Comprehensive EBV-specific T cell profiling and multi-omics analysis of nasopharyngeal carcinoma CD8 T cells reveals novel cancer-associated phenotypes

Nandita Kumar

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Evan W. Newell, Chair

David M. Koelle

Edus H. Warren

Program Authorized to Offer Degree:

Molecular Medicine & Mechanisms of Disease

University of Washington

Abstract

Comprehensive EBV-specific T cell profiling and multi-omics analysis of nasopharyngeal carcinoma CD8 T cells reveals novel cancer-associated phenotype

Nandita Kumar

Chair of the Supervisory Committee:
Evan W Newell

Nasopharyngeal carcinoma (NPC) is a high mortality, Epstein-Barr virus (EBV)-driven cancer with limited treatment options. Although EBV is highly antigenic and detectable in nearly all NPC tumor cells, it remains unclear why NPC tumors evade immune targeting despite high levels of immune infiltration. Here, we investigate the phenotypes of peripheral EBV-specific T cells with the purpose of identifying clinically relevant, cancer-associated EBV-specific T cell phenotypes. We developed a 34-marker mass cytometry and multiplexed MHC-I tetramer panel to detect and phenotype CD8 T cells of up to 55 antigen-specificities, including EBV and other known viral and non-viral epitopes. We used this approach to profile peripheral blood mononuclear cells (PBMCs) from treatment-naïve NPC patients (n=51). Compared to other antigen-specific T cells detected in

NPC patients and EBV-specific T cells in healthy individuals, EBV-specific T cells in NPC patients expressed high levels of activation markers (CD38, HLA-DR), co-inhibitory markers (CD39, TIGIT, PD-1), and migratory markers (ITB7, CD103). Using unsupervised clustering, we found that the frequency of EBV-specific T cells in a phenotypic cluster characterized by high CD103, CD39, CD71, HLA-DR, and CD38 expression positively correlated with plasma EBV-DNA levels and gross tumor volume. Remarkably, we observed that regardless of antigen specificity, NPC patients had a higher frequency of total CD8 T cells with this activated/exhausted cluster phenotype, and CD8 frequencies of this cluster were also positively correlated with plasma EBV-DNA titers, tumor stage (TNM), and EBV early antigen antibody titers. To investigate the gene expression profiles and TCR clonal diversity of this cancer-associated phenotype, we performed single-cell multi-omics analysis on peripheral CD8 T cells from 17 patients from this cohort. We identified that cell subsets within this phenotype had distinct gene expression patterns indicative of potential associations with the tumor and we also assessed the clonality of these subpopulations. Finally, we assess connections between the NPC tumor microenvironment using multiplexed immunohistochemistry (mIHC) and bulk tumor gene expression and TCR sequencing. We found that peripheral activated/exhausted T cell phenotypes are correlated with mIHC tumor PD-1 expression, neutrophil levels identified using gene expression data, and have high TCR clonal sharing with NPC tumors. Overall, this research identified unique cancer-associated peripheral EBV-specific T cell phenotypes in the context of nasopharyngeal carcinoma that are positively correlated with advanced disease. Frequencies of peripheral T cells with this phenotype could also be useful as a blood-based T cell correlate for NPC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This work would not have been possible without the invaluable contributions from so many individuals who have been pivotal to my scientific journey. They have all been instrumental in fostering my curiosity, encouraging my growth, and paving the way forward so that I could get to where I am today. No amount of gratitude is enough to sufficiently convey my thanks.

My parents were the first to encourage my excitement about the natural world, they championed all my interests from birds to fish. In third grade, my neighbor Amy, a microbiology graduate student, showed me how to make wet mounts and use her childhood microscope, introducing me to a world beyond what meets the eye. Teachers like Mrs. Paula Schmidt and Ms. Susan Duncan saw potential in me and helped me see that in myself. The compassion shown to me by my AP Biology teacher Mr. Dan Ivie was a guiding light that allowed me to illuminate my passion for the biological sciences during otherwise bleak times. In college, I was given the lifechanging opportunity to do research in Dr. Jadwiga Giebultowicz's lab. The immeasurable investment she made in my mentorship and the skills that I gained from my work in her lab provided me with a strong scientific foundation. As a research technician, I got the opportunity to work with many amazing scientists including Dr. Ursula Sandau, Dr. Danielle Osborne, and Dr. Bahar Shamloo. They provided a safe sisterhood of women in STEM and welcomed me to join the fold. I learned so much from working with Dr. Shamloo, especially about balancing science with creativity and appreciating all facets of life.

I am immensely thankful to have gotten the opportunity to do my dissertation work in Dr. Evan Newell's lab. Evan is an exceeding kind, generous, and encouraging mentor and I could not have gotten through all the ups and downs of the last five and half years without him. I feel very privileged to know that I can count on his support no matter what is going on, even a global

pandemic. I have learned so much from him not just about science, but also the power of empathetic leadership, for which I am eternally grateful. When I first joined the lab, Summer Zhuang, Timothy Bi, Dr. Shamin Li, and Dr. Yannick Simoni made me feel so welcome and taught me so much. I am so grateful to all of the members of the lab that I have worked with ever since. In every iteration, the Newell lab has been the best bunch of scientists and people, including Dr. Heeju Ryu, Dr. Amy Codd, Ana Jojic, Dr. Hugh MacMillan, Dr. Allie Donlan, Ernest Moelhman, Long Nguyen, Dr. Ju Kim, Dr. Daniel Jones, Korok Sarkar, Michelle Wurscher, Dr. Ye Zheng, Guoyue Xue, Tony Chour, Dr. David Glass, Anthony Cessna, Dr. K. Rachael Park, Dr. Rachel Zemek, Josh Yee, Dr. Damian Oyong, Dr. Carolyn Shasha, Emily Park, and Laura Islas. I am so grateful to all forms of support that I received from you whether it was help learning lab techniques or computational skills or providing a listening ear, good vibes, or delicious treats. It would take yet another thesis to full convey the full extent of my gratitude, and I am grateful to count you among my friends. In addition, I would like to thank my committee, Dr. Hootie Warren, Dr. David Koelle, Dr. Denise Galloway, and Dr. Ram Savan. I am also thankful to my collaborators, Dr. Amit Jain, Dr. Zhiwei Liu, Dr. Melvin Chua, Dr. Joe Yeong, and Dr. Darren Lim for helping me take my research to the next level. I am also grateful to all Fred Hutch core facilities staff, particularly Rebecca Reeves, Ben Janoschek, and Dolores Covarrubias. I am grateful for the support of my graduate program including Dr. Bill Mahoney and Megan Barker, as well as all my fellow M3D students who have provided lots of solidarity and friendship. My non-science-related supporters also deserve a huge thanks. I am so grateful to have such a supportive family, thank you Amma, Appa, Medha, and Thatha. I would like to thank my friends for always being there for me, including Maddie, Kamna, Pauline, Gillian, Allison, and so many others. Lastly, I would like to thank the patients and their families for their invaluable contributions to this work.

DEDICATION

To Thanga Paati and Lakshmi Paati and all those that came before.

Chapter 1. INTRODUCTION

1.1 EPSTEIN-BARR VIRUS

1.1.1 *Virology*

Epstein-Barr virus (EBV) is a human herpes virus (HHV4) that leads to life-long chronic infection in >90% of the human population[1]. EBV was the first tumor virus to be identified and it was first isolated from a Burkitt lymphoma tumor sample in 1964[2,3]. EBV infection is primarily transmitted through saliva, and most individuals infected with EBV are infected early in life. However, EBV infection later in life is common in developed countries, where EBV-induced infectious mononucleosis (IM) is more prevalent[3]. During primary EBV infection, epithelial cells and B cells in the oropharyngeal tonsils are infected with EBV, and a small number of B cells remain latently infected post primary infection[3]. To mediate viral entry, EBV uses unique receptors on the surface of B cells and epithelial cells[3].

The viral life cycle of EBV can be segregated into two distinct phases: the lytic and latent cycles. Many EBV gene products are expressed during both these cycles, and each phase of the life cycle has a unique program of gene products that are expressed. In latency, the EBV genome exists as an episome that is tethered by EBNA1 to the host's DNA, and various lytic cycle gene products such as BRLF1 and BZLF1 are expressed in order to initiate lytic cycle reactivation, which is needed for viral genome replication and infection[3]. EBV latency is maintained in many different programs, and the unique subset of gene products expressed during each program define each latency phase. Latency program 0 has the fewest expressed viral gene products and is maintained in quiescent B cells, whereas subsequent latency programs (Latency 1, Latency 2, Latency 3) have an increasing variety in the gene products that are expressed (EBNA1, LMP1,

LMP2A etc.). While EBV infection in healthy individuals is latent and largely asymptomatic, sporadic lytic reactivation occurs in order to ensure that a quorum of B cells remain latently infected[3].

1.1.2 *EBV-specific Immune Responses*

In healthy individuals, EBV infections are largely asymptomatic and only occasionally lead to disease because EBV infection is usually readily controlled by host immune responses. In terms of innate immune responses, EBV gene products have unique pathogen-associated molecular patterns (PAMPs) that can be recognized and targeted by innate immune receptors called pattern recognition receptors (PPRs). Some known EBV PAMPs include gp350 dUTPase which is targeted by a PRR called toll-like receptor 2 (TLR2), and EBERs which are targeted by toll-like receptor 3 (TLR3)[3,4]. In terms of adaptive immune responses, CD8 T cells play a crucial role in controlling lytic reactivation of EBV-infected B cells and these EBV-specific T cell responses are well studied[5]. EBV peptides are presented to CD8 T cells by major histocompatibility class I (MHC Class I) molecules, which then leads to a large expansion of EBV-specific T cells, which can be quantified using MHC Class I tetramers loaded with EBV-specific peptides. Studies have shown that T-cell responses differ based on whether the source of antigen is from lytic or latency-associated genes, which is logical since these two types of antigens are presented at different timings and contexts[2,5]. In fact, previous studies from our group using mass cytometry have shown that EBV T cell phenotypes differ based on whether these cells are specific for lytic or latent antigens[6]. Here, it was observed that latent EBV-specific T cells had more central-memory-like phenotypes as compared to lytic EBV-specific T cells that had more effector-like phenotypes[6]. While the mechanisms of immunological control of EBV viral infection and infectious

mononucleosis is well-studied, relatively few studies have investigated EBV-specific T cell responses in the context of nasopharyngeal carcinoma (NPC) and other EBV-related diseases.

1.1.3 *EBV-associated diseases*

EBV plays a role in the pathogenesis of infectious mononucleosis, and post-transplant EBV reactivation. Importantly, EBV infection is a key etiological factor in the pathogenesis of NPC and other EBV-associated cancers, which total ~200,000 cases each year[7]. In addition to NPC, other related EBV-driven carcinomas of nasopharyngeal type or lymphoepitheliomas with similar characteristics have been reported in tissues such as the thymus, tonsil, lungs, stomach, skin, and uterine cervix[8]. Furthermore, EBV infection is responsible for the development of many other cancers such as Hodgkin's lymphoma, Burkitt lymphoma, and some forms of gastric cancer (also common in Asia)[7]. Additionally, lymphoproliferative diseases driven by EBV often arise in immunocompromised patients, including individuals with human immunodeficiency virus (HIV) and cancer patients post-bone marrow transplant. In recent years, the link between autoimmune diseases such as multiple sclerosis (MS) and systemic lupus erythematosus (SLE) has also been strengthened[3,9]. Thus, the study of T-cell responses in the context of EBV-driven cancer has broad implications across many diseases and immunologically vulnerable populations.

1.2 NASOPHARYNGEAL CARCINOMA

1.2.1 *Etiology and Risk Factors*

NPC is a common form of head and neck cancer with approximately 129,000 new cases of NPC reported worldwide in 2018[10]. NPC is categorized into three different histological types that include keratinizing, non-keratinizing, and undifferentiated[7]. Non-keratinizing NPC is

endemic in East and Southeast Asia where >70% of NPC cases are reported[7]. The overall 5-year survival rate for non-keratinizing NPC is only 51%, which represents an unmet clinical need[11]. The reason for this geographic distribution is not clear and likely involves environmental and/or virological factors[12,13]. Research in recent years has suggested that variations in the EBV viral genome could confer increased risk for NPC, especially in endemic regions[14]. In addition, certain HLA-types have been identified that confer increased risk of NPC, such as HLA-A*02 and HLA-B*46 [15]. Other risk factors for NPC include smoking and salty fish consumption, which is prevalent in endemic regions[7]. In terms of the demographics of NPC, this cancer occurs more commonly in males than females. However, compared to all other risk factors EBV infection is thought to confer the highest risk of NPC and is even viewed as the causal agent of this cancer[7].

1.2.2 *Nasopharyngeal Carcinoma Tumorigenesis and Epstein-Barr-virus*

NPC arises from the epithelial lining of the nasopharynx and nearly all cases of non-keratinizing NPC are driven by EBV infection[7]. Although EBV is usually maintained in Latency 2 in NPC, suggesting that the tumors would mostly express a subset of latency-associated EBV genes[16,17], sporadic expression of lytic cycle antigens in this tumor have also been reported[18-22]. Tumorigenesis is thought to occur due to a series of aberrant genetic mutations in epithelial cells, which in combination with latent EBV infection, leads to the clonal expansion and proliferation of EBV infected tumor cells. Therefore, all NPC tumor cells are infected with EBV, which can be detected pathologically by staining for EBV-encoded small RNAs (EBERs), which are abundantly expressed by EBV-infected cells[3,7].

1.2.3 *Standard of Care for Nasopharyngeal Carcinoma*

In terms of diagnosis and prognosis assessment for NPC, currently few biomarkers exist that are predictive of patient clinical status or outcomes. One potential prognostic/predictive factor that has been well studied and is used as a part of clinical evaluation is circulating EBV-DNA levels[23]. EBV DNA detected in NPC patients is thought to originate from the tumor and therefore presents a relatively easy way to detect circulating free tumor DNA (ctDNA)[24]. High levels of EBV-DNA have been associated with more advanced disease and persistently high levels of the EBV-DNA titer post-treatment are associated with poor prognostic outcomes[7,25,26].

In addition to EBV-DNA titers, plasma EBV antibody levels for EBV viral antigens like viral capsid antigen (VCA), early antigen (EA), and EBV nuclear antigen 1 (EBNA1) have been preliminarily evaluated for the diagnosis of NPC and for population screening to identify individuals at risk for developing NPC[27]. However, there is no consensus around the utility of these titers since these antibodies can be found in healthy individuals, in addition to their low sensitivity/specificity[7]. One systematic study by Tan et al., comparing plasma VCA IgA, EBNA-1 IgA, EA IgA, and EA IgG titers across stage in NPC patients found that there was not a clear correlation between stage and high antibody titers, however VCA IgA and EA IgG titers as compared to EBNA-1 IgA titers were superior in detecting early stage NPC[26]. Another study assessing the utility of VCA IgA and VCA IgG in predicting survival in the context of NPC found that these antibody titers were not predictive of survival[28], and similarly, a study assessing EA IgA and IgG titers found that EA serology was elevated in both patients with active disease and in remission[29]. In the context of other EBV-related diseases, a cohort of EBV-associated gastric carcinoma showed increased antibody responses for lytic antigens compared to latent

antigens[30], and unique patterns of EBV antibodies were also found in the context of extranodal natural killer/T-cell lymphoma (NKTCL)[31].

Treatment of NPC post-diagnosis is largely limited to radiation and chemotherapy, and surgical resection is not commonly used due to the sensitive region in which this tumor occurs. Various radiation methods can be used with intensity-modulated radiotherapy (IMRT) being the most common approach[7]. In addition to radiation, concurrent, adjuvant, and induction chemotherapy are also often implemented. While methods of administering radiation and chemotherapy have greatly improved and have high efficacy, patients treated with these therapeutic modalities often experience adverse effects, which can greatly reduce quality of life. Therefore, additional treatment options are still necessary to improve not just patient outcomes but also patient quality of life.

1.2.4 *Nasopharyngeal Carcinoma Immunology and Immunotherapy*

Compared to many other solid tumors, most NPC tumors are heavily infiltrated with lymphocytes and express PD-L1[7], and therefore NPC is a promising candidate for new immunotherapeutic treatments. Both checkpoint-blockade and adoptive cellular immunotherapy approaches are actively being tested in many NPC clinical trials[32-35]. However, the roles for EBV-specific T cells in the control of NPC have yet to be determined. Ongoing studies are using adoptive transfer of EBV-specific T cells and have shown some promise[36-39]. Checkpoint blockade immunotherapy is also being tested for NPC, but anti-PD-1 therapy alone using nivolumab has shown a low rate of efficacy (overall response rate ~20%)[33]. Additional trials for immune checkpoint blockade therapy have been conducted for NPC using therapeutics such as pembrolizumab and camrelizumab[7]. In recent studies, using toripalimab in conjunction with

chemotherapy increased progression-free survival and was given FDA approval for treatment of advanced NPC[40,41].

Due to the low rates of response to anti-PD-1 monotherapy alone, the safety and efficacy of a combinatorial approach of anti-PD1 and anti-CTLA4 therapy (Ipilumimab + Nivolumimab) have been recently assessed in a phase II clinical trial (NCT03097939)[42]. The results of this study lead to the conclusion that combined anti-PD-1/CTLA4 therapy is safe and achieved durable responses in recurrent and metastatic NPC patients (best overall response rate = 38%). In this clinical trial, patients with low baseline EBV-DNA levels had a higher median progression-free-survival rate in comparison with patients with high EBV-DNA levels, and no confirmed partial responses were observed in any patients with high EBV-DNA levels. This suggests that EBV-DNA levels could be a marker of response to immune checkpoint blockade therapy in NPC patients.

