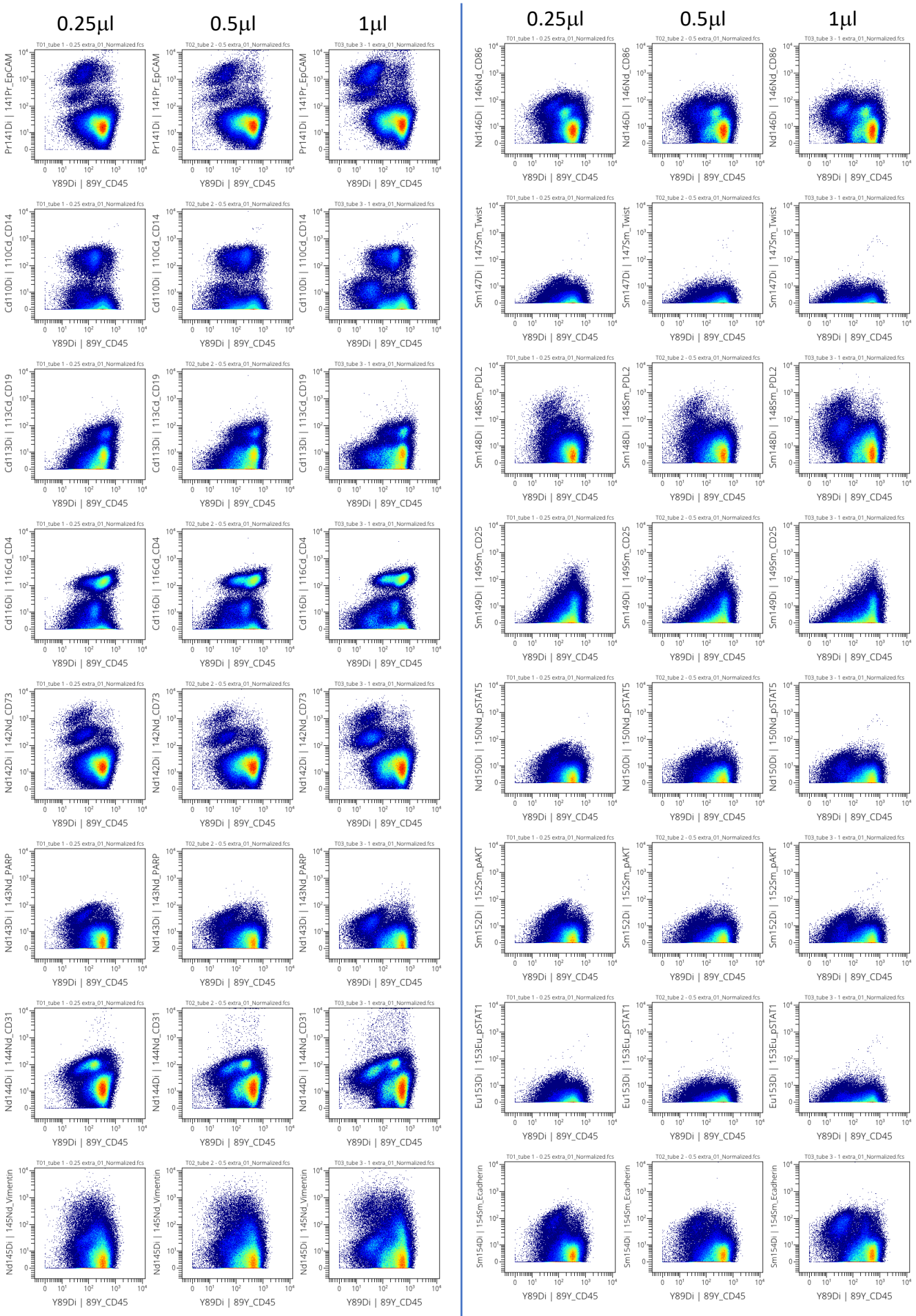
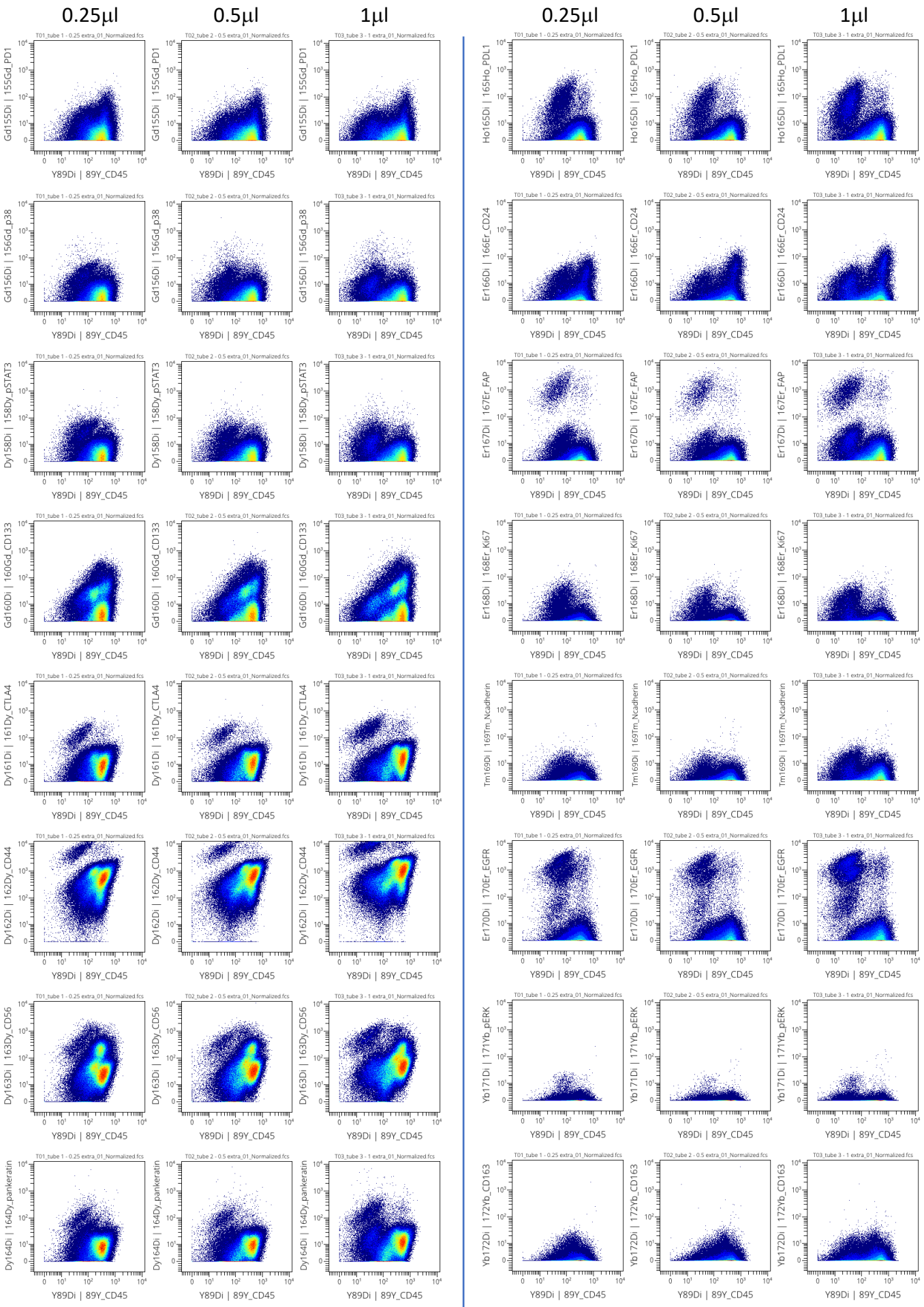
Tube 1.	cytodelics Tube 2.	Tube 3.	cytodelics Tube 4.	Transfix Tube 5.
				