1.3 Cancer Immune landscapes of the Tumor and Periphery

1.3.1 *Immunology of the Tumor and Connections with the Periphery*

A huge area of interest within cancer immunology, particularly in the context of solid tumors is understanding how immune cells are recruited to the tumor, what happens to immune cells once they have entered and adapted into the tumor microenvironment, what interactions T cells have with tumor cells and with other immune cells, and how immune cells egress from the tumor to lymphoid tissues and the blood. In recent years, there has been increasing focus on not only understanding tumor-immune dynamics within the tumor, but also to understand the systemic dynamics of immune responses mounted to target the tumor. This is exemplified by the model provided by the cancer-immunity cycle[43], which is a framework by which we can understand how intra-tumoral T cell responses lead to external trafficking of T cells to the

tumor.[43] Here we will summarize the paradigm described by Chen and Mellman as it provides a useful method of viewing immune responses in the blood, lymphoid tissues, and tumor in concert. The cycle begins with the release of cancer-specific antigens from the tumor into the periphery where they are taken up and presented by antigen presenting cells (APCs). These APCs then present cancer-specific antigens to T cells in the periphery and lymphoid tissue. T cells then become activated post antigen priming and traffic to the tumor when they can mount anti-cancer immune responses causing tumor cell death, releasing cancer-antigens that restart the cycle and broaden immune responses. If this cycle worked perfectly each renewed immune response would be more targeted and the tumor would be eliminated. However, we know that is not the case. In fact, the utility of this model is that each of these steps represents a point at which anti-tumor immune responses can and do fail. Fleshing out how the cancer-immunity cycle works, and more importantly does not work, can help us better understand the cancer immune landscape and tailor immunotherapies to aid where the host immune responses are failing. There has been extensive research on each one of these steps, and yet, much is still unknown when it comes to the immune dynamics of each step.

1.3.2 *Current Methods of Studying Peripheral Tumor-specific T cells*

A major challenge in the field of cancer immune monitoring is that tumor-specific T cells are difficult to study in peripheral blood samples[44]. Additionally, a better understanding of the characteristics and roles of tumor-specific T cells remains an important challenge for improving efficacy of immune-based cancer therapies[45]. Tumor-specific T cells can be difficult to detect in the blood due to their very low frequencies[46,47] and in the tumor because of the variable and often high abundance of non-tumor specific bystander T cells that infiltrate tumors[48,49]. In addition, only a few generalizable cancer antigens have been identified such as NY-ESO-1[50], and many tumor-

specific T cell antigens are likely neo-antigens which arise due to mutations that occur in tumor cells and are therefore personal to each patient. Furthermore, identifying T cells specific for these patient-specific neo-antigens relies on sequencing the patient tumor exomes and using imperfect prediction algorithms to impute these neoantigen peptide sequences from this data[50]. Therefore, this process of identifying tumor neo-antigens can be very costly, labor intensive, and time consuming[50]. Because all NPC tumor cells are infected with EBV, and peripheral CD8 EBV-specific T cells are typically abundant and easily identifiable, we have a unique opportunity to study peripheral viral antigen-specific T cells that could likely also be tumor-specific in the context of this EBV-driven cancer.

1.4 SUMMARY

The goal of this work was to better understand the role of EBV-specific T cell responses in nasopharyngeal carcinoma (NPC) immunopathology. In the following chapters, we test the hypothesis that peripheral EBV-specific T cells are reflective of anti-tumor immune responses and that the phenotypic and clonal characteristics/dynamics of peripheral EBV-specific T cells are indicative of clinical disease status for NPC patients. Overall, the aim of these studies was to not only improve our understanding of the NPC immune landscape, but also potentially identify T cell populations associated with immunotherapeutic response that could be tested in future NPC immunotherapy trials.

MATERIALS AND METHODS

1.5 Patient Sample Acquisition and Processing

1.5.1 *Human Samples*

Peripheral blood and tumor tissue samples were obtained from newly diagnosed, untreated patients with nasopharyngeal carcinoma (n=51) (Table 1). In addition, peripheral blood samples from healthy, HIV negative adult donors (n=12) were obtained through the HIV Vaccine Trials Network, FHCC). All patient and healthy donor samples were obtained and used with informed consent and in adherence to institutional review board approval.

1.6 Mass Cytometry

1.6.1 *Heavy Metal-labelled Antibody and Streptavidin Conjugation*

Purified antibodies were purchased and labelled with heavy metals using the antibody conjugation protocol provided by Standard Biotools (Table 2). Streptavidin was conjugated to heavy metal labels as previously described[6]. All metal-labelled reagents were tested prior to use and each metal-labelled antibody was titrated to identify the optimal concentration.

1.6.2 *Multiplexed MHC Tetramer Cocktail Preparation*

For mass cytometry staining, the multiplexed MHC tetramer cocktail was prepared as previously described[51]. Briefly, 8 unique heavy metal-labelled streptavidins were generated. For each experimental batch, a large panel of known viral and tumor antigen-associated epitopes for the HLA-A alleles of patients within the batch were each assigned to a unique combination of three metal-labelled streptavidins (8 choose 3) (see Table 3). Using an automated pipetting device, three metal labelled streptavidins were sequentially added into individual wells of a 96-well plate

as per the previously assigned combinations. In a corresponding 96-well plate, each epitope was peptide exchanged with the appropriate biotinylated, UV-cleavable MHC Class I monomer for HLA-A allele monomers (HLA-A*02:01, HLA-A*11:01, and HLA-A*24:02, Immune Monitoring Core, FHCC) by UV light exposure and incubated overnight. HLA-B monomers were not used since only HLA-B*07 monomer was available, and this HLA-B allele was not prevalent in our cohort. Each peptide-exchanged MHC monomer was then tetramerized to the corresponding mix of three metal labelled streptavidins and this reaction was quenched by the addition of free biotin. Finally, all tetramers were combined, concentrated, filtered (0.1 μm filter, Millipore), and diluted to generate the final MHC tetramer cocktail with equal concentration of tetramers across all epitopes. 100μl of 0.1 mg/mL MHC Class I Monomer and 5 μL of 1mM peptide were used for each peptide exchange, 30μL of 50 mg/mL mixed metal labelled streptavidin were added to each peptide-exchanged monomer, and 100 μl of the final cocktail was used to stain each patient.

1.6.3 *Mass Cytometry Staining and Data Acquisition*

Peripheral blood samples were stained for mass cytometry as previously described[51]. Briefly, for each experimental batch, frozen PBMC samples (including one healthy donor leukapheresis sample to use as an internal batch control and bridge sample) were thawed into pre-warmed thawing medium buffer (90%RPMI+10%FBS), transferred into a V-bottom 96 well plate, and incubated for 5 minutes with DNase (final concentration: 0.1 mg/mL) in thawing medium. Next, the cells were washed two times with FACS buffer (90%PBS+10%FBS) and incubated with the multiplexed MHC tetramer cocktail for 1 hour at room temperature. After incubation, the cells were washed once with FACS buffer and additionally with PBS and then resuspended in diluted cisplatin and incubated for 5 minutes at room temperature. After washing the cells in FACS

buffer, the cells were incubated for 20 minutes on ice in the dark with diluted anti-γδTCR-PE (Invitrogen), washed with FACS buffer, and incubated again in a cocktail of metal-labelled antibodies for an additional 20 minutes (see Table 2). After 2 washes in FACS buffer and 1 wash in PBS, cells were fixed in 2% paraformaldehyde and incubated overnight. On the day of acquisition, cells were washed 3 times with permeabilization buffer, incubated with intracellular antibody cocktail for 30 minutes, and washed twice with permeabilization buffer and once with PBS. Each sample was then barcoded with two in-house prepared metal-labelled barcodes as previously described[52] for 30 minutes on ice. After barcoding, cells were washed twice with Milli-Q water, stained with DNA intercalator (Cell-ID™ Intercalator-Ir, Standard Biotools) for 10 minutes at room temperature, and washed twice with Milli-Q water. Cells were then resuspended in Milli-Q water and prior to acquisition cells were strained using mesh-capped FACS tubes. and EQ Four Element Calibration Beads were added (#201078, Standard Biotools). Data was acquired on the Helios™ CyTOF® system (Standard Biotools).

1.6.4 *Mass Cytometry Data Pre-processing*

Raw data from each experimental batch were normalized for time-associated signal fluctuation based on EQ Four Element Calibration bead signal using previously developed software [53]. Then, data from each sample was concatenated, and by manual gating using FlowJo v10 software, calibration beads were removed, DNA Intercalator staining and Residual measurements were used to identify singlets, and CD45 and Cisplatin staining were used to identify viable immune cells. Using previously described methods[54,55], spillover signal between heavy metal channels was corrected for gated cells from each samples using an script that returned spillover corrected .fcs files for each sample (R (http://www.r-project.org), R Studio (https://www.r-studio.com/)). Next, zero-values were randomized to a uniform distribution

between negative one and zero using R. Finally, technical inter-batch variation in antibody surface marker expression was normalized using *ADTnorm* by peak and valley alignment and scaling of each metal-labelled antibody channel based on the known biological expression pattern of the marker (i.e., CD3 has bi-modal expression, CD4 has tri-modal expression, PD-1 has gradient expression lacking a defined positive peak etc.) [56]. Mass cytometry data from all patient and healthy donor samples were processed in this manner and used for downstream analysis.

1.6.5 *Mass Cytometry Data Analysis*

For hypothesis testing, immune cell subsets were first quantified using biaxial gating (FlowJo v10). For downstream CD8 T cell analysis, antigen-specific T cell populations and surface marker phenotypic populations were calculated as a frequency of total CD8 T cells, and not using absolute lymphocyte counts and therefore does not account for the presence and variability of patient lymphopenia. In addition, total CD8 T cells and antigens-specific T cell populations from each sample were used for high dimensional clustering using *Phenograph*[57], followed by dimensionality reduction visualization using uniform manifold approximation and projection (UMAP) [58,59]. The resulting cluster frequencies for each sample were then used for an unbiased analysis of differential cellular composition[60,61] among the NPC patient cohort and healthy donor leukapheresis samples. In addition, the cluster frequency of each antigen-specific T cell population (i.e., percent of the antigen-specific T cells within the population in each cluster) was also calculated.

1.7 SINGLE-CELL MULTI-OMICS

1.7.1 *Oligo-barcoded MHC-I Tetramer Cocktail Generation*

To detect CD8 T cell antigen-specificity in our single cell RNA-sequencing experiments, MHC class I tetramers were generated using oligo-barcoded streptavidins (Biolegend, TotalSeq™-C). Peptide epitopes for each patient are described in Table 5. First, peptide exchange was performed by transferring 0.5µL of a 1mM peptide stock to 10 uL of 0.1 mg/mL biotinylated MHC class I monomers in a 96-well plate. Then, each well was pipette-mixed, and the plate was exposed to 10 minutes of UV light and rotated after five minutes. The peptide exchanged peptide-monomers were incubated overnight at 4°C in the dark. To assemble the tetramers, 1 µl of 50 µg/mL oligo barcoded streptavidin was added, mixed by pipetting up-and-down, and incubated at room temperature in the dark for 10 minutes. Two more additions of streptavidin were added with 10-minute incubations between additions. The tetramers were quenched by subsequently adding 1µl of biotin solution and pipette to stop the reaction and the sealed plates were incubated for ten minutes. In addition, to remove mucosal-associated invariant T cell (MAIT) contamination, an MR-1 tetramer was generated using the same protocol, but tetramerized to a fluorophore conjugated streptavidin. To generate the tetramer cocktail, all tetramers for each individual patient sample were combined, filtered using a 0.4 µm filter tube, and diluted to a final volume of 150 µL in HBSS+0.04% BSA.

1.7.2 *CITE-seq Antibody and Tetramer Pre-staining and FACS Sorting*

Oligo-barcoded antibodies listed in Table 4 and fluorophore surface marker antibodies for FACS sorting of total CD8 T cells were combined into two separate cocktails in PBS+10% FBS, filtered using a 0.4 µm filter tube, and stored on ice. Cryopreserved PBMC samples were thawed

as described above for mass cytometry staining, washed with HBSS+0.04% BSA, and resuspended with wide-bore pipette tips and transferred to a cluster tube. To reduce non-specific antibody staining, the cells were incubated with Fc block (10 µL Tru-Stain Fc block in 40 µL HBSS-0.04% BSA) on ice for 10 minutes. For some batches, patient samples were oligo-hash-tagged by adding hashtag antibodies (Biolegend TotalSeqC) with each oligo-hashtag was assigned to a single patient so that cells from each patient sample could be identified by oligo-hashtag staining. Cells were then incubated on ice for 20 minutes and combined for further staining. Patient samples were then stained with the oligo-barcoded tetramer cocktail after washing once with HBSS+0.04% BSA and adding the previously generated corresponding tetramer cocktail (see above). Cells were stained for 1 hour in the dark at 4°C and resuspended often with wide-bore pipette tips to prevent cell clumping. After one wash with PBS+10% FBS, the surface antibody cocktail was added, cells were pipette-mixed, and were incubated in the dark for 30 minutes. After two washes with PBS+10% FBS, cells were stored on ice and stained with viability marker 7-AAD prior to FACS sorting (3µL in 47µL PBS+10% FBS). CD8 T cells were sorted into PBS+10% FBS using the BD FACSSymphony S6 sorter (BD Biosciences) and kept on ice prior to single-cell gel-bead capture.

1.7.3 *Single-cell RNA-sequencing Library Preparation*

Post sorting, single cell libraries were prepared using the Chromium Single Cell 5' Reagent Kits User Guide (v2 Chemistry Dual Index) with Feature Barcoding technology for Cell Surface Protein and Immune Receptor Mapping from 10X Genomics. Briefly, gel bead capture was performed by loading cells recovered from FACS sorting into Chip K, targeting ~60,000 cells based on the number of cells sorted provided by the FACSSymphony S6 sorter. Gel bead emulsions were created using the Chromium Controller, and the reverse transcription incubation was performed in

accordance with the manufacture protocol. Next, complimentary DNA (cDNA) amplification was performed, and the resulting cDNA was used to generate gene expression, feature barcode (CITE-seq), and TCR library for each profiled sample. Libraries were then sequenced using Illumina sequencing with 2 by 92 base pair paired-end reads with a target of sequencing 20,000 reads per cell for each sample.

1.7.4 *Single-cell Data Pre-processing*

All sequencing libraries were aligned to the hg38 (GRCh38) genome using Cell Ranger (v.6.0.0), and the resulting data from each batch was initial loaded into a Seurat[62] object for downstream data quality control steps. Cells from each batch were filtered to remove cells with low and high RNA read counts (<200 and >4,000) and nonviable or dying cells with high mitochondrial DNA counts (>10% mitochondrial DNA). Datasets from each batch was further normalized using the *NormalizeData* Seurat function. Next, *DoubletFinder*[63] was used to remove doublets by approximating the number of doublets using the number of droplets captured in each well based on provided 10X Genomics reference guidelines. Data from patient samples that were loaded into more than one well was merged, and samples that were hashtagged were manually demultiplexed by removing any cells that had expression of more than one hashtag and assigning the remaining cells to their corresponding patient hashtag code. Any contaminating mucosal-associated invariant T (MAIT) cells, CD4 cells, or B cells were removed using CD4, CD20, and CD161 feature barcode expression. All Seurat objects for each batch were then integrated using *Harmony*[64], and TCR genes and genes expressed in fewer than 10 cells were removed from the dataset. This batch corrected and integrated Seurat object was then using for downstream analysis.

1.7.5 *Single-cell Data Analysis*

Unsupervised clustering of the gene expression and surface marker protein expression data from the batch integrated Seurat object was performed using Louvain clustering, and UMAP plots were generated to visualize clustering using only the gene expression data, only the surface marker expression data, and using both the gene expression and surface marker expression data. Dot plots to identify surface marker expression by cluster was created by using the *AverageExpression*() Seurat function, and by calculating percent positivity by manual gating each marker using all cells in the integrated dataset to define gating thresholds. Differential surface marker expression analysis was performed using the *FindMarkers()* Seurat function using a log foldchange cut-off of 0.25, and data was visualized using ComplexHeatMap[65] package. Differential gene expression analysis was performing using the *FindAllMarkers()* Seurat function, and the top 10 differentially expressed genes by highest average foldchange from each cluster were visualized using the *DoHeatmap()* Seurat function. The tissue resident memory (Trm) gene signature was calculated using data from Kumar et. al.[66] and the *AddModuleScore()* Seurat function, and the final score was calculated by subtracting the score calculated from the downregulated genes from score calculated from the upregulated genes. scRepertoire[67] was used to add TCR gene usage and clonality data to the integrated Seurat object and to visualized difference in clonality and clonal sharing between clusters. Manual gating of cells with feature barcode expression levels of oligo-tagged tetramers was used to identify antigen-specific T cells, and this data was added to the integrated dataset using the cell barcodes for cells that had positive tetramer feature barcode expression. The coding scheme for the tetramers can be found in Table 5. scCustomize[68], Seurat, and ggplot2 were used for data visualization.

1.8 Peripheral Blood Plasma Serology

Serum samples were process as previously described[27]. Briefly, serum sample testing was carried out using multiplex serology in a blinded manner. Six IgG and seven IgA antibodies against 11 antigens (VCAp18, EBNA1, BMRF1 [EA-D], BPLF1, BXLF1, LF2, BZLF1, BORF1, BFRF1, BGLF2, and BRLF1) previously shown to be strongly associated with NPC risk and whose combined score has been shown to be highly predictive of NPC risk [10], were targeted by the multiplex serology assay. Serum samples were pre-incubated at 1:50 dilution for IgA testing (final dilution 1:100) and at 1:5000 dilution for IgG testing (final dilution 1:10 000) in a pre-incubation buffer based on phosphate-buffered saline (PBS) with 2 mg/mL casein and additionally containing 2g/L of lysate proteins of *Escherichia coli* overexpressing glutathione-S-transferase-tag, 5g/L polyvinyl alcohol and 8g/L polyvinyl-pyrrolidone. Median fluorescence intensity (MFI) values were obtained for individual antibodies measured.

1.9 Tumor T-cell receptor Sequencing

Nucleic acid extraction was performed using the Oncomine TCR Beta-SR Assay kit (ThermoFisher) from FFPE specimens as follows: 1) Library preparation and clean up were performed as per protocols provided in the manual. 2) Sequencing was done on the S5 Gene studio using the 540 or 550 chips. 3) Torrent suite 5.10 and Ion Reporter 5.10 versions was used to run the results and analysis. Acquired TCR sequences were then mapped to peripheral TCRs sequences added to the batch corrected, integrated Seurat object using *scRepertoire*[67] based on exact amino acid sequence matches. Tumor samples with fewer that 100 recovered sequences were removed from the analysis, correction for sequencing depth was not performed. Sequences captured per samples and other TCR summary statistics are provided in Table 7.

1.10 TUMOR BULK RNA SEQUENCING

RNA was extracted from 3-4 pooled 15-micron thick curls from FFPE blocks using the TruSeq RNA Exome kit (Illumina). Data quality control (QC) included positive and negative sample controls and evaluation of probe binding density and image quality. Samples outside of acceptable ranges were flagged for manual review of raw data and necessary troubleshooting. Following QC, for analysis, transcriptomic raw counts were TPM normalized. Tumor purity for each sample assessed, however RIN scores were unknown. To assess the cellular composition within the tumor, *MCP-Counter* was used to quantify the relative presence of immune cell subsets and cancer-associated fibroblasts within each sample[69].