				
				
				

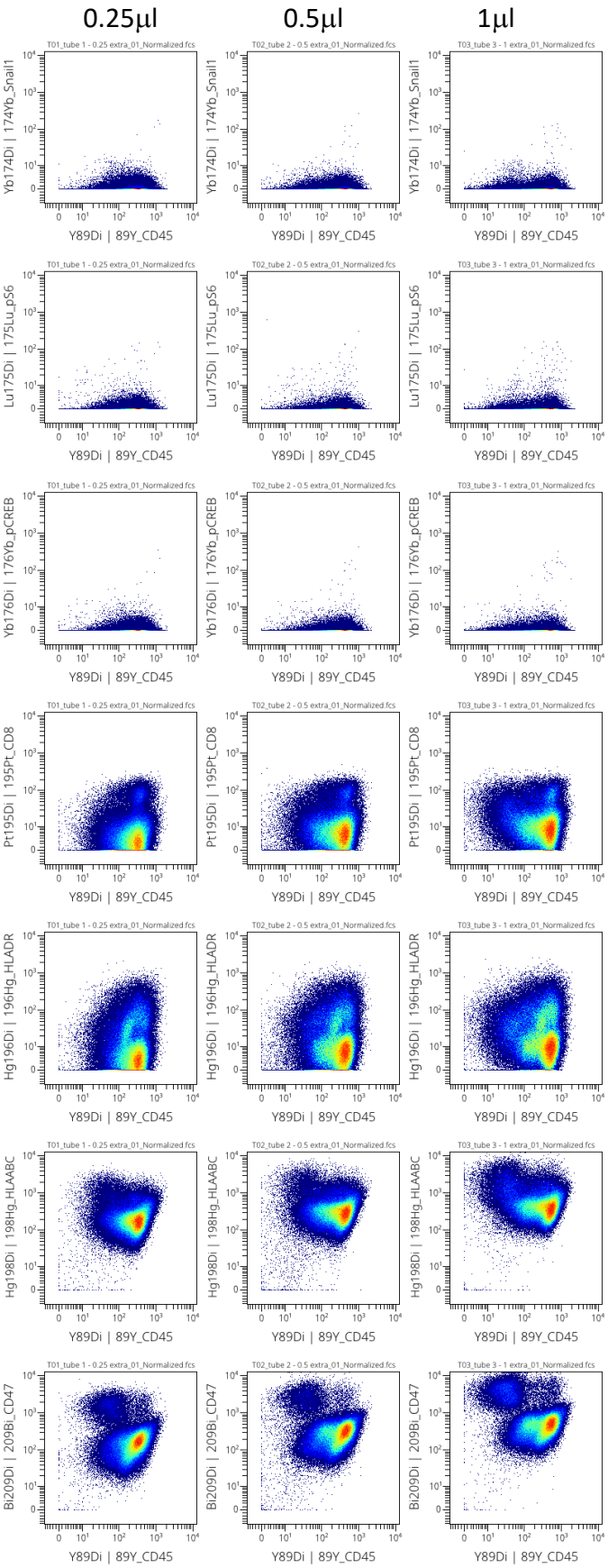
Figure A1.4: Extracellular and intracellular marker titration

Biaxial plots presenting mass cytometry experiments to confirm spillover and antibody titration of the final panel design. As before, a control cell mix of HNSCC cell line (FaDu) and healthy donor PBMCs was used. Sequential panel markers are presented on the y-axis, plotted against a common x-axis of marker CD45. For extracellular markers, three titrations of concentration were tested (0.25, 0.5 and 1μl) and two concentrations for intracellular markers (0.5 and 1μl). To assess intracellular phospho-antibody staining, stimulated versus unstimulated cells were compared.