1.11 TUMOR MULTIPLEXED IMMUNOHISTOCHEMISTRY STAINING

Tumor biopsy FFPE slides were stained with the Opal Multiplex fIHC kit (Akoya Bioscience), as previously described[70-73]. Viable tumor regions were selected and images were captured for each case under a Vectra 3 pathology imaging system microscope (Akoya Bioscience) then analyzed and scored using inForm software (version 2.4.2)[74-76] and HALO TM (Indica Labs)[77-79].

Chapter 2. HIGH-DIMENSIONAL PROFILING OF THE NASOPHARYNGEAL CARCINOMA IMMUNE PERIPHERY USING MASS CYTOMETRY

2.1.1 *Introduction*

The overarching goal of this work was to survey the nasopharyngeal carcinoma immune periphery to identify circulating T cells indicative of tumor-related immune responses that could be useful correlates of disease. To this end, we collected cryopreserved peripheral blood mononuclear cells (PBMCs) and plasma from the patient peripheral blood samples and generated formalin fixed, paraffin-embedded (FFPE) tumor biopsies from a cohort of Singaporean individuals with untreated nasopharyngeal carcinoma (Fig. 1a). Various immune profiling methods were then used to collect data from these samples to deeply dissect the connections between the NPC tumor and the peripheral immune landscape and also to investigate correlations to disease status using known clinical parameters (EBV-DNA titer, gross tumor volume, survival, and stage (TNM)). Mass cytometry (CyTOF) (n=51), single-cell multi-omics (10X Genomics/CITE-seq) (n=17), and plasma serology (Luminex) (n=50) were performed on the cryopreserved PBMC and plasma samples and TCR sequencing (n=10), bulk RNA-seq (n=13), and multiplexed immunohistochemistry (mIHC) (n=15) were performed on the FFPE samples. This chapter focuses on results and conclusions drawn from the CyTOF data.

2.1.2 *Contributions*

Dr. Amit Jain (National Cancer Centre Singapore) consented and directed collection and processing of patient samples. N.K. and E.W.N conceived and designed the experiments. N.K. performed the experiments, data analysis, and wrote the experimental findings.

2.1.3 *Results*

Mass Cytometry peripheral immune cell profiling identifies expression of unique NPC-associated EBV-specific T cell phenotypic markers.

To comprehensively survey the unperturbed peripheral NPC immune landscape, we collected and processed peripheral blood samples from a cohort of newly diagnosed, untreated Singaporean patients (n=51) (Table 1). This cohort had a median age of 54 (range:23-89), median disease stage of III (range: I-IVB) and consisted of 43 males and 8 females. NPC is more prevalent in males compared to females, which explains the sex distribution skew of this cohort[7]. Using cryopreserved PBMCs from this cohort, we performed multiplexed major histocompatibility complex Class I (MHC-I) tetramer staining and mass cytometry experiments to deeply profile the phenotypes of the NPC peripheral immune compartment. Our primary focus was to study peripheral EBV-specific CD8 T cells to test the hypothesis that EBV-specific T cells in NPC patients have unique phenotypes that may be reflective of their role in anti-tumor immune responses.

To perform deep phenotypic profiling of antigen-specific CD8 T cells and the peripheral immune compartment as a whole, we developed and optimized a 34-marker panel of metal-labelled antibodies (Figure 1B). This included markers to detects all major $CD45^+$ immune subsets, including $CD14^+$ monocytes, $CD19^+$ B cells, $CD56^+$ NK cells, and $CD3^+$ T cells. We could also identify T cell subsets including $CD4^+$ helper T cells, $CD8^+$ cytotoxic T cells, γδ T cells, and MAIT cells ($CD161^+$, $Vα7.2^+$) (Table 2). Not only were we able to identify these broad subsets, we were also able to create cellular phenotypic profiles using markers of T cell memory (CD45RA, CD45RO, CCR7, CD27, CD127), activation/proliferation (CD38, HLA-DR, CD71, CD94, perforin, Granzyme K (GZMK), CD25, CD69, 2B4 (CD244), CD28), trafficking (CD103, CXCR5

and ITβ7) exhaustion (PD-1, TIGIT, CD39), and late differentiation/senescence (CD57, KLRG1) (Table 2).

To maximize the number of detectable antigen-specificities, we leveraged our previously developed multiplexed MHC-1 staining approach for mass cytometry[51] (see methods) to generate metal-barcoded MHC Class I tetramers for up to 55 different epitopes (Figure 1b). This tetramer panel consisted of previously identified EBV epitopes for the HLA-A types found in our patient cohort (HLA-A*02, HLA-A*24, HLA-A*11). The EBV epitopes included both lytic cycle antigens (i.e., BMLF1, BMRF1 etc.) and latent cycle antigens (i.e., LMP1, LMP2, EBNA3B etc.) (Table 3). To compare the phenotypes of EBV-specific T cells to other virus-specific T cells, we also included 29 epitopes for other viruses such as cytomegalovirus (CMV), influenza (Flu), human papillomavirus (HPV), parvovirus (Parvo), and adenovirus (AdV) (Figure 1B) (Table 3). Additionally, we included 6 non-viral tumor-associated (non-mutated) epitopes for antigens such as NY-ESO-1, MART1/MelanA, and MAGE that we assessed as an aggregate (Non-viral) (Figure 1b). In total we were able to identify 134 different CD8 antigen-specific T cell populations using of at readily detectable frequencies among the 51 NPC patients profiled (Figure 1C). This was achieved by using 8 metal-labelled streptavidins and assigning each epitope to a unique predetermined code of three metal labelled streptavidins. The metal-labelled streptavidins were then mixed in accordance with the code and then tetramerized to their corresponding epitope. A cocktail of these tetramers was then used to stain the cells (see methods). After data acquisition, we were then able to cleanly recognize numerous antigen-specific T cell populations within a single patient by first gating for CD8 T cells among the total $CD45^+$ compartment (Figure 2a), and CD8 T cells that exclusively bound to all three metal-labelled streptavidins in the predefined code to identify each antigen-specific T cell population by their corresponding epitope (Figure 1D).

Using a manual gating approach with this logic, we first gated out cells that were positive for the five streptavidins that were not present in the code, and then gated cells that were positive for the remaining three streptavidins. Only populations that met a cutoff of at least 5 cells were retained for downstream analysis.

First, we wanted to determine if EBV-specific T cells have distinct phenotypic profiles based upon whether they have lytic cycle or latent cycle antigen specificity. In total, we identified 64 detectable EBV-specific T cell populations among all NPC patients and 26 EBV-specific T cell populations amongst a cohort of 12 healthy donor control leukapheresis samples (HD) which we profiled using the same mass cytometry and multiplexed MHC tetramer staining approach. We noticed that the frequency of lytic and latent cycle antigen-specific T cell populations was similar between both the NPC patient and the HD samples, with there being slightly more latent-specific T cell populations than lytic-specific T cell populations in both groups. In addition, no one epitope was overrepresented in either the NPC EBV-specific T cell populations or the HD EBV-specific T cell populations (Figure 3a). We then compare the expression of phenotypic markers on lytic-specific T cells to latent-specific T cells in NPC patients and found that latent-specific T cells had higher expression of CCR7, CD45RO, CD27, CD28, CXCR5, TIGIT, and PD-1, while lytic-specific T cells had higher expression of CD45RA and CD94 (Figure 3B). This suggests that latent specific T cells have a more central memory-like phenotype while lytic-specific T cells appear more TEMRA-like, and even when subdivided by epitope specificity do not appear to have differential phenotypic profiles based on any effector-like phenotypes, such as exhaustion or activation (Figure 3C). Moreover, these same differential phenotypes were also observed on lytic and latent EBV-specific T cells identified in the HD samples, where CCR7 and CD45RO expression was higher on latent specific T cells as compared to lytic-specific T cells (Figure 3D).

These phenotypic differences between lytic and latent EBV-specific T cells in healthy individuals has also been noted in previous work by our group[6]. Taken together, these findings suggests that differences in lytic and latent-specific T cell phenotypes are likely intrinsic to whether the EBV-specific T cell populations target lytic or latent cycle antigens and are less reflective of whether either lytic-specific T cells or latent-specific T cells are more relevant to NPC tumor-related immune responses.

We further considered whether phenotypes of EBV-specific T cells differed based on epitope HLA-restriction. We found that EBV-specific T cells with HLA-A*11 epitopes had higher expression of CD27, CD28 and CD45RO and lower expression of perforin and CD16 (Figure 3e). However, this is confounded by the fact that HLA-A*A11 EBV epitope-specific T cell populations were overwhelmingly also latent-specific (20 of 22) as compared to the HLA-A*A2 and HLA-A*A24 EBV epitope-specific T cells, which were more evenly divided between lytic and latent-specific T cell populations (Figure 3f). What is more is that these markers are the same as those shown to be highly expressed on latent-specific T cells as compared to lytic specific T cells. Therefore, these phenotype differences are likely not based on the HLA-type of the epitope, but more attributable to whether the EBV-specific T cell population is latent- or lytic-antigen specific.

Since there were not any phenotypic differences among EBV-specific T cells based on epitope-specificity between NPC patients and healthy individuals, we broadened our analysis to compare the phenotypes of EBV-specific T cells in NPC patients to the phenotypes of antigen-specific T cell populations with other viral or non-viral antigen specificities. Here we found that compared to all other antigen-specific T cells found in NPC patients, EBV-specific T cells had significantly higher expression of activation markers (CD38 and HLA-DR), co-inhibitory markers (CD39, PD-1, TIGIT) and migratory/trafficking markers (CXCR5, CD103, ITβ7) (Figure 1e). We

also observed that other antigen-specificities also had unique phenotypic marker expression patterns, with Flu-specific T cells displaying high levels of CD127 and CD28, and CMV-specific T cells expressing high levels of CD57 and KLRG1 (Figure 1b). This suggests antigen-specific T cell populations in NPC patients have unique phenotypic profiles based on their antigen specificity, and that EBV-specific T cells have a unique activated/exhausted phenotype that also suggests that these cells have distinct patterns of T cell migration.

Next, we wanted to determine if the distinct EBV-specific phenotypic T cell profiles that we identified in NPC patients were also found on EBV-specific T cells identified in healthy individuals. To this end, we preliminarily compared the frequencies of EBV-specific T cell populations found in both NPC and HD peripheral blood samples as a frequency of total CD8 T cells and found that NPC patients overall have higher frequencies of EBV-specific T cells than healthy individuals (Figure 2c). Then, we performed manual gating on the HD EBV-specific T cell populations and compared them to the marker expression frequencies of NPC EBV-specific T cells. Strikingly, we found that NPC EBV-specific T cells and HD EBV-specific T cells had markedly different phenotypes, namely significantly increased expression of activation markers (CD71, HLA-DR, CD38), co-inhibitory/exhaustion markers (CD39, TIGIT, PD-1), migratory/trafficking markers (ITB7, CD103), and late differentiation markers (Perforin, KLRG1, CD57) (Figure 1f). We also found that NPC EBV-specific T cells had higher expression of 2B4 (CD244), CD45RO, CD56, and GZMK, and HD EBV-specific T cells had elevated expression of CCR7, CD27, CD28, and CD127 (Figure 2d). It was remarkable that the phenotypic differences we observed between NPC EBV-specific T cells and HD EBV-specific T cells largely involved the same markers that were differentially highly expressed on NPC EBV-specific T cells as compared to other antigen-specific T cells in NPC patients. Notably, we observed that in addition

to the unique activation, exhaustion, and trafficking profile of NPC EBV-specific T cells, these cells also had more late-differentiation-like phenotypes, while HD-EBV-specific T cells tended towards more central-memory-like phenotypes. Overall, we identified that compared to other antigen-specific T cells within NPC patients and EBV-specific T cells in healthy individuals, EBV-specific T cells in NPC patients have a unique NPC-associated phenotypic profile related to activation, exhaustion, and trafficking.

Unsupervised CD8 T cell Clustering reveals complex NPC-associated EBV-specific T cell phenotypes

Since we determined that NPC EBV-specific T cells had increased expression of various surface markers, we wanted to assess whether there were multiple distinct phenotypic profiles of NPC EBV-specific T cells that could be delineated based on their expression of unique combinations of these markers, i.e., complex phenotypes. To do this, we performed unsupervised Louvain clustering on the surface marker expression data from the total CD8 T cell population from all NPC patients and healthy donors and identified 22 different clusters of CD8 T cells based on their unique phenotypic marker profiles. We then used dimensionality reduction to visualize the clustered total CD8 T cells from the NPC and healthy donors and generated a uniform manifold approximation and projection (UMAP) plot that represented these data (Figure 4a). Using this method, we were not only able to categorize CD8 T cells into broad memory compartments (naïve, central memory (T_{cm}), effector memory (T_{eff}), and TEMRA, we could also identify and visualized phenotypic clusters within these compartments based on their unique combination of marker expression (i.e., cluster 14: $2B4^+$, $KLRG1^+$,$CD57^+$ T_{eff} etc.) to identify the complex phenotype of each cluster (Figure 4b) (Figure 4c).

To identify the breadth of complex phenotypes specifically found on NPC EBV-specific T cells, we also included each antigen-specific T cell population that we identified for all NPC

patients and HD in the CD8 clustering analysis and visualized the cluster distribution of all antigen-specific T cells when overlaid onto the total CD8 T cell UMAP (Figure 4d) (Figure 4f). We also calculated the frequency of cells within each antigen-specific T cell population that was assigned to each cluster and compared the cluster frequency of all antigen-specific T cell populations by antigen-specificity (Figure 4d) (Figure 4e). We found that EBV-specific T cells, as compared to all other antigen specificities, were significantly enriched in cluster 4 ($CD39^+$, $CD103^+$, HLA-DR^+, $CD71^+$, $CD38^+$, T_{eff}), cluster 7 ($TIGIT^+$, PD-1^+, $2B4^+$, $KLRG1^+$, $CD57^+$, T_{eff}), cluster 11 ($CXCR5^+$, T_{eff}), and cluster 13 ($ITB7^+$, $CD103^+$, T_{eff}) (Figure 4e). We also found that CMV-specific T cells were enriched in cluster 1 ($2B4^+$, $KLRG1^+$, TEMRA), cluster 2 ($2B4^+$, $KLRG1^+$, $CD57^+$, $CD94^+$,TEMRA), cluster 3 ($2B4^+$, $KLRG1^+$, $CD57^+$, T_{eff}/TEMRA), cluster 5 ($2B4^+$, $KLRG1^+$, $CD57^+$, TEMRA), cluster 6 ($2B4^+$, $KLRG1^+$, $CD57^+$, $Perforin^+$, TEMRA), and Flu-specific T cell were enriched in cluster 12($CD27^+$, $CD28^+$, $CD127^+$, T_{CM}) and cluster 15($CD27^+$, $CD28^+$, $CD127^+$, $ITB7^+$, T_{CM}). Overall, we elucidated that EBV-specific T cells in NPC patients have unique complex T_{eff} phenotypes, whereas CMV-specific T cells have various TEMRA complex phenotypes and Flu-specific T cells have T_{CM} complex phenotypes.

Subsequently, we explored differences in complex phenotypes between NPC EBV-specific T cells and HD EBV-specific T cells (Figure 4f). When we compared EBV-specific T cell cluster frequencies between NPC patients and HD individuals, we found that HD EBV-specific T cells were significantly enriched in cluster 1($2B4^+$, $KLRG1^+$, TEMRA), cluster 12 ($CD27^+$, $CD28^+$, $CD127^+$, T_{CM}) and cluster 15($CD27^+$, $CD28^+$, $CD127^+$, $ITB7^+$, T_{CM}) (Figure 4g). NPC EBV-specific T cells were significantly enriched in cluster 3 ($2B4^+$, $KLRG1^+$, $CD57^+$, T_{eff}/TEMRA), cluster 4($CD39^+$, $CD103^+$, HLA-DR^+, $CD71^+$, $CD38^+$, T_{eff}), cluster 7 ($TIGIT^+$, PD-1^+, $2B4^+$, $KLRG1^+$, $CD57^+$, T_{eff}), cluster 14 ($2B4^+$, $KLRG1^+$, $CD57^+$, T_{eff}), and cluster

16($CD94^+$, $CD56^+$, TEMRA) (Figure 4g). This analysis showed that EBV-specific T cells have complex phenotypes that are phenotypically distinct from EBV-specific T cells in healthy individuals, and overall HD EBV-specific T cells are largely T_{cm}-like whereas, NPC EBV-specific T cells have unique T_{eff} and TEMRA complex phenotypes not found in healthy individuals (Figure 4h).

Taken together, when comparing the complex phenotypes of NPC EBV-specific T cells to other antigen-specific T cells found in NPC patients and to EBV-specific T cells found in HD individuals, we see that NPC EBV-specific T cells are significantly enriched in cluster 4 and cluster 7 in both comparisons. Since this enrichment is only observed on EBV-specific T cells in NPC patients and not on any other antigen-specificity in NPC patients or in HD individuals, it is possible that EBV-specific T cells with these unique effector phenotypes in the peripheral blood could indicate presence of circulating T cells that are relevant to immune responses to the NPC tumor.

2.1.4 *Conclusions and Discussion*

The work in this chapter focused on using high dimensional mass cytometry phenotyping approaches to comprehensively profile the peripheral antigen-specific T cell landscape of NPC patients. Our aim was particularly to study the phenotypes of peripheral EBV-specific T cells to assess whether these T cells that target a virus that infects all tumor cells in this cancer type have any unique hallmarks of tumor-related immune interactions, and if they could provide any clues about the tumor immune microenvironment. To accomplish this, we used both manual gating and unsupervised clustering approaches to identify phenotypes that were exclusive to EBV-specific T cells in NPC patients. Both of these approaches found that these cells had uniquely high expression levels of exhaustion, activation, and migration related markers. The clustering

analysis also allowed us to identify how these markers were expressed in distinct complex phenotypes and brought into focus two clusters (7: $TIGIT^+$, $PD\text{-}1^+$ and 4: CD013, CD39, $HLA\text{-}DR^+$, $CD38^+$, $CD71^+$) with high frequencies of EBV-specific T cells in comparison to all other antigen-specificities in NPC patients and EBV-specific T cells in HD individuals.