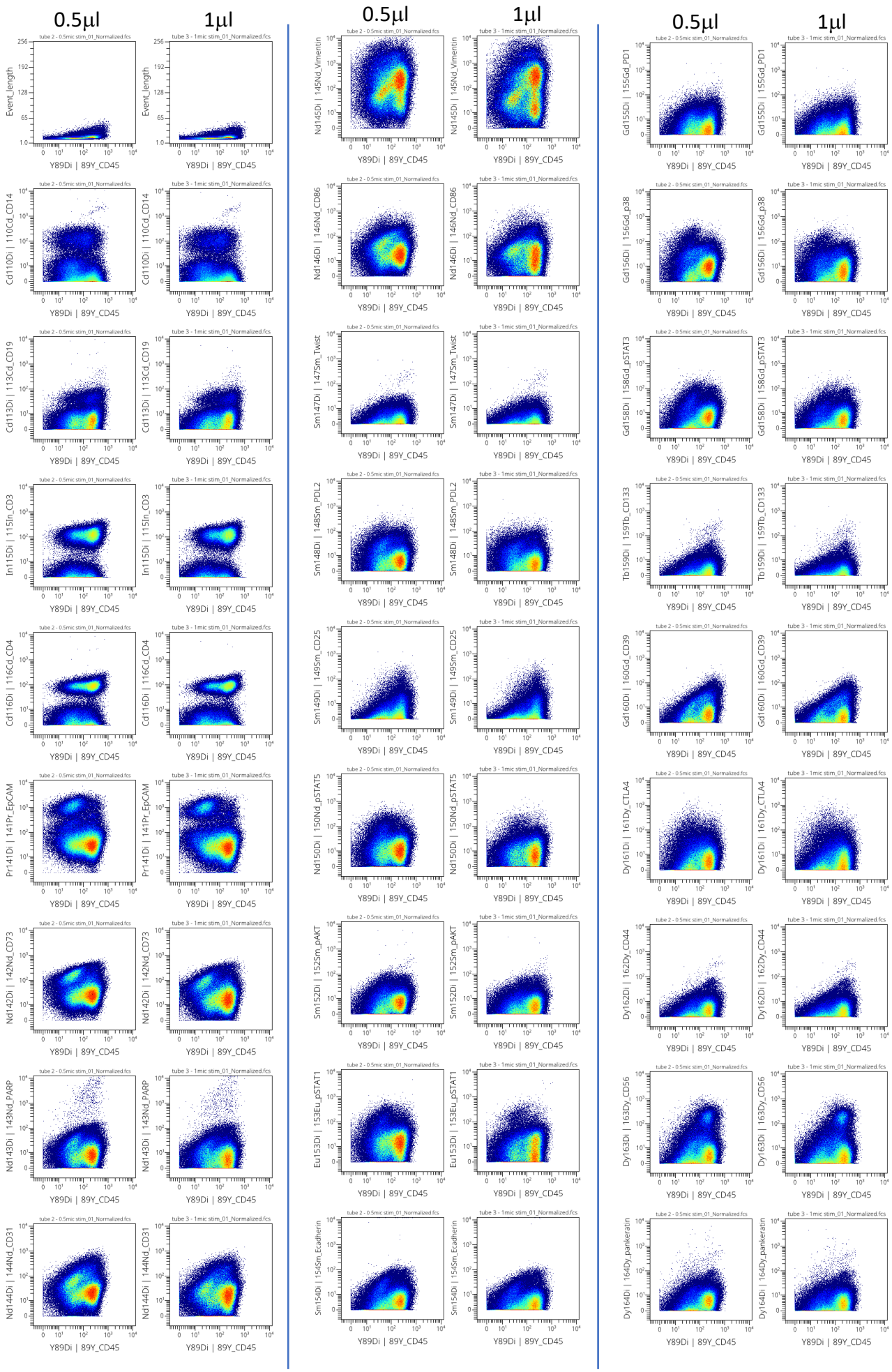
Extracellular marker titration (unstimulated cells)

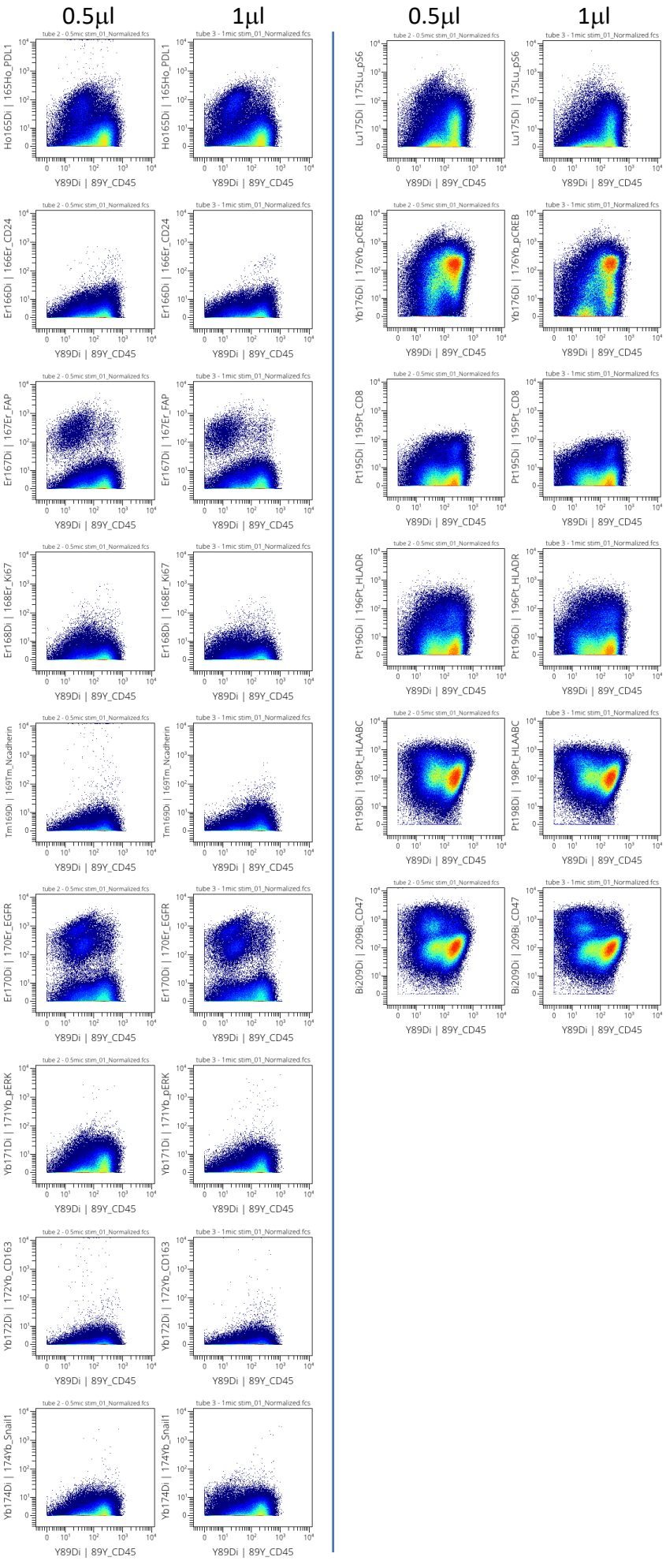



0.25µl
0.5µl
1µl
0.25µl
0.5µl
1µl
T01_tube 1 - 0.25 extra_01_Normalized.fcs
T02_tube 2 - 0.5 extra_01_Normalized.fcs
T03_tube 3 - 1 extra_01_Normalized.fcs
Gd155Di | 155Gd_PD1
Ho165Di | 165Ho_PDL1
Gd156Di | 156Gd_p38
Er166Di | 166Er_CD24
Dy158Di | 158Dy_pSTAT3
Er167Di | 167Er_FAP
Gd160Di | 160Gd_CD133
Er168Di | 168Er_Ki67
Dy161Di | 161Dy_CTLA4
Tm169Di | 169Tm_Ncadherin
Dy162Di | 162Dy_CD44
Er170Di | 170Er_EGFR
Dy163Di | 163Dy_CD56
Yb171Di | 171Yb_pERK
Dy164Di | 164Dy_pankeratin
Yb172Di | 172Yb_CD163
Y89Di | 89Y_CD45