Figure 1:

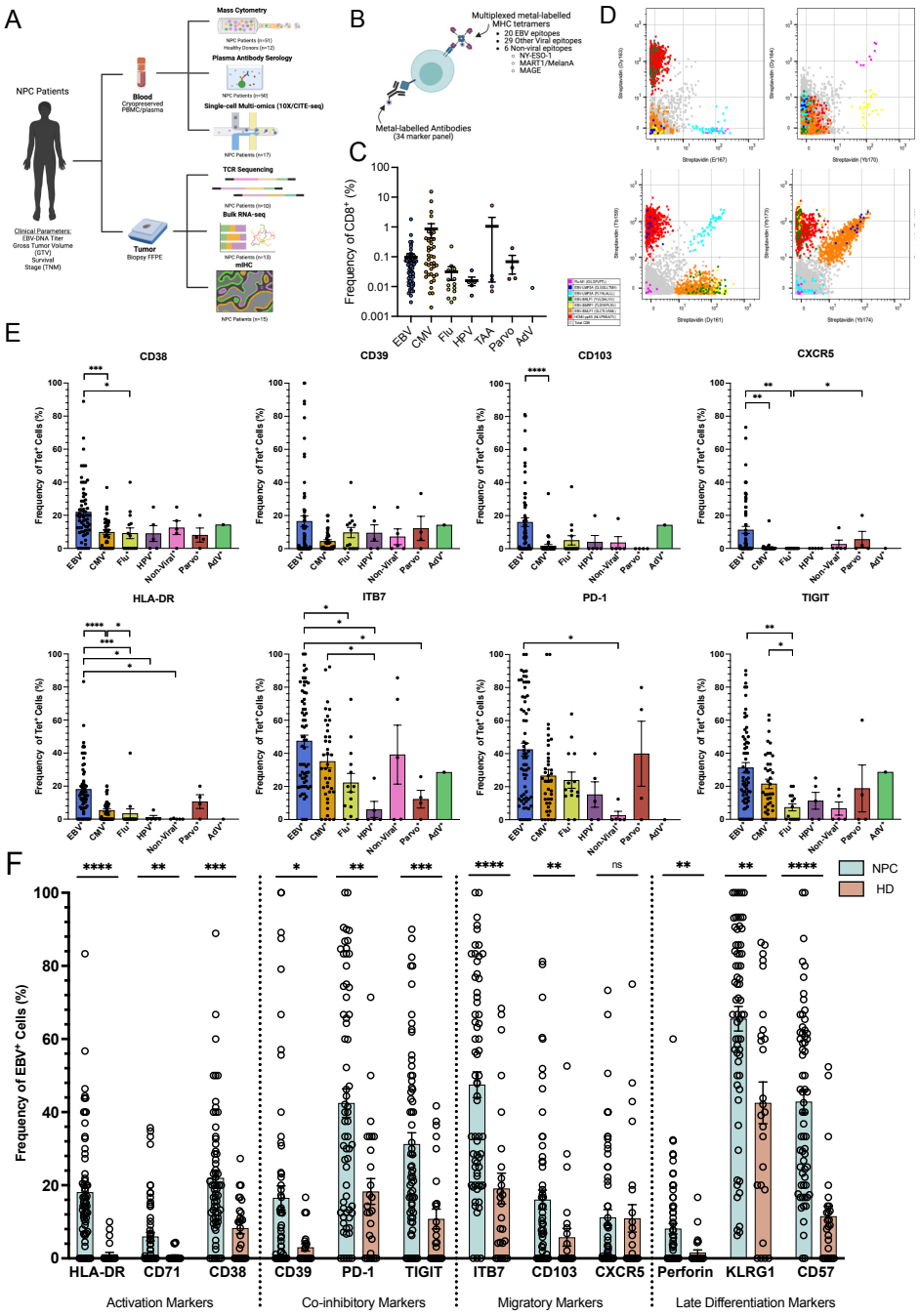

Figure 1. Mass cytometry peripheral immune cell profiling reveals unique NPC-associated EBV-specific T cell phenotypes.

a. Experimental overview. **b.** Mass cytometry (CyTOF) staining schematic of antibody and epitope panel overviews **c.** Frequencies of each antigen-specific T cell population as a percent of the total CD8 cell population in NPC patients. **d.** Representative biaxial plots showing gated antigen-specific T cell populations from a single patient (CIE_001). **e.** Selected markers that were highly expressed on EBV-specific T cells compared to other antigen specific T cell populations (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*, error Bars: Mean with SEM). **f.** Selected markers that are differentially expressed on EBV-specific T cells in NPC patients as compared to EBV-specific T cells in healthy donors. (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*, error Bars: Mean with SEM).

Figure 2:

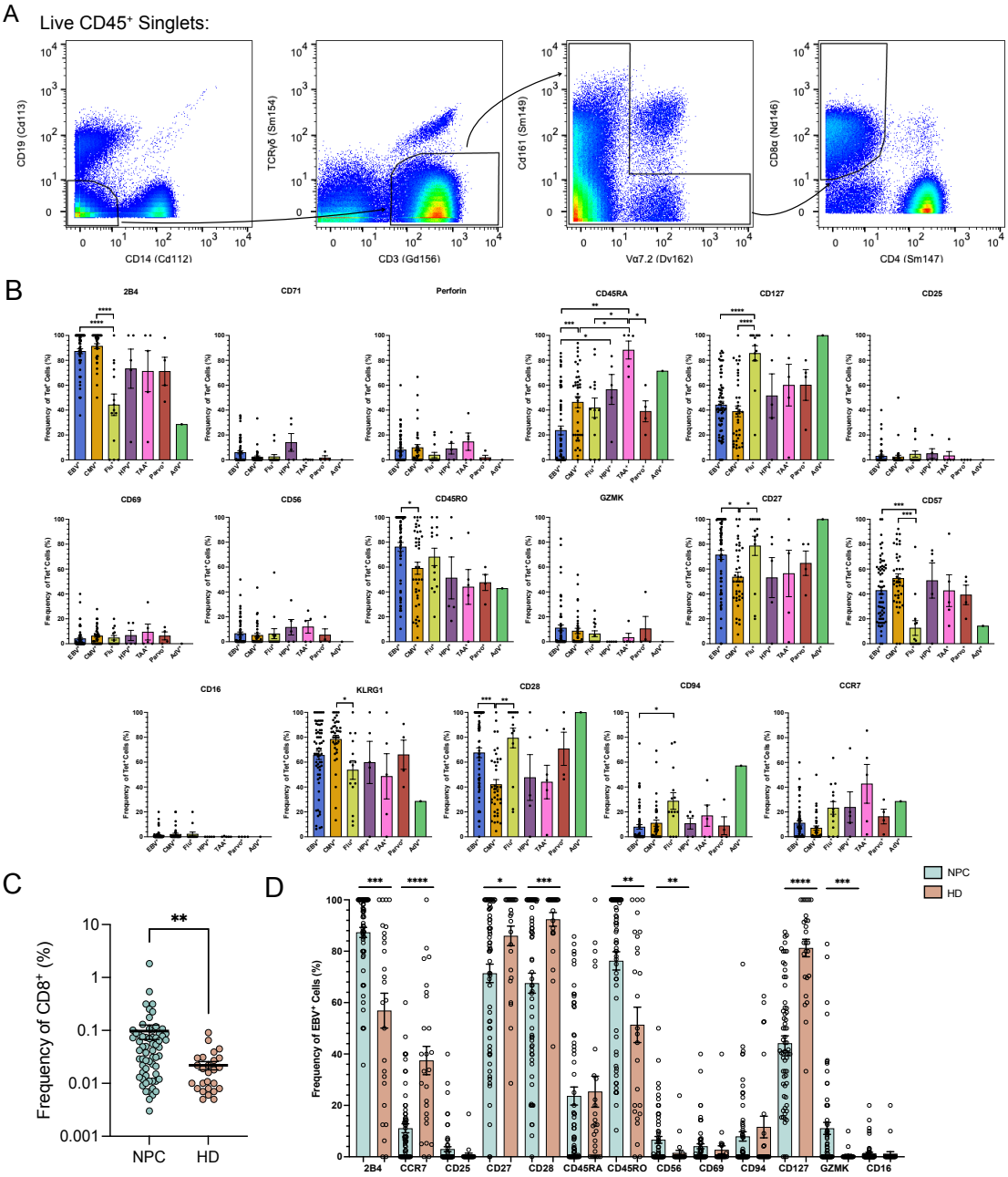

Figure 2. Comprehensive analysis of Antigen-specific T cell phenotypes. a. CD8 T cell gating strategy for total live CD45+ singlets. b. Frequencies of marker expression for individual antigen-specific T cell populations compared by specificity. (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*, error Bars: Mean with SEM). **c.** Comparison of EBV-specific T cell frequencies in total CD8 T cells between NPC patients and healthy donors. (Mann Whitney U Test, (p: <0.0001 ****, <0.0002 ***, <0.0021 **, <0.00332*, error Bars: Mean with SEM). d. Frequency of marker expression (not shown in Figure 1) on EBV-specific T cells in NPC patients compared to EBV-specific T cells in healthy donors. (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*, error Bars: Mean with SEM).

Figure 3:

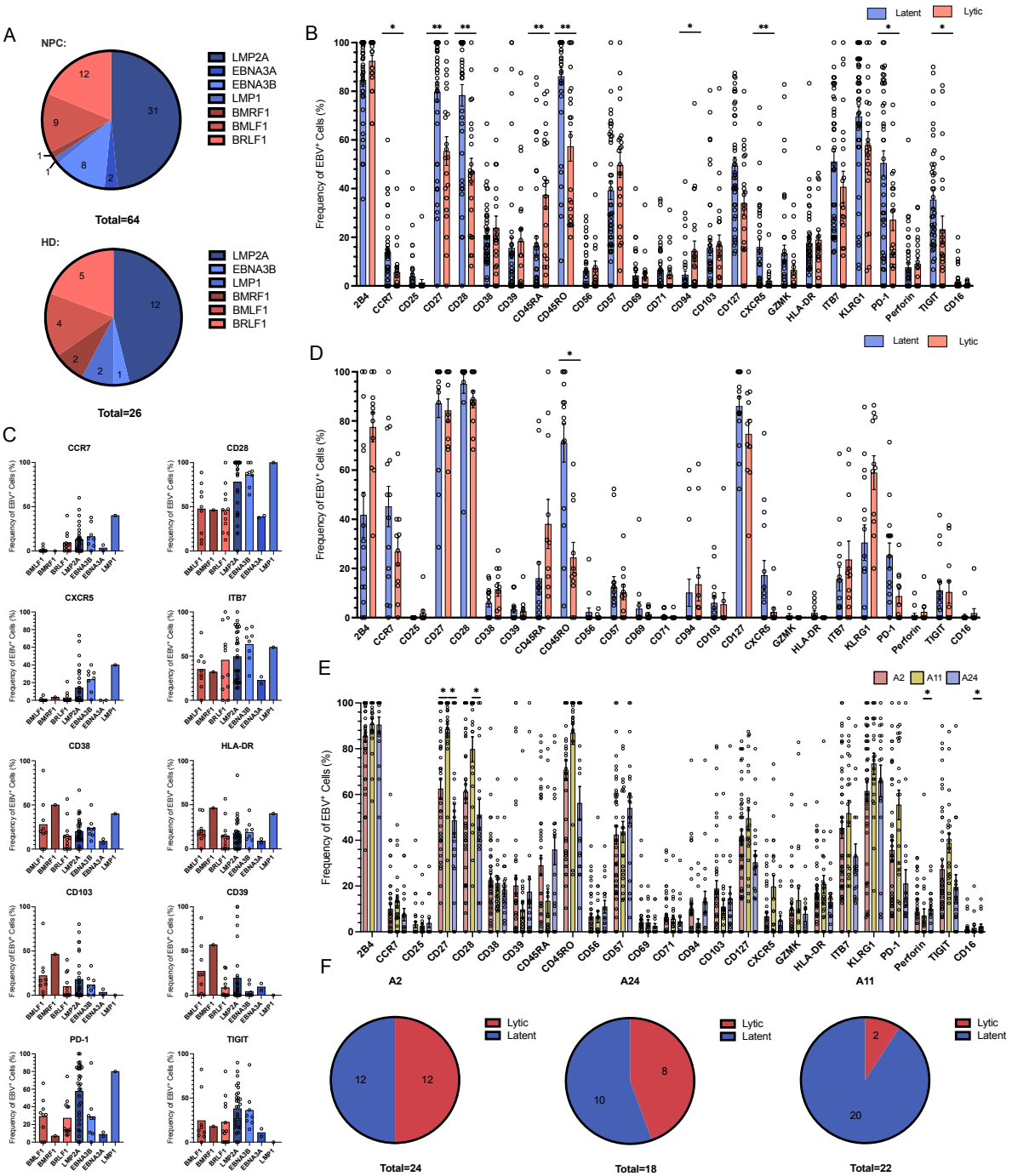

Figure 3. Comparison of Latent and lytic EBV-specific T cell phenotypes. a. Summary of lytic and latent epitope-specific T cell populations in NPC patients (n=64) and healthy donors (n=26). **b.** Comparison of latent and lytic-specific T cell populations in NPC patient peripheral blood based on phenotypic marker expression (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*), error Bars: Mean with SEM). **c.** Comparison of phenotypic marker on EBV-specific T cell populations delineated by epitope-specificity. **d.** Comparison of latent and lytic-specific T cell populations in healthy peripheral blood based on phenotypic marker expression (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*), error Bars: Mean with SEM). **e.** Comparison of phenotypic marker expression on EBV-specific T cell populations delineated by epitope HLA-type. (Mann Whitney U Test, FDR correction (p: <0.0001 ****, <0.001 ***, <0.01 **, <0.05*), error Bars: Mean with SEM). **f.** Summary of latent and lytic specific-T cell populations based on epitope HLA-A type.

Figure 4:

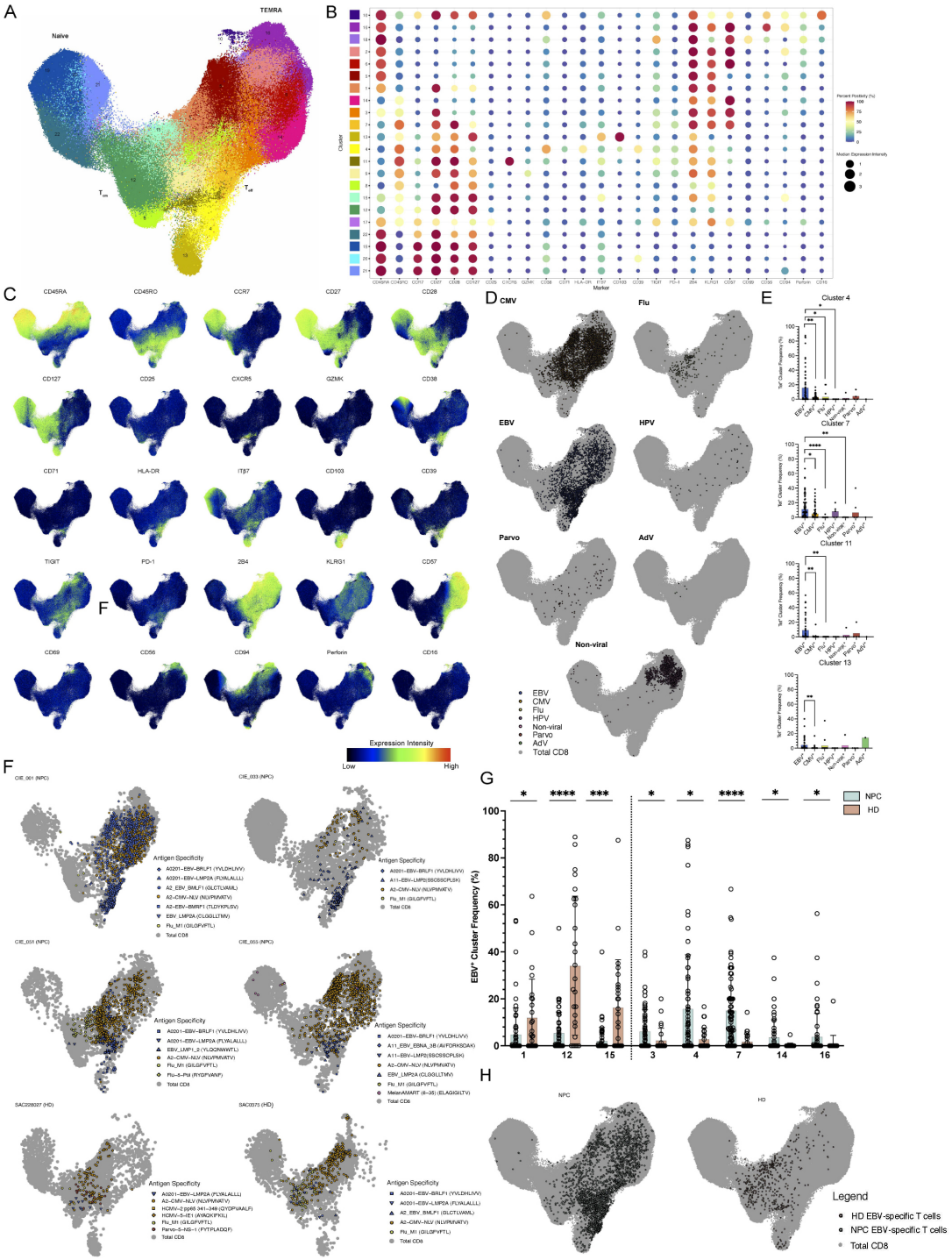

Figure 4. Unsupervised CD8 T cell Clustering Identifies EBV-specific NPC-associated complex phenotypes. **a.**UMAP of total CD8 T cells from NPC and healthy donors. **b.** Bubble plot of cluster identities based on phenotypic marker expression intensity (size) and percent positivity (color) **c.** Expression intensity of each phenotypic marker overlaid on UMAP plot **d.** Distribution of all antigen-specific T cells overlaid on UMAP plot by antigen specificity. **e.** Select comparisons of cluster frequencies of antigen-specific T cell populations between antigen specificities in NPC patients. (Mann Whitney U Test, FDR correction (p: <0.0001****, <0.001***, <0.01**,<0.05*)) **f.** Representative UMAPs for 4 NPC patients (CIE_001, CIE_003, CIE_051, CIE_055) and 2 healthy donors (SAC03075,SAC228027) and overlays of all antigen specific T cells identified for these individuals. **g.** Select cluster frequency comparisons between NPC and HD EBV-specific T cells (Mann Whitney U Test, FDR correction (p: <0.0001****, <0.001***, <0.01**,<0.05*)). **h.** All EBV-specific T cells from HD and NPC overlaid onto total CD8 T cell UMAP.

Table 1: Summary of NPC Patient Cohort Clinical Parameters and Data Collection

	Demographic Information			HLA-typing		EBV-DNA Titer Information					Diagnostic Information				Treatment Modality	Survival (months)	Data Collection Information					
ID	Sex	Race	Age (years)	A1	A2	Baseline EBV-DNA Titer	TP1 (months post baseline)	TP1 EBV-DNA Titer	TP2 (months post baseline)	TP2 EBV-DNA Titer	WHO Type	GTV	T	N	M			PBMC CyTOF	Plasma Serology	PBMC ssRNA-seq	Tumor mIHC	Tumor RNA-seq	Tumor TCR-seq
CH001	M	Chinese	47	A*02:07:01	A*02:01:01	115,050 copies/ml (5.06 log)	2	Undetectable	4	Undetectable	1	97.6	3	3	0	0	61	Y		Y			
CH002	F	Chinese	59	A*02:07:01	A*24:02:01	1,402 copies/ml (3.15 log)	2	<265 copies/ml (<2.42 log)	3	<265 copies/ml (<2.42 log)	2	102.7	4	3	0	0	62	Y		Y				
CH003	M	Chinese	55	A*02:01:01	A*26:01:01	<265 copies/ml (<2.42 log)	2	Undetectable			2	12.2	1	1	0	0	63	Y		Y	Y	Y	Y	Y
CH004	M	Chinese	53	A*33:03:01	A*33:03:01	460 copies/ml (2.66 log)	2	Undetectable			2		1	1	0	2	57	Y		Y				
CH005	M	Chinese	75	A*11:01:01	A*02:03:01	867 copies/ml (2.94 log)	2	Undetectable			2	11.1	1	0	0	2	62	Y		Y		Y	Y	Y
CH006	M	Chinese	65	A*02:03:01	A*02:03:01	447 copies/ml (2.65 log)	2	Undetectable			2	12	1	1	0	0	60	Y		Y	Y			
CH007	M	Chinese	64	A*11:01:01	A*33:03:01	Undetectable	2	<265 copies/ml (<2.42 log)	5	Undetectable	2	23.3	2	0	0	2	61	Y		Y	Y			
CH010	M	Chinese	49	A*02:03:01	A*33:03:01	2,364 copies/ml (3.37 log)	2	Undetectable			2	3.9	1	1	0	2	60	Y		Y	Y	Y	Y	Y
CH011	F	Chinese	47	A*11:01:01	A*02:03:01	16,213 copies/ml (4.21 log)	2	Undetectable			2	68.2	4	2	0	0	56	Y		Y				
CH012	M	Chinese	45	A*24:02:01	A*11:01:01	<265 copies/ml (<2.42 log)	2	Undetectable			2	13.8	2	0	0	2	57	Y		Y				
CH013	M	Chinese	50	A*11:01:01	A*24:02:01	Undetectable	2	Undetectable			1	44	3	3	0	1	47	Y		Y				
CH014	M	Chinese	63	A*24:02:01	A*24:02:01	<265 copies/ml (<2.42 log)					1		2	2	0	0	57	Y		Y	Y			
CH015	F	Malay	36	A*11:01:01	A*33:03:01	587,500 copies/ml (5.77 log)	3	>192,300 copies/ml (5.28 log)	6	>192,300 copies/ml (5.28 log)	2		4	3	1	3	10	Y		Y	Y			
CH016	M	Chinese	89	A*11:01:01	A*02:07:01	2,262 copies/ml (3.35 log)	7	Undetectable		NA (deceased)	2		4	2	0	2	9	Y		Y				
CH017	M	Chinese	68	A*02:01:01	A*29:02:01	19,208 copies/ml (4.28 log)	2	Undetectable			2	21	2	1	0	0	48	Y		Y	Y	Y	Y	
CH018	M	Chinese	48	A*11:02:01	A*24:02:01	7,360 copies/ml (3.87 log)	2	<265 copies/ml (<2.42 log)	6	<265 copies/ml (<2.42 log)	2	98.8	4	3	0	1	59	Y		Y	Y	Y	Y	Y
CH019	M	Chinese	48	A*11:01:01	A*33:03:01	<265 copies/ml (<2.42 log)	2	Undetectable			2		3	2	0	1	58	Y		Y				
CH021	M	Chinese	67	A*02:01:01	A*02:07:01	16,025 copies/ml (4.20 log)	2	Undetectable			2	73.8	3	3	0	1	34	Y		Y				
CH023	M	Chinese	51	A*11:01:01	A*02:07:01	3,163 copies/ml (3.50 log)	2	Undetectable			2	37.8	3	1	0	0	59	Y		Y		Y	Y	Y
CH024	M	Chinese	58	A*11:01:01	A*26:01:01	5,810 copies/ml (3.76 log)	2	Undetectable	5	Undetectable	3	24.9	2	1	0	0	57	Y		Y		Y	Y	
CH025	M	Chinese	70	A*24:02:01	A*02:06:01	<265 copies/ml (<2.42 log)	2	Undetectable			2	14.6	1	1	0	2	57	Y		Y				
CH026	M	Malay	50	A*11:01:01	A*33:03:01	72,750 copies/ml (4.86 log)	2	Undetectable			2	16	1	3	0	0	49	Y		Y				
CH027	M	Chinese	56	A*24:02:01	A*02:07:01	Undetectable	2	Undetectable			3		1	0	0	2	57	Y		Y		Y	Y	
CH028	M	Chinese	48	A*24:20	A*02:06:01	477 copies/ml (2.68 log)	2	Undetectable			2	84.4	4	1	0	0	50	Y		Y		Y		Y
CH029	F	Chinese	53	A*11:01:01	A*02:01:01	38,550 copies/ml (4.59 log)	2	12,235 copies/ml (4.09 log)	4	Undetectable	2	94.2	4	3	0	1	58	Y		Y				
CH030	M	Chinese	26	A*11:01:01	A*24:02:01	52,600 copies/ml (4.72 log)	2	<265 copies/ml (<2.42 log)	5	Undetectable	3	66	4	2	0	1	58	Y		Y		Y	Y	Y
CH031	M	Chinese	52	A*11:01:01	A*02:07:01	464 copies/ml (2.67 log)	2	Undetectable			2		3	2	1	0	57	Y		Y			Y	
CH032	F	Chinese	64	A*02:07:01	A*02:07:01	5,933 copies/ml (3.78 log)	2	Undetectable			2	18.2	2	1	0	0	54	Y		Y	Y			
CH033	M	Chinese	23	A*02:07:01	A*11:01:01	9,470 copies/ml (3.98 log)	2	Undetectable			2		3	2	0	0	52	Y		Y	Y			
CH034	M	Chinese	67	A*11:01:01	A*02:03:01	42,700 copies/ml (4.63 log)	2	Undetectable		NA (deceased)	2	26.1	3	3	0	0	13	Y		Y	Y			
CH035	F	Chinese	70	A*02:07:01	A*26:01:01	804 copies/ml (2.91 log)	2	<265 copies/ml (<2.42 log)			3		1	2	0	0	50	Y		Y				
CH036	M	Malay	54	A*24:10:01	A*24:02:01	9,700 copies/ml (3.99 log)	2	<265 copies/ml (<2.42 log)			2		4	2	0	0	10	Y		Y				
CH037	M	Chinese	68	A*26:01:01	A*02:01:01	<265 copies/ml (<2.42 log)	2	Undetectable			2		2	0	0	2	54	Y		Y				
CH038	M	Chinese	57	A*33:03:01	A*24:02:01	1,560 copies/ml (3.19 log)	2	536 copies/ml (2.73 log)	5	Undetectable	1		3	0	0	2	54	Y		Y	Y			
CH039	M	Chinese	67	A*11:01:01	A*33:03:01	452 copies/ml (2.65log)	2	Undetectable			2	19.8	2	1	0	2	54	Y		Y		Y	Y	
CH040	M	Chinese	74	A*24:02:01	A*02:07:01	925 copies/ml (2.97 log)	2	Undetectable			2		2	0	0	2	33	Y		Y				
CH041	M	Chinese	47	A*02:06:01	A*02:01:01	48,125 copies/ml (4.68 log)	2	Undetectable			2	22.7	1	3	0	0	52	Y		Y				
CH042	F	Vietnamese	35	A*11:01:01	A*34:01:01	925 copies/ml (2.97 log)	2	Undetectable			2	21.3	2	3	0	0	53	Y		Y		Y	Y	
CH043	M	Chinese	60	A*11:01:01	A*33:03:01	>192,300 copies/ml (>5.82 log)					2		4	3	1	3	7	Y		Y	Y	Y		Y
CH045	M	Malay	62	A*26:01:01	A*24:07	106,675 copies/ml	2	<265 copies/ml (<2.42 log)	5	Undetectable	2		3	3	1	3	54	Y		Y				
CH046	F	Chinese	56	A*02:07:01	A*24:02:01	29,625 copies/ml (4.47 log)	2	Undetectable			2	9.1	1	2	0	0	10	Y		Y				
CH047	M	Chinese	46	A*11:02:01	A*33:03:01	58,075 copies/ml (4.76 log)	5	15,910 copies/ml (4.20 log)		NA (deceased)	2		3	3	0	2	1	Y		Y				
CH049	M	Malay	42	A*02:03:01	A*24:10:01	84,550 copies/ml	2	Undetectable			1	14.3	1	2	0	0	53	Y		Y	Y			
CH050	M	Chinese	51	A*24:02:01	A*01:01:01	1,338 copies/ml (3.13 log)	4	Undetectable			1		1	2	0	2	51	Y		Y				
CH051	M	Chinese	60	A*02:01:01	A*24:02:01	1,066 copies/ml (3.03 log)	3	Undetectable			2		1	2	0	1	47	Y		Y	Y	Y		Y
CH052	M	Chinese	61	A*24:02:01	A*33:03:01	2,650 copies/ml (3.42 log)	2	Undetectable			2	10.1	2	1	0	0	50	Y		Y				
CH054	M	Chinese	33	A*02:07:01	A*33:03:01	1,838 copies/ml (3.26log)	2	Undetectable			2	26.6	3	1	0	0	52	Y		Y				
CH055	M	Malay	72	A*11:01:01	A*02:01:01	132,225 copies/ml (5.12 log)	2	<265 copies/ml (<2.42 log)		NA (deceased)	2		4	3	1	2	13	Y		Y	Y			
CH057	M	Chinese	51	A*24:02:01	A*33:03:01	<265 copies/ml (<2.42 log)	2	Undetectable			2		1	1	0	2	51	Y		Y				
CH058	F	Malay	62	A*02:11:01	A*29:01:01:0	48,900 copies/ml (4.69 log)	1	1,561 copies/ml (3.19 log)	4	Undetectable	2		3	3	0	2	49	Y		Y				
CH059	F	Chinese	49	A*01:01:01	A*02:07:01	<265 copies/ml (<2.42 log)	2	Undetectable			1	17.2	1	1	0	0	51	Y		Y		Y	Y	Y

Legend

WHO Type	
0	Type I: Keratinizing squamous cell carcinoma
1	Type II: Non-keratinizing differentiated carcinoma
2	Type III: Non-keratinizing undifferentiated carcinoma
3	Unclassified

Modality	
0	concurrent chemoradiation therapy
1	concurrent chemoradiation therapy followed by adjuvant chemotherapy
2	radiation therapy
3	chemotherapy

Table 2: Mass Cytometry Phenotyping Antibody Panel

Metal Isotope	Marker	Antibody Clone
89	CD45	HI30
102	Barcode	
104	Barcode	
106	Barcode	
108	Barcode	
110	Barcode	
112	CD14	M5E2
113	CD19	HIB19
114	CD56	NCAM16.2
115	CD57	HCD57
139	Barcode	
141	CD25	M-A251
142	HLA-DR	L243
143	ITβ7	FIB504
144	TIGIT	741182 (MAB7898)
145	CD69	FN50
146	CD8α	SK1
147	CD4	SK3
148	CD45RO	UCHL1
149	CD161	HP-3G10
150	KLRG1	13F2F12
151	CD27	LG.7F9
152	CD127	A019D5
153	CD103	B-Ly7
154	TCRγδ-PE	5A6.E9
	anti-PE	PE001
155	2B4	C1.7
156	CD3	OKT3
157	CD94	DX22
158	CD28	CD28.2
159	Streptavidin	
160	PD-1	eBioJ105
161	Streptavidin	
162	Vα7.2	3C10
163	Streptavidin	
(select batches)	LAYN	328002
164	Streptavidin	
(select batches)	Granzyme B	CLB-BB11
165	CXCR5	RF8B2
166	CD71	OKT9
167	Streptavidin	
168	CCR7	150503 (MAB197)
169	CD45RA	HI100
170	Streptavidin	
171	Granzyme K	GM6C3
172	CD39	A1
173	Streptavidin	
174	Streptavidin	
(select batches)	Granzyme A	CB9
175	Perforin	B-D48
176	CD38	HIT2
191	DNA Intercalator	
193	DNA Intercalator	
195	Cisplatin (Viability)	
209	CD16	3G8

Table 3: List of All Epitopes used for Multiplexed mIHC Class I Tetramer Staining

	EBV Peptides	Peptide Sequence	Lytic/Latent	HLA-A
1	A2_EBV_BMLF1	GLCTLVAML	lytic	A2
2	A2_EBV_BALF4	FLDKGTYTL	lytic	A2
3	A2_EBV_BMRF1	TLDYKPLSV	lytic	A2
4	A0201_EBV_EBNA3B	LLDFVRFMGV	latent	A2
5	EBV_LMP2A	CLGGLLTMV	latent	A2
6	EBV_LMP1_2	YLQQNWWTL	latent	A2
7	EBV_LMP1_1	YLLEMLWRL	latent	A2
8	A0201_EBV_BRLF1	YVLDHLIVV	lytic	A2
9	A0201_EBV_LMP2A	FLYALALLL	latent	A2
10	A2402-EBV-BRLF1	DYCNVLNKEF	lytic	A24
11	A2402-EBV-LMP-2	TYGPVFMCL	latent	A24
12	EBV-1	IYVLVMLVL	latent	A24
13	EBV-10-EBNA3A nuclear protein	RYSIFFDYM	latent	A24
14	EBV-5	TYPVLEEMF	lytic	A24
15	EBV-6-BMLF1	DYNFVKQLF	lytic	A24
16	EBV-9-EBNA-3B nuclear protein	TYSAGIVQI	latent	A24
17	A11_EBV_EBNA_3B	AVFDRKSDAK	latent	A11
18	A11_EBV_EBNA4	IVTDFSVIK	latent	A11
19	A11_EBV_BRFL1	ATIGTAMYK	lytic	A11
20	A11_EBV_LMP2	SSCSSCPLSK	latent	A11

	Viral Peptides	Peptide Sequence
1	A2_HIV_gag	SLYNTVATL
2	Flu_M1	GILGFVFTL
3	A2_CMV_NLV	NLVPMVATV
4	A2_CMV_IE1	VLEETSVML
5	CMV_IE1_2	FMDILTTCVET
6	CMV_pp65	QMWQARLTV
7	HBV_SAg	WLSLLVPFV
8	A0201_HPV_E7	YMLDLQPETT
9	A11_HPV_E6	NTLEQTVKK
10	A11_HBV_core	YVNVNMGLK
11	A11_H1N2_PB2	KVYKTYFEK
12	A1101_HIVGag_349_359	ACQGVGGPGHK
13	A1101_HIV_Env	ELYKYKVVK
14	HCMV-1 pp65 113-121	VYALPLKML
15	HCMV-2 pp65 341-349	QYDPVAALF
16	HCMV-5-IE1	AYAQKIFKIL
17	HCMV-6	FTSQYRIQGKL
18	HPV-1-E6	VYDFAFRDL
19	HPV-5-NS-1	FYTPLADQF
20	A2402-HPV-L1	FFPTPSGSI
21	A2402-HPV-E7	LCVQSTHVDI
22	Flu-8-Pol	RYGFVANF
23	Flu-1-M1	AYQKRMGVQM
24	Flu-4-H1N1-PB	FYRYGFVANF
25	Flu-5-PB-H1N1	LYASPQLEGF
26	Flu-7-NP-H1N1	PFERATVMAAF
27	A24-HBV-core 117-125	EYLVSFGVW
28	A24-HBV-pol	KYTSFPWLL
29	A24-Adenovirus 5 Hexon 37	TYFSLNNKF

	Cancer Antigen Peptides	Peptide Sequence
1	MAGEA3 112_120	KVAELVHFL
2	MAGEA-10 254_262	GLYDGMEHL
3	MAGEA1 278_286	KVLEYVIKV
4	MAGEA2 157_166	YLQLVFGIEV
5	NYESO1 157_165_9C	SLLMWITQC
6	MelanAMART 26_35	ELAGIGILTV

Chapter 3. CLINICAL CORRELATIONS BETWEEN PERIPHERAL EBV-SPECIFIC AND TOTAL CD8 T CELL PHENOTYPES AND CLINICAL PARAMETERS

3.1.1 *Introduction*

In the previous chapter, we found that EBV-specific T cells in NPC patients have unique phenotypes not found on other antigen-specificities or on EBV-specific T cells in healthy individuals. This led us to hypothesize that the phenotypes of EBV-specific T cells could be related to tumor-related immune responses. In this chapter we assess whether these T cell phenotypes are associated to the clinical status of the tumor by performing correlation analysis between the frequencies of peripheral T cells with these unique T cell phenotypes and clinical parameters, such as stage (TNM), gross tumor volume (GTV), survival, and plasma EBV-DNA titer. We also preliminarily assessed whether these frequencies are correlated to EBV plasma antibody titers measured in these patients.

3.1.2 *Contributions*

Dr. Amit Jain (National Cancer Centre Singapore) collected the clinical data for the NPC cohort and provided the cryopreserved plasma samples. Plasma serology experiments was performed by a team lead by Dr. Zhiwei Liu (National Institutes of Health). N.K. performed the data compilation, correlation analysis with mass cytometry data, and wrote the findings.