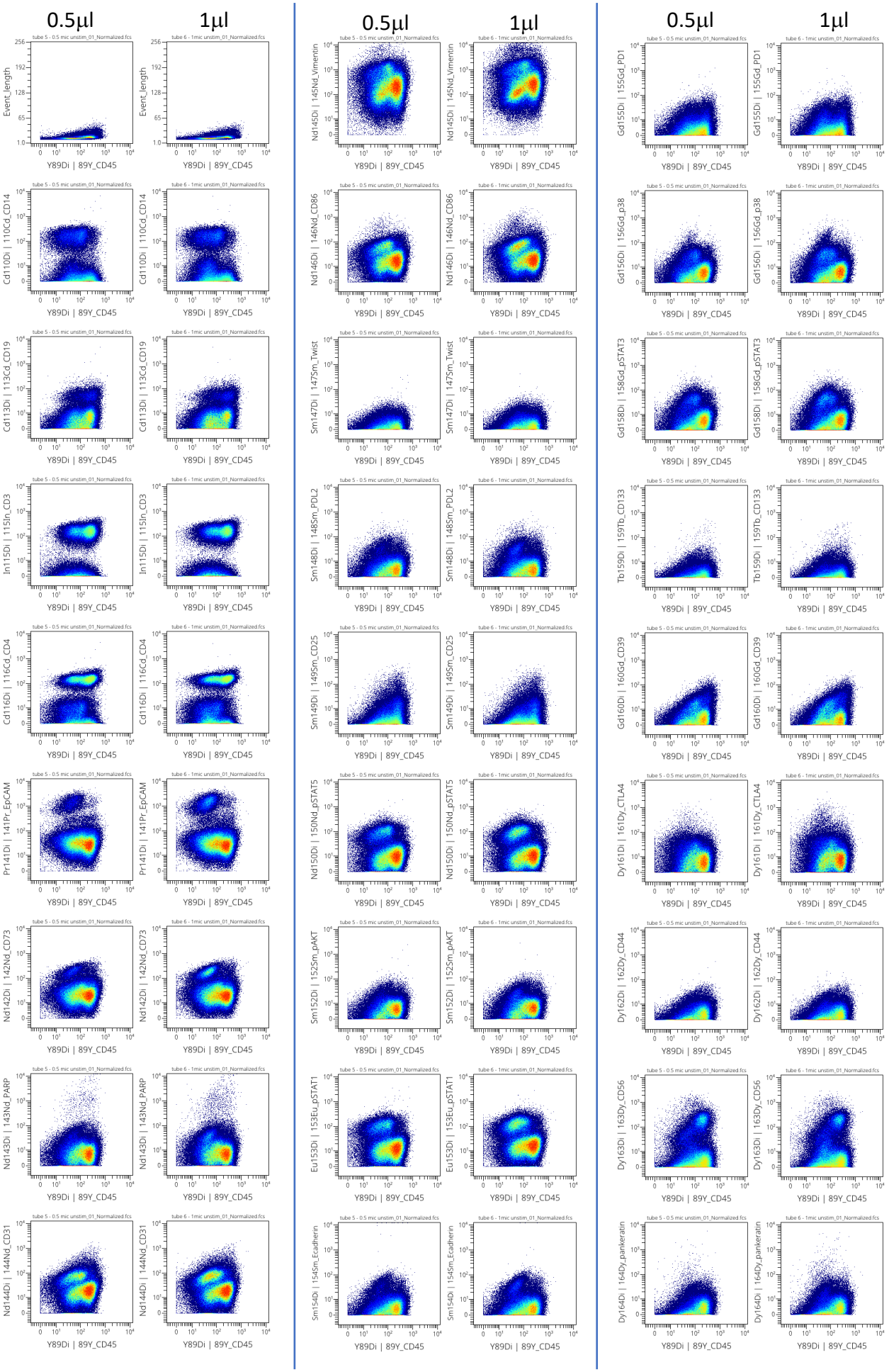

0.25µl
0.5µl
1µl
T01_tube 1 - 0.25 extra_01_Normalized.fcs
T02_tube 2 - 0.5 extra_01_Normalized.fcs
T03_tube 3 - 1 extra_01_Normalized.fcs
Yb174Di | 174Yb_Snail1
Lu175Di | 175Lu_pS6
Yb176Di | 176Yb_pCREB
Pt195Di | 195Pt_CD8
Hg196Di | 196Hg_HLADR
Hg198Di | 198Hg_HLAABC
Bi209Di | 209Bi_CD47
Y89Di | 89Y_CD45