3.1.3 *Results*

Peripheral EBV-specific T cell phenotypes are correlated to NPC Clinical Parameters

First, we performed correlation analysis on the manually gated marker frequencies and cluster frequencies of EBV-specific T cells to known clinical parameters for the NPC patient cohort (Table 1). We started by looking at correlations to plasma EBV-DNA titer (log(copies/mL), which is one of the only known peripheral biomarkers for NPC that is currently used in the clinic. Thought to be akin to circulating tumor DNA, plasma EBV-DNA is not found in healthy individuals and is correlated with advanced disease. We found that frequencies of EBV-specific T cells expressing CD71, CD25, TIGIT, CD39, and ITB7 were positively correlated with EBV-DNA titer (Figure 5a). We also found that expression of CD16, CD56, and CD94 on EBV-specific T cells was negatively correlated with EBV-DNA titer (Figure 5b). When we looked to correlations between EBV cluster frequencies and EBV-DNA titer, we found that cluster 4 was significantly positively correlated with EBV-DNA titer and cluster 16 was significantly negatively correlated with EBV-DNA titer (Figure 5c). To better visualized how EBV-DNA levels are correlated to EBV-specific T cell phenotypes, we stratified patients into two groups: patients with low EBV-DNA titer and patients with high EBV-DNA titer using median EBV-DNA titer as a cutoff. We then aggregated all EBV-specific T cells from patients with low EBV-DNA titer and patients with high EBV-DNA titers and overlaid these cells onto the total CD8 T cell UMAP generated in Chapter 2 (Figure 5d). Using these overlays, it was readily appreciable that EBV-specific T cells in patients with low EBV-DNA titers were very sparse in the cluster 4 region of the UMAP, whereas EBV-specific T cells in patients with high EBV-DNA titers were found in high abundance in this region (Figure 5d).

Next, we assessed clinical correlations between gross tumor volume (GTV) and EBV-specific T cell phenotype frequencies and found that there was a positive correlation between expression of CD39, CD71, and CD38 on EBV-specific T cells and GTV (Figure 5e). In

concordance to these findings, we found that there was a significant positive correlation with GTV and cluster 4 frequencies of EBV-specific T cells. In addition, we also found that there was a slight negative correlation with GTV and cluster 13 ($CD103^+$, $ITB7^+$, T_{eff}) (Figure 5f). Similar to the analysis with EBV-DNA titer, we stratified patients based on GTV by the median patient GTV and overlaid EBV-specific T cells from these two groups on the total CD8 UMAP and noticed that patients with higher GTV had high cell densities in cluster 4 as compared to the patients with lower GTV (Figure 5g).

Finally, we evaluated correlations between stage (TNM) and peripheral EBV-specific T cell phenotypes, and we did not find any correlations between the size of the tumor (T) or metastasis (M). However, we did find that there was a weak, but significant, positive correlation between nodal involvement and EBV-specific T cell cluster 4 frequencies (Figure 5h). We also found that there were negative correlations between cluster 1 and cluster 13 EBV-specific T cell frequencies and nodal involvement (N) (Figure 5h). When we generated EBV-specific T cell overlays by grouping patients with stages N0 and N1 and patients with N2 and N3, it was evident that patients with advanced levels of nodal involvement have higher cluster 4 cell densities (Figure 5i).

Peripheral CD8 T cell phenotypes are correlated with clinical parameters and plasma antibody titers

Since we found there were distinct EBV-specific T cell phenotypes that were correlated with many clinical parameters (namely cluster 4 phenotypes), we evaluated whether the marker expression frequencies or cluster frequencies of bulk peripheral CD8 T cells agnostic of antigen-specificity would be correlated with clinical parameters. When we compared the cluster frequencies of total CD8 T cells for each patient to their EBV-DNA titer levels, we found that

there was a strong positive correlation between EBV-DNA and bulk cluster 4 frequencies (Figure 6a). When looking at correlations to stage, we found that TIGIT and CD71 expression and cluster 4 frequencies among total CD8 T cells were all significantly positively correlated with stage (T) (Figure 6b)(Figure 6c). Additionally, cluster 4 frequencies were also positively correlated with nodal involvement (N) and metastasis (M) (Figure 6d) (Figure 6e). These correlations with cluster 4 were also very evident when we stratified patients by these clinical parameters and plotted separate UMAPs for these two groups (Figure 6h).

We also preliminarily assessed correlations between survival and cluster 4 frequencies. Because survival rates were relatively high at the last known follow-up, we were not able to draw definitive correlations between survival and peripheral CD8 phenotypes. However, we first assessed differences in survival between patients with high EBV-DNA titers and low EBV-DNA titers since EBV-DNA titer is known to be correlated with advanced clinical status (Figure 6f). The survival curves of these two groups showed that patients with high EBV-DNA titer had poorer survival in the first two years post-diagnosis, but there were not enough mortality events to identify an effect long-term using these data (Log-rank Mantel-Cox test p-value: 0.2426). We then performed similar analysis for cluster 4 frequencies and survival by stratifying patients with low cluster 4 frequencies and high cluster 4 frequencies and compared the survival of these groups (Figure 6g). This showed similar trends to the EBV-DNA titer analysis where patients with high cluster 4 frequencies had lower survival rates in the first two years post-survival, but this trend is not conclusive long term (Log-rank Mantel-Cox test p-value: 0.4717) (Figure 6g).

Then, we compared the levels of plasma EBV antibody titers for patients in this cohort to known clinical parameters and cluster frequencies of total CD8 T cells. We found that Early Antigen (EA) EBV antibody titers were positively correlated with EBV-DNA titer, stage (T),

and cluster 4 frequencies (Figure 6i). Previous studies have shown concordance between EBV-DNA titer and EA antibody levels[80], so a positive correlation between cluster 4 frequencies with EA antibody titers is concordant with the positive correlation observed between EBV-DNA titer and cluster 4 frequencies.

Finally, we wanted to see if the total frequency of cluster 4 and its defining phenotypic markers were different between NPC patients and HD individuals, so we compared bulk cluster 4 frequencies and CD38, CD39, HLA-DR, CD71, and CD103 frequencies between these two groups. We found that NPC patients had significantly higher bulk CD8 cluster 4 T cell frequencies as compared to HD individuals (Figure 6j). We also found that NPC patients have higher frequencies of $CD39^+$, $CD38^+$, $HLA\text{-}DR^+$, and $CD71^+$ peripheral CD8 T cells as compared to HD peripheral T cells (Figure 6k). We, however, did not see a significant difference in CD103 expression between these two groups (Figure 6k).

3.1.4 *Summary*

The experimental findings from this chapter were focused on our analysis of correlations between clinical parameters and peripheral EBV-specific and bulk T cell phenotypes. In terms of peripheral EBV-specific T cell phenotypes, we identified that the frequency of EBV-specific cells in cluster 4 ($CD39^+$, $CD103^+$, $HLA\text{-}DR^+$, $CD71^+$, $CD38^+$, $T_{eff)}$ and frequencies of EBV-specific T cells expressing CD71, CD25, TIGIT, CD39, and ITB7 were positively correlated with patient EBV-DNA titer. We also found a negative correlation between the frequency of EBV-specific T cells in cluster 16 ($CD94^+$, $CD56^+$, TEMRA) and EBV-specific T cells expression CD16, CD56, and CD94 and EBV-DNA titer. When looking at correlations with GTV, we found a positive correlation with the expression of CD39, CD71, and CD38 and cluster 4 ($CD39^+$, $CD103^+$, $HLA\text{-}DR^+$, $CD71^+$, $CD38^+$, $T_{eff)}$ among EBV-specific T cells. We also found

a positive correlation between nodal involvement and cluster 4 EBV-specific T cell frequencies. Overall, this shows a consistent positive correlation with the frequency of EBV-specific T cells in cluster 4 and clinical parameters associated with advanced disease (EBV-DNA titer, GTV, and nodal involvement). Furthermore, we also saw similar positive correlations between bulk CD8 cluster 4 frequencies and EBV-DNA titer, stage (TNM), and EA antibody levels. We also see that compared to healthy donors, NPC patients have higher bulk cluster 4 frequencies, as well as expression of markers that define cluster 4. This indicates that agnostic of antigen-specificity, the frequency of this phenotypic population in the NPC immune periphery is consistently correlated with increased disease burden.

Figure 5:

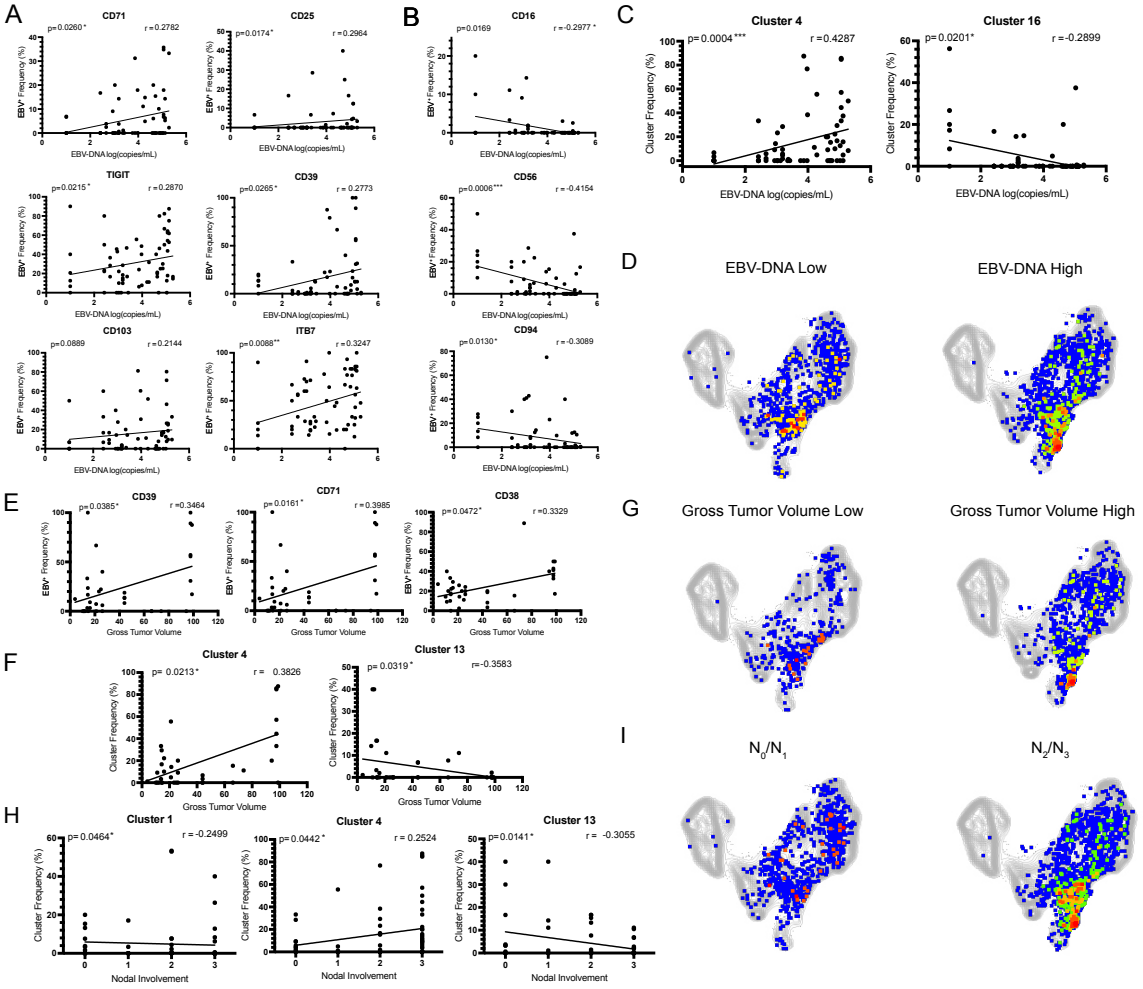

Figure 5. Peripheral EBV-specific T cell phenotypes are correlated to NPC clinical parameters.
a. Expression of phenotypic markers on EBV-specific T cells that have a positive correlations with patient EBV-DNA titer (log(copies/mL). **b.** Expression of phenotypic markers on EBV-specific T cells that have a negative correlations with patient EBV-DNA titer (log(copies/mL). **c.** EBV-specific T cell cluster frequencies that have correlations with patient EBV-DNA titer (log(copies/mL). **d.** CD8 T cell UMAP plots with all EBV-specific T cells from patients with low EBV-DNA titer (left) and high EBV-DNA titer (right), stratified by median patient EBV-DNA titer (2.22 log(copies/mL). **e.** Expression of phenotypic markers on EBV-specific T cells that have a positive correlations with gross tumor volume (GTV). **f.** EBV-specific T cell cluster frequencies that have correlations with GTV. **g.** CD8 T cell UMAP plots with all EBV-specific T cells from patients with low GTV (left) and high GTV (right), stratified by median patient GTV (median: 22). **h.** EBV-specific T cell cluster frequencies that have correlations with stage (N). **g.** CD8 T cell UMAP plots with all EBV-specific T cells from patients with low GTV (left) and high GTV (right), stratified by median patient GTV (median: 22). Statistics: Spearman correlation; r = Spearmen's rank correlation coefficient, p>0.05 (ns), ≤ 0.05 (*), ≤ 0.01 (**), ≤ 0.001 (***),≤ 0.0001 (*****).

Figure 6:

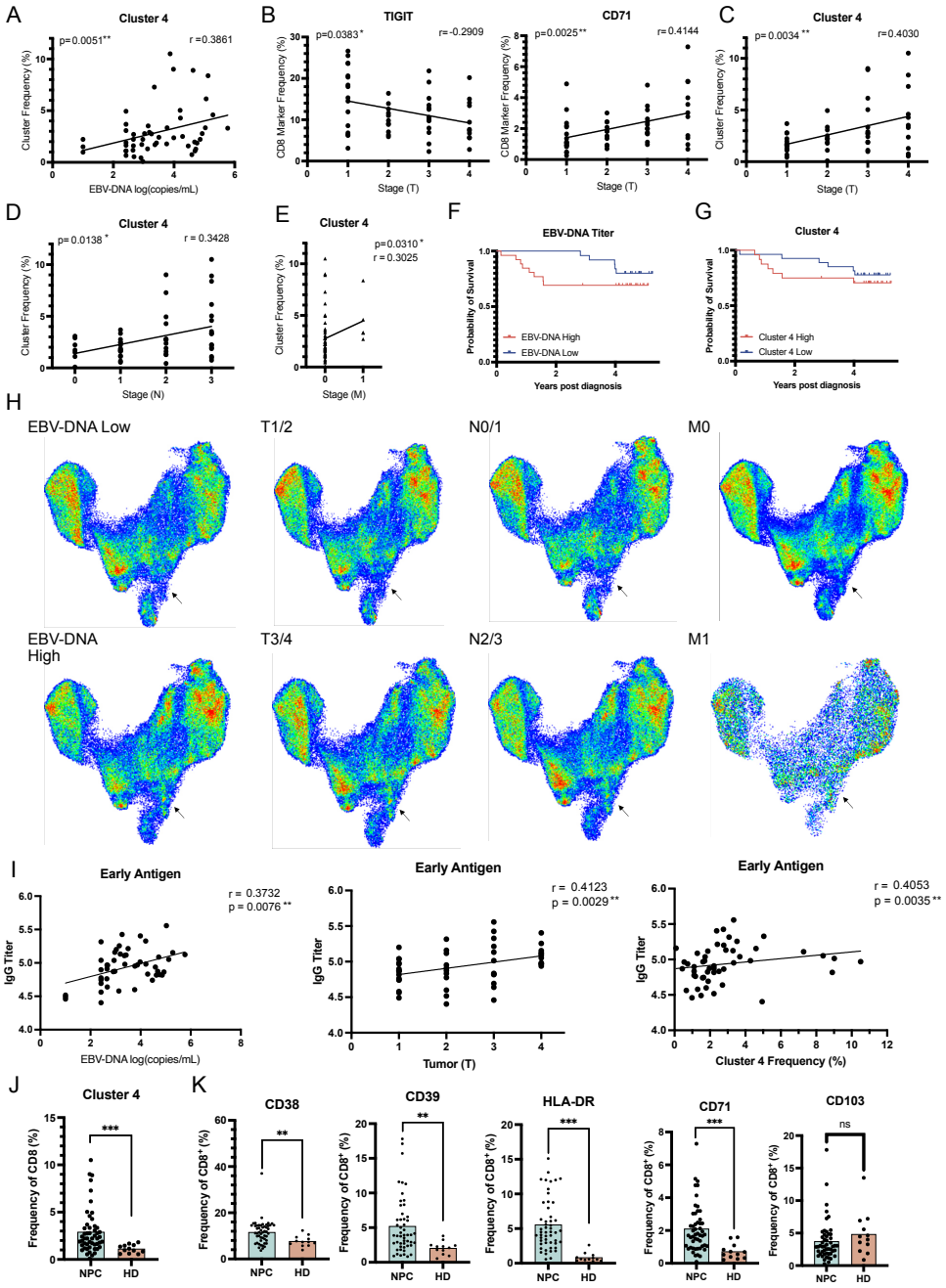

Figure 6. Peripheral Bulk CD8 T cell phenotypes are correlated clinical parameters and plasma antibody levels. **a.** Total CD8 T cell cluster frequencies that are positively correlation with plasma EBV-DNA titer. **b.** Total CD8 T cell frequencies of phenotypic markers that are correlated to stage (T). **c.** Total CD8 T cell cluster frequencies that are positively correlation with stage(T). **d.** Total CD8 T cell cluster frequencies that are positively correlation with stage(N). **e.** Total CD8 T cell cluster frequencies that are positively correlation with stage(M). **f.** Survival curves comparing patients with low EBV-DNA titer levels to patients with high EBV-DNA titer levels. **g.** Survival curves comparing patients with low bulk cluster 4 frequencies to patients with high bulk cluster 4 frequencies. **h.** Bulk CD8 T cell UMAPs stratified by patients clinical parameters. **i.** Correlations between EBV-DNA titer, stage (T), and bulk cluster 4 frequencies and EBV Early antigen plasma antibody titers. **j.** Comparison of bulk cluster 4 frequencies between NPC patients and healthy donors. **k.** Comparison of bulk frequencies between NPC patients and healthy donors of cluster 4 defining phenotypic markers.Statistics:(a, b,c,d,e,i: Spearman correlation; r = Spearman's rank correlation coefficient, p>0.05(ns), ≤ 0.05 (*), ≤ 0.01 (**), ≤ 0.001 (***),≤ 0.0001 (*****)), (j,k: Mann Whitney U Test , FCR Correction, p>0.05(ns), ≤ 0.05 (*), ≤ 0.01 (**), ≤ 0.001 (***),≤ 0.0001 (*****))

Chapter 4. ASSESSING NPC-ASSOCIATED CD8 T CELL PHENOTYPES USING SINGLE-CELL MULTI-OMICS

4.1.1 *Introduction*

In the previous chapters we identified an NPC-associated activated/exhausted T_{eff} phenotype that is exclusively expressed on NPC EBV-specific T cells and that is correlated with advanced disease. In this chapter, we wanted to further investigate the cell surface marker, gene expression, and TCR diversity of this NPC-associated phenotype to better understand any dynamics between this population and the tumor. To achieve this, we used a 5'10X/CITE-seq multi-omics approach to profile CD8 T cells from a subset of 17 NPC patients from the total 51-patient cohort (Figure 7a). In this chapter we will describe the findings from these experiments.

4.1.2 *Contributions*

Dr. Amit Jain provided the cryopreserved PBMC samples. N.K. and E.W.N designed the experiments. N.K. conducted the experiments. N.K. and Long Nguyen performed the computational data analysis. N.K. wrote the findings

4.1.3 *Results*

We deeply profiled the peripheral CD8 T cell immune landscape of 17 NPC patients in our cohort by first sorting total CD8 T cells from each patient, staining the cells with a cocktail of oligo-tagged surface marker antibodies (CITE-seq) (Table 4), and a custom oligo-tagged tetramer cocktail designed based on previously identified antigen specificities found in the mass cytometry data (Table 5) (Figure 7a). Next, we created surface marker antibody, gene

expression, and TCR libraries for each patient, and after sequencing, quality control steps, and batch correction, we generated a data set of 77,555 cells from all 17 patients (Figure 7a) (Figure 7b). We generated UMAPs for this data set using just the surface marker expression for easy comparison the CyTOF data (Figure 8a), as well as a combined UMAP considering both gene expression and surface marker expression using a weighted-nearest neighbor approach (Figure 7b). We identified 22 different clusters using Louvain clustering based on the gene expression and surface marker expression of these cells (Figure 7b) (Figure 8a).

First, we wanted to identify cells and clusters with the clinically relevant NPC-associated phenotype that we had identified in previous chapters. Therefore, we assessed each cluster based on their surface marker protein expression profiles to identify effector memory clusters that had high expression of CD39, CD103, HLA-DR and CD71 which were the hallmarks of cluster 4 identified in the CyTOF data (Figure 7c, d) (Figure 8b). We found four clusters that had high expression of one or many of these surface markers. These clusters were as follows: cluster 8 (CD39 expression only), cluster 12 ($CD103^{+}$, $CD39^{+}$), cluster 13 ($CD71^{high}$), and cluster 14 ($CD38^{+}$, $HLA\text{-}DR^{+}$) (Figure 7e, f) (Figure 8e).

Next, we wanted to leverage our large panel of surface marker antibodies to identify any other surface markers that delineate these NPC-associated clusters (Table 4). We first performed differential marker analysis on each of these clusters as compared to all other clusters to see which markers were most highly expressed by each of these four clusters (Figure 8f). We then took all the differentially expressed markers from all four of these clusters and clustered their marker expression by phenotypic cluster to see what the overall program of expression was for each of these clusters (Figure 7g). We found that compared to the other three NPC-associated clusters, cluster 8 (CD39 only) had the highest expression of CD28, CD25, CD27, CD127, and

CD73, and overall had less of an effector-like phenotype. We found that cluster 12 ($CD103^{+}$, $CD39^{+}$) had the highest expression of CD103, CD45RO, CD45RB, CCR6, and CXCR3, and had the most trafficking/ migration related phenotype. Cluster 13 ($CD71^{high}$) had the highest expression of activation markers such as CD70, HLA-DR, CD38, 2B4, and Cluster 14 ($CD38^{+}$, $HLA\text{-}DR^{+}$) in addition to also having high expression of activation markers, also had the highest expression of PD-1 and CD45RA suggesting a more late-differentiation/exhaustion-like phenotype. Overall, here we found that these NPC-associated clusters had four different phenotypes based on surface marker expression cluster 8 (central-memory like), cluster 12 (migratory), cluster 13 (proliferating), cluster 14 (activated/exhausted).

Subsequently, we wanted to assess differences in gene expression between these four clusters using differential gene expression analysis (DEG) since it was clear that surface marker expression was not the only factor driving cluster differences (Figure 8 g, h). Particularly, it was clear the surface marker expression was driving a large amount of the differential profile of cluster 8, but that it was not weighted very heavily in terms of differentiating cluster 13 (Figure 8h). We found only a small number of differentially upregulated genes for cluster 8, but these included *ENTPD1* (encodes CD39), and we found downregulation of genes such as *GZMH*, *GNLY*, *NKG7*, which is consistent with lower levels of cytotoxicity. When examining cluster 12, we found upregulation of *ZNF683* (encodes transcription factor HOBIT), *ITGAE* (encodes CD103), *CD69* (encode CD69), which are all genes related to tissue residency. Cluster 13 had upregulation of cell cycle and proliferation related genes such as *MKI67* and Cluster 14 had high expression of activation genes: *CD38*, *HLA-DRA*, *GZMA*, and *CD74* (Figure 7h). This furthers the surface marker findings that these four clusters also have very distinct gene expression profiles.

Because of the many hallmark genes of tissue residency that were expressed in cluster 12, we wanted to explore the hypothesis that these cells may have a tissue residency origin by using a known gene expression signature to score the cluster by their tissue residency (Trm) gene score[66] (Figure 7i). We found the NPC-associated clusters had higher Trm gene scores than most other clusters, which included cluster 12 (Figure 7j).

Next, we wanted to assess the TCR clonality of these clusters, so we calculated the clone size of the TCRs in these clusters to see which clusters had the highest clonal diversity (Figure 7k). We found that clusters 13 and 14 had the largest clones while clusters 12 and 8 had fewer hyper-expanded clones (Figure 7k). This is consistent with our finding that cluster 13 and cluster 14 have proliferating and activated phenotypes, respectively, and they have high levels of clonal expansion. Next, to better understand NPC-associated cluster clonal dynamics, we examined the clonal sharing between these clusters and found that cluster 8 was the most private cluster with the fewest TCRs shared with the other three clusters (Figure 7l). Interestingly, we found the highest degree of clonal sharing between cluster 13 and cluster 14 suggesting high likelihood of shared TCR dynamics. We also observe modest levels of clonal sharing between cluster 12 and cluster 14, but very little clonal sharing with the other two clusters (Figure 7l). This suggests that cells in cluster 13 and cluster 14 may be expanding due to related priming events.

Then, we examined the cluster frequencies of the antigen-specific T cells we identified using oligo-tagged MHC Class I tetramers (Figure 9a, b). We found EBV-specific T cells of 8 different epitope specificities, CMV-specific T cells of two epitope specificities, and Flu-specific T cells of two epitope specificities. Although the antigen-specific T cells we identified were not as numerous as those found in the mass cytometry dataset, we found high levels of concordance between the phenotypes of these cells based on their antigen specificity. As in the mass

cytometry data, we largely saw that CMV-specific T cells had more late differentiation-like phenotypes and Flu-specific T cells had a central memory-like phenotype (Figure 9a, b). EBV-specific T cells largely had effector and central memory-like phenotypes, and overall, we found that cluster 12 ($CD103^+$, $CD39^+$ had the highest number and percent of EBV-specific T cells (although many of these EBV-specific T cells were largely from one epitope specificity) (Figure 9 d,e).

Lastly, we wanted to use the phenotypes of the NPC-associated clusters that we identified in the single cell multi-omics analysis to assess clinical correlations between these subpopulations of NPC-associated cluster 4 ($CD39^+$, $CD103^+$, HLA-DR^+, $CD71^+$, $CD38^+$, T_{eff}) identified in the mass cytometry data. To do this, we used the main differentially expressed surface markers from each of the 4 single cell clusters to create a 1-2 marker phenotypic shorthand that could easily be used to gate these cells from the mass cytometry data. In the mass cytometry data, we gated $CD39^+$, $CD28^+$ cells as a proxy for 10X cluster 8-like cells, $CD39^+$, $CD103^+$ as a proxy for 10X cluster 12-like cells, $CD71^+$ as a proxy for 10X cluster 13-like cells, and HLA-DR^+, $CD38^+$ as proxy for 10X cluster 14-like cells. To assess whether there was enrichment of these phenotypes within CyTOF cluster 4, we manually gated for these phenotypes among all CD8 T cells used in the clustering analysis. We found that frequencies of these phenotypes were the highest among cells in cluster 4 with 86.40% of $CD39^+$, $CD103^+$ cells, 50.50% of HLA-DR^+, $CD38^+$ cells, 49.00% of $CD71^+$ cells, and 43.9% of $CD39^+$, $CD28^+$ cells residing in cluster 4. We then used these surface markers to gate for EBV-specific T cell population frequencies with these phenotypes as well as bulk CD8 frequencies of these phenotypes. We then identified correlations between the frequencies of these sub-population phenotypes and clinical parameters to identify which subpopulations could be driving the

positive clinical correlations that we see between mass cytometry cluster 4 frequencies and clinical parameters (Table 6). Readily, it was clear that there was no correlation between cluster 8-like cells ($CD39^+$, $CD28^+$) and any clinical parameters. However, there was a distinct dichotomy between which of the other three clusters were correlated to clinical outcomes based on whether their expression frequencies were measured on EBV-specific T cells or bulk CD8 T cells. We found that cluster 12-like cells ($CD39^+$, $CD103^+$) were strongly, positively correlated with EBV-DNA titer, GTV, and stage (N) when expressed on EBV-specific T cells, however there were no correlations with the total bulk CD8 frequencies of these cells and these same clinical parameters, with only a correlation with metastasis approaching statistical significance ($p=0.0757$). On the other hand, we saw that frequencies of cluster 13-like cells ($CD71^+$) were positively correlated with EBV-DNA titer when expressed on EBV-specific T cells, GTV both when expressed on EBV-specific T cells and total CD8 T cells, and stage (T) when expressed only on total CD8 T cells. We also saw a positive correlation between the frequencies of cluster 14-like cells ($CD38^+$, HLA-DR^+) and GTV when expressed on EBV-specific T cells and stage (T) when expressed by total CD8 T cells, suggesting that frequencies of this population are connected to the size of the tumor. Overall, it was notable to observe which of these sub-population were driving correlations with clinical parameters, particularly based on whether these phenotypes were only expressed by EBV-specific T cells or in the total peripheral CD8 T cell compartment. (Table 6)

4.1.4 *Summary*

In this chapter, we used single cell multi-omics to better understand the clinically relevant, peripheral NPC-associated CD8 T cell phenotypes that we had identified in previous chapters. We identified four clusters of cells with the same surface marker profile as that

identified by mass cytometry and found that they each had different surface marker and gene expression profiles including central memory-like (cluster 8), migratory (cluster 12), proliferating (cluster 13), and activated (cluster 14). When we examine these cells resembling the CyTOF cluster 4 using a multi-omics approach, we see that they actually have diverging biological roles. We also can appreciate how interrelated these subpopulations might be based on clonal expansion and TCR sharing since we see more clonal expansion and sharing among the proliferating and activated clusters as compared to the migratory and central memory-like clusters. Then, we were also able to validate the finding from the mass cytometry data that EBV-specific T cells are enriched in these clusters in this modality, and we saw the highest enrichments in the $CD103^+$, $CD39^+$ migratory cluster 12. Lastly, we used the phenotypes of the NPC-associated clusters that we found in the single-cell multi-omics data to assess how these subpopulations contributed to correlations with clinical parameters. Taken together, these experiments gave us deeper insights into the cellular biology of these peripheral NPC-associated CD8 T cells and provided validation and further insights for findings in previous chapters.

Figure 7:

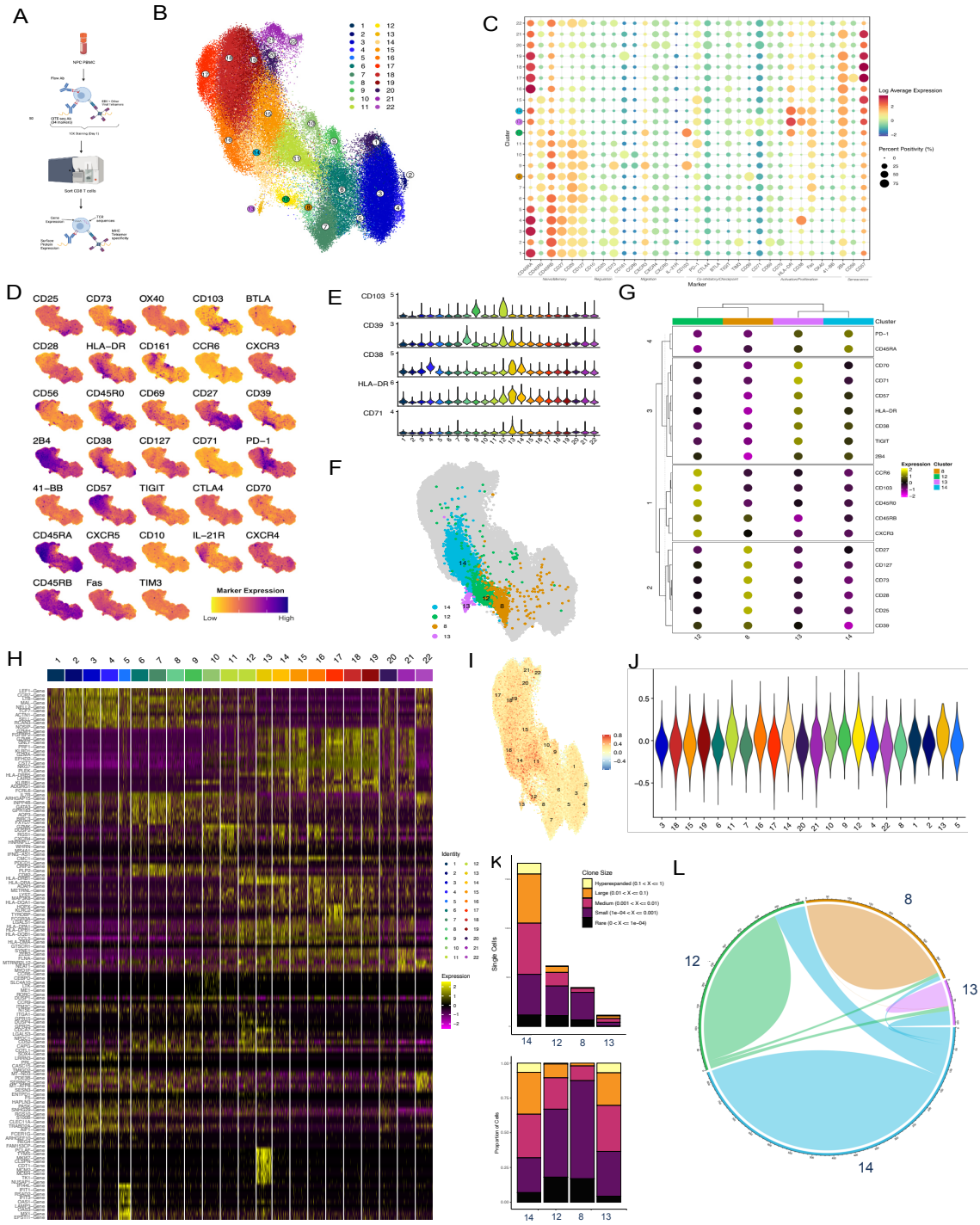

Figure 7. Single Multi-omics Analysis of NPC Peripheral CD8 T cells Identifies Surface Marker, Gene Expression, and TCR Diversity Profiles of NPC-associated phenotypic clusters.
a. Experimental overview. **b.** wnnUMAP of total CD8 T cells in dataset labelled by Louvain clusters with NPC-associated phenotypic clusters highlighted. **c.** Dot plot of surface marker expression by cluster, (color = Log average marker expression, dot size = percent positivity) **d.** Representative wnnUMAPs colored by surface marker expression intensity. **e.** Violin plots show expression of NPC-associated phenotypic markers by cluster. **f.** wnnUMAP with NPC-associated clusters highlighted. **g.** Clustered differentially expressed surface markers for NPC-associated clusters (k-means = 4). **g.** Differentially expressed genes by cluster. **h.** wnnUMAP color by tissue residency gene score (low = blue, high = red). **i.** Violin plots showing tissue residency gene scores by cluster. **j.** TCR clonal diversity stratified by clone size between NPC-associated phenotypic clusters. **k.** Chord diagram showing clonal sharing between NPC-associated phenotypic clusters.