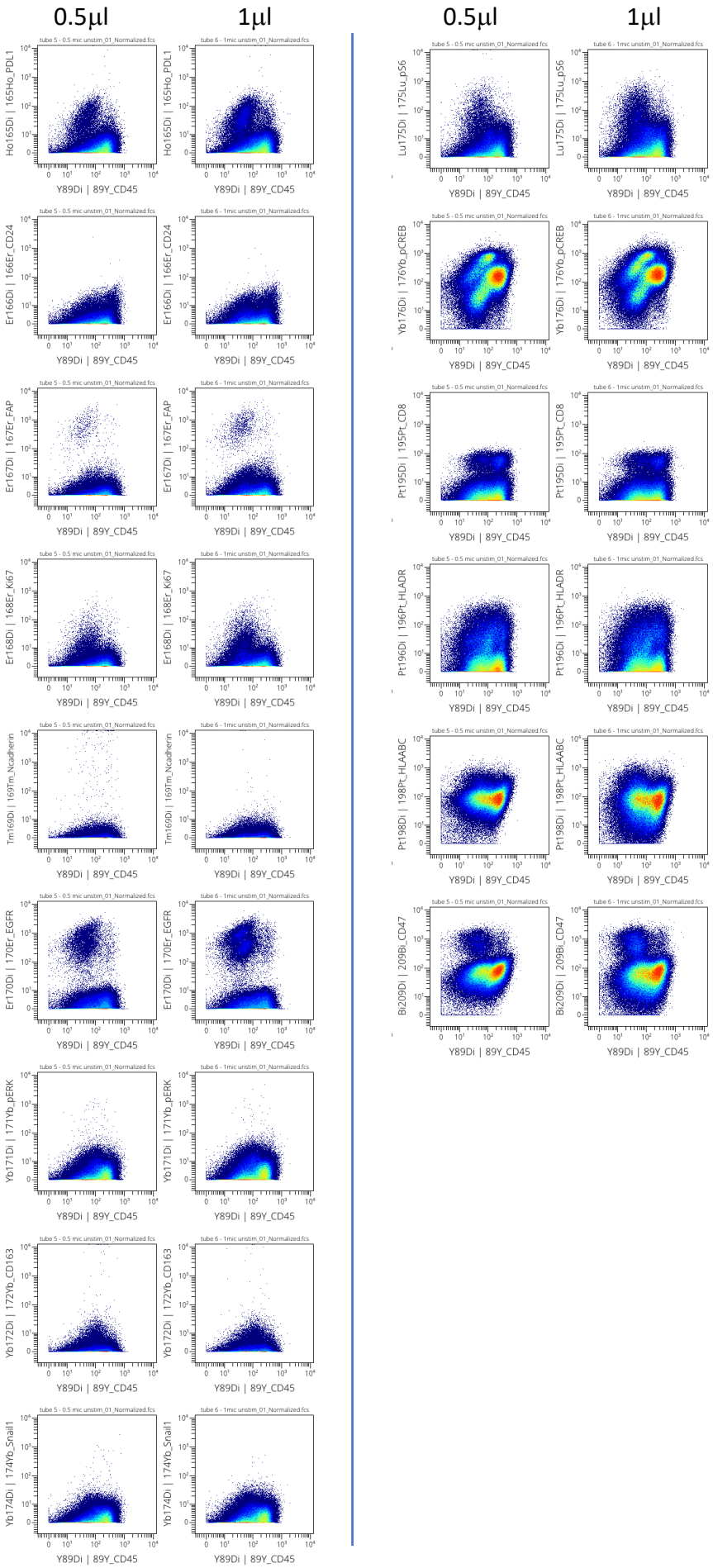
Intracellular marker titration (stimulated cells)




0.5µl
1µl
0.5µl
1µl
tube 2 - 0.5mic stim_01_Normalized.fcs
tube 3 - 1mic stim_01_Normalized.fcs
Ho165Di | 165Ho_PDL1
Er166Di | 166Er_CD24
Er167Di | 167Er_FAP
Er168Di | 168Er_Ki67
Tm169Di | 169Tm_Ncadherin
Er170Di | 170Er_EGFR
Yb171Di | 171Yb_pERK
Yb172Di | 172Yb_CD163
Yb174Di | 174Yb_Snail1
Lu175Di | 175Lu_pS6
Yb176Di | 176Yb_pCREB
Pt195Di | 195Pt_CD8
Pt196Di | 196Pt_HLADR
Pt198Di | 198Pt_HLAABC
Bi209Di | 209Bi_CD47
Y89Di | 89Y_CD45

Intracellular marker titration (unstimulated cells)




0.5µl
1µl
0.5µl
1µl
Ho165Di | 165Ho_PDL1
Er166Di | 166Er_CD24
Er167Di | 167Er_FAP
Er168Di | 168Er_Ki67
Tm169Di | 169Tm_Ncadherin
Er170Di | 170Er_EGFR
Yb171Di | 171Yb_pERK
Yb172Di | 172Yb_CD163
Yb174Di | 174Yb_Snail1
Lu175Di | 175Lu_pS6
Yb176Di | 176Yb_pCREB
Pt195Di | 195Pt_CD8
Pt196Di | 196Pt_HLADR
Pt198Di | 198Pt_HLAABC
Bi209Di | 209Bi_CD47
Y89Di | 89Y_CD45