Figure 8:

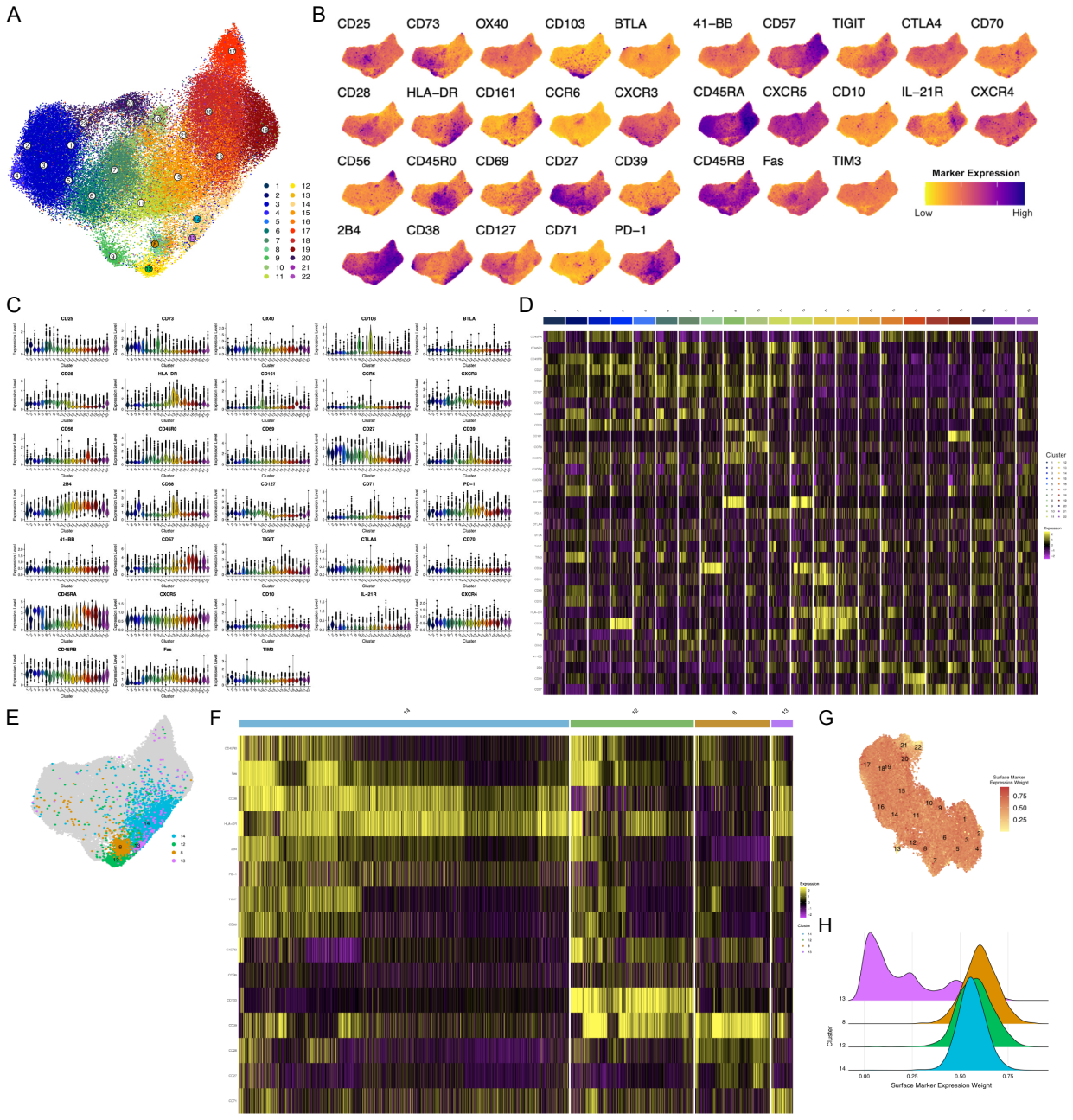

Figure 8. Surface Marker Expression Analysis of Single-cell Multi-omics Clusters. a) Louvain clustering of single-cell multi-omics data only using surface marker expression data (CITE-seq antibody expression). **b)** Surface marker expression marker overlays of each marker on the surface marker-only UMAP. (low expression= yellow, high expression=high). **c)** Violin plots of surface marker expression by cluster, box plots represent media value, interquartile range and outlier values. **d)** Heatmap of surface marker expression of all markers by cluster (low expression = purple, intermediate expression = black, high expression yellow). **e)** Surface marker expression-only UMAP with NPC-associated phenotypic clusters highlighted. **f)** Heatmap of differentially expressed surface markers of NPC-associated phenotypic clusters, with all cells from NPC-associated phenotypic clusters shown (low expression = purple, intermediate expression = black, high expression yellow). Differentially expressed markers identified by comparing each cluster to all other clusters using Wilcoxon Rank Sum Test and logFC threshold=0.25. **g)** wnnUMAP showing surface marker expression weight of each cells (low weight = yellow, high weight = red). **h)** Ridge plot of surface marker expression weight of NPC-associated phenotypic clusters.

Figure 9

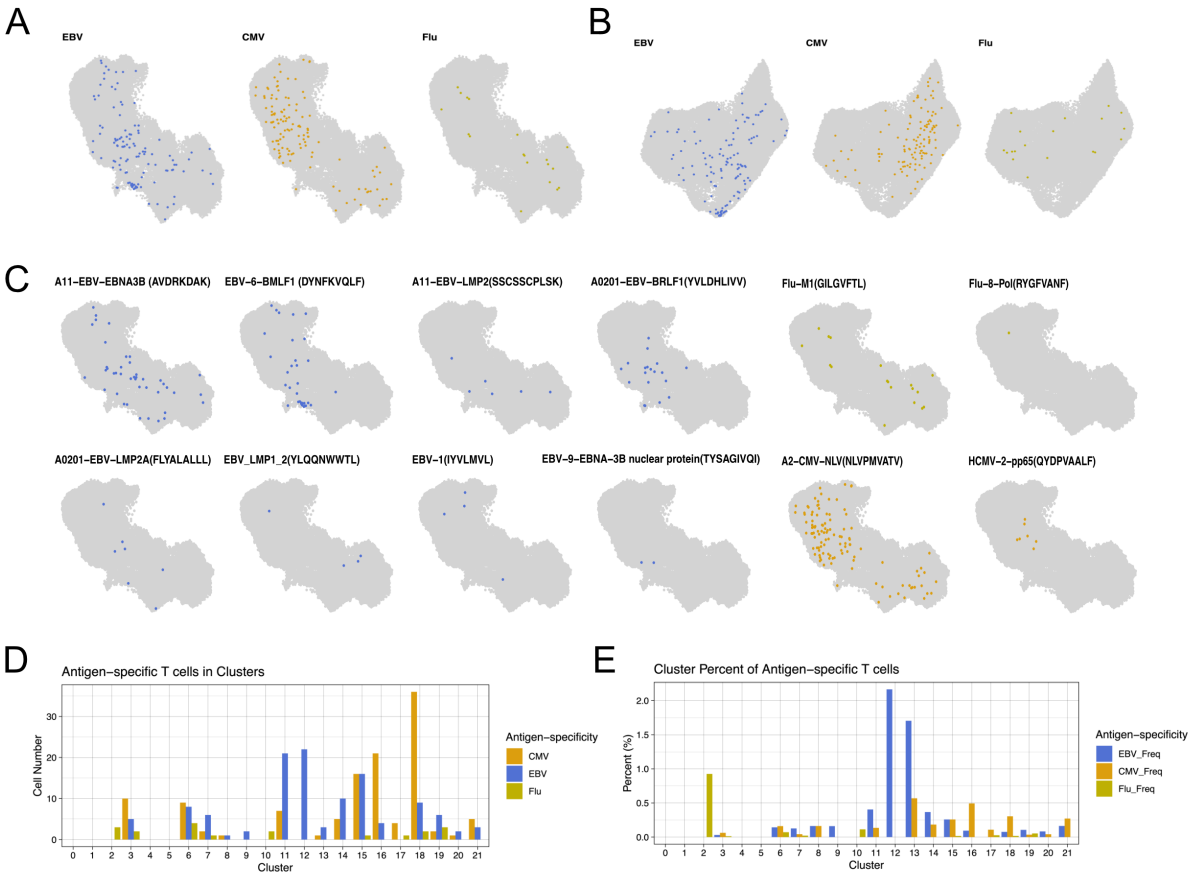

Figure 9. Multi-omics Antigen-specific T cell profiling using oligo-tagged MHC Class I Tetramers. **a.** All antigen-specific T cells identified in 10/CITE-seq dataset using oligo-tagged MHC Class I Tetramers overlaid onto wnnUMAP by antigen specificity. **b.** All antigens-specific T cells overlaid onto surface marker expression-only UMAP y antigen specificity. **c.** Antigen-specific T cell populations each individually overlaid onto the wnnUMAP. **d.** Number of antigen-specific T cells found in each cluster by antigen specificity. **e.** Percent of cluster of each antigen-specificity.

Table 4: CITE-seq panel

	Total Seq #	Marker	Clone
1	C5146	CD197 (CCR7)	MAB197
2	C5148	KLRG1	13F12F2
3	C0085	CD25	BC96
4	C0577	CD73	AD2
5	C0158	CD134 (OX40)	BerACT35
6	C0145	CD103	Ber-ACT8
7	C0170	CD272 (BTLA)	MIH26
8	C0179	CX3CR1	K0124E1
9	C0386	CD28	CD28.2
10	C0159	HLA-DR	L243
11	C0149	CD161	HP-3610
12	C0143	CD196 (CCR6)	G034E3
13	C0072	CD4	RPA-T4
14	C0140	CD183 (CXCR3)	G025H7
15	C0080	CD8a	RPA-T8
16	C0084	CD56	QA17A16
17	C0087	CD45R0	UCHL1
18	C0146	CD69	FN50
19	C0184	CD27	O323
20	C0176	CD39	A1
21	C0189	CD244 (2B4)	C1.7
22	C0389	CD38	HIT2
23	C0390	CD127	A019D5
24	C0394	CD71	CY164
25	C0088	CD279 (PD-1)	EH12.2H1
26	C0355	CD137(41-BB)	4B4-1
27	C0168	CD57	QA17A04
28	C0089	TIGIT	A15153G
29	C0151	CD152 (CTLA4)	BN13
30	C0027	CD70	113-16
31	C0063	CD45RA	HI100
32	C0144	CD185 (CXCR5)	J252D4
34	C0100	CD20	2H7
35	C0062	CD10	HI10a
36	C0985	IL-21R	17A12
37	C0366	CD184 (CXCR4)	12G5
38	C0844	CD45RB	MEM-55
39	C0156	CD95 (Fas)	DX2
40	C0169	CD366 (TIM3)	F38-2E2
41	C0141	CD195 (CCR5)	J418F1
42	C0397	CD193 (CCR3)	5E8
43	C0083	CD16	3G8
46	C0396	CD26	BA5b
47	C0830	CD319 (CRACC)	162.1
48	C0152	CD223 (Lag-3)	11C3C65
49	C0815	CCR10	6588-5
50	C5154	CLA	HECA-452

Table 5: 10X/CITE-seq. Tetramer panel

20210413 CIE_001			
PE			
Target	TotalSeq Code	Peptide	
A0201-EBV-BRLF1	TotalSeqC0953	YVLDHLIVV	
A2-EBV-BMRF1	TotalSeqC0952	TLDYKPVSV	
A2-CMV-NLV	TotalSeqC0951	NLVPMVATV	
APC			
Target	TotalSeq Code	Peptide	
EBV-LMP2A	TotalSeqC0956	CLGGLLTMV	
A2-EBV-BMLF1	TotalSeqC0957	GLCTLVAML	
A0201-EBV-LMP2A	TotalSeqC0958	FLYALALLL	
20210721 CIE_018			
PE			
Target	TotalSeq Code	Peptide	
HCMV-2 pp65	TotalSeqC0951	QYDPVAALF	
APC			
Target	TotalSeq Code	Peptide	
EBV-6-BMLF1	TotalSeqC0956	DYNFKVQLF	
A11-EBV-EBNA3B	TotalSeqC0958	AVFDRKSDAK	
20211018 CIE_018			
PE			
Target	TotalSeq Code	Peptide	
HCMV-2 pp65	TotalSeqC0951	QYDPVAALF	
APC			
Target	TotalSeq Code	Peptide	
EBV-6-BMLF1	TotalSeqC0956	DYNFKVQLF	
A11-EBV-EBNA3B	TotalSeqC0958	AVFDRKSDAK	
20211018 CIE_033			
PE			
Target	TotalSeq Code	Peptide	
A2-CMV-NLV	TotalSeqC0951	NLVPMVATV	
APC			
Target	TotalSeq Code	Peptide	
A11-EBV-LMP2	TotalSeqC0956	SSCSSPLSK	
20211202 CIE_038			
Target	TotalSeq Code 1 PE	TotalSeqCode 2 APC	Peptide
EBV-9-EBNA-3B	TotalSeqC0951	TotalSeqC0956	TYSAGIVQI
EBV-1	TotalSeqC0952	TotalSeqC0957	IYVLVMVL
EBV-10-EBNA3A	TotalSeqC0953	TotalSeqC0958	RYSIFFDYM
EBV-5	TotalSeqC0954	TotalSeqC0977	TYPVLEEMF
EBV-6-BMLF1	TotalSeqC0963	TotalSeqC0978	DYNFVKQLF
20220223 CIE_014,CIE_015,CIE_049	no tetramers		
20220615 Hashtagged			
Target	TotalSeq Code 1	TotalSeqCode 2	Peptide
A0201-EBV-LMP-2A	TotalSeqC0951	TotalSeqC0956	FLYALALLL
A2-EBV-BMLRF1	TotalSeqC0952	TotalSeqC0957	GLCTLVAML
EBV-LMP2A	TotalSeqC0953	TotalSeqC0958	CLGGLLTMV
A11EBVLMP2	TotalSeqC0954	TotalSeqC0977	SSCSSCPLSK
A11-EBV(EBNA3B)	TotalSeqC0955	TotalSeqC0978	AVFDRKSDAK
A0201EBVBRLF1	TotalSeqC0956	TotalSeqC0979	YVLDHLIVV
20220715 CIE_018			
Target	TotalSeq Code 1 PE	TotalSeqCode 2 APC	Peptide
EBV-6-BMLF1	TotalSeqC0951	TotalSeqC0956	DYNFKVQLF
HCMV-2 pp65	TotalSeqC0952	TotalSeqC0957	QYDPVAALF
A11-EBV-EBNA3B	TotalSeqC0953	TotalSeqC0958	AVFDRKSDAK
20220715 CIE_006			
Target	TotalSeq Code 1 PE	TotalSeqCode 2	Peptide
A2-CMV-NLV	TotalSeqC0966	TotalSeqC0979	NLVPMVATV

Table 6:

Correlations between frequency of CyTOF EBV-specific T cells with Cluster 4-phenotypes and Clinical Parameters

	% $CD38^+$,$HLA\text{-}DR^+$	$CD39^+$,$CD103^+$	$CD39^+$,$CD28^+$	$CD71^+$	Cluster 4 Frequency of EBV^+ cells
EBV_DNA titer	0.2099	0.0005	0.1345	0.0260	0.0004
GTV	0.0074	0.0013	0.211	0.0161	0.0213
Stage (T)	0.1151	0.0914	0.682	0.0678	0.0830
Stage (N)	0.2633	0.0148	0.1774	0.0773	0.0442
Stage (M)	0.5535	0.9479	0.7366	0.1638	0.8851

Correlations between frequency of CyTOF Bulk CD8 T cells with Cluster 4-phenotypes and Clinical Parameters

	$CD38^+$,$HLA\text{-}DR^+$	$CD39^+$,$CD103^+$	$CD39^+$,$CD28^+$	$CD71^+$	Cluster 4 Frequency of Bulk CD8
EBV_DNA titer	0.1796	0.2588	0.5765	0.2720	0.0154
GTV	0.1547	0.2268	0.1479	0.0100	0.0934
Stage (T)	0.0130	0.1718	0.2631	0.0025	0.0034
Stage (N)	0.0927	0.1839	0.5909	0.0590	0.0138
Stage (M)	0.1756	0.0757	0.5738	0.2323	0.0310

Table 6. P-values of Spearman correlation between Clinical Parameters and Gated frequencies of peripheral EBV-specific T cells (top) and Bulk CD8 T cells (bottom) with NPC-associated clusters phenotypes in Mass Cytometry Data. Spearman correlation between mass cytometry EBV-specific and bulk cluster 4 frequencies and clinical parameters is includes as a comparison. (p-values <0.05 highlighted in green).

Chapter 5. EVALUATING CONNECTIONS BETWEEN NPC TUMOR AND IMMUNE PERIPHERY

5.1.1 *Introduction*

In this last chapter, we aim to better understand connections between the NPC tumor microenvironment, the immune periphery, and the clinical disease status of the patient (Figure 1a). Directly studying the NPC tumor microenvironment is often difficult since these tumors are not often resected and therefore, sizeable tumor samples are very rarely available. In this chapter, we rely upon data from the tumor FFPE punch biopsies which are used by pathologists to diagnose and monitor the cancer. In previous chapters we found a unique peripheral NPC-associated CD8 T cell phenotype that is correlated with advanced disease and further dissect the biology of these cells using single-multi-omics into four immunologically distinct subpopulations within this cluster based on their surface marker, gene expression, and TCR clonality profiles: central memory-like, migratory, proliferating, and activated. In this chapter, we aim to use bulk TCR sequencing of tumor FFPE biopsies to compare to the peripheral CD8 TCR dataset collected in the previous chapter to assess which peripheral CD8 T cell clusters had the highest degree of clonal sharing with the tumor. In addition, we used mIHC staining and tumor bulk gene expression data to assess the tumor microenvironment of each tumor to see if characteristics of the tumor have any correlation to the characteristics of the immune periphery.

5.1.2 *Contributions*

Dr. Joe Yoeng (A*STAR, Singapore) and his team collected the bulk TCR data, perform the mIHC staining, and scored the mIHC data. Dr. Melvin Chua (National Cancer Centre,

Singapore) and his team collected the tumor bulk RNA-seq data. N.K. compiled the data and N.K. and Hugh MacMillan analyzed the data. N.K. wrote the findings.

5.1.3 *Results*

We first performed clonal overlap analysis between the bulk tumor TCR sequences that we obtain from the tumors that also had corresponding peripheral TCR data from our peripheral multi-omics analysis (n=6). We identified TCR sequencies for 43,881 cells of the 77,555 cells in the single cell multi-omics dataset, and the tumor TCR sequence read depth and clonality by patient is enumerated in Table 7. Our aim was to identify if any of the Louvain clusters that we identified in the multi-omics data or the four NPC-associated cluster (8,12,13, and 14) had enrichment of TCRs found in the tumor. We first mapped the peripheral TCR data with tumor TCRs that were exact amino acid sequence matches found in both datasets (Figure 10a). We then overlaid these TCRs that are shared between the blood and the tumor onto the wnnUMAP to assess the cluster distribution of these shared TCR sequences (Figure 10a). We then quantified the number of shared TCRs by cluster both as the total number of shared TCRs in each cluster as well as the percent of TCRs in each cluster that were also mapped to the tumor (Figure 10b, c). Intriguingly, we found that cluster 11 had the highest number of shared tumor TCRs, however when looking at the percent of shared TCRs by clusters, which adjusts for cluster cell number differences, we found that cluster 12 ($CD103^+$, $CD39^+$) had the highest percent of shared TCRs with the tumor. Intriguingly, the other NPC-associated clusters had much lower frequencies of shared tumor TCRs, with cluster 13 having low levels of clonal overlap and clusters 8 and 14 having moderate levels of clonal overlap with the tumor (Figure 10c).

Next, we assessed the tumor microenvironment of these NPC tumors using bulk RNA sequencing (n=13), and mIHC staining (n=15). We first evaluated the cell type composition of

these tumors using the bulk RNA sequencing data and the MCP-Counter package. This method allowed us to approximate and create a score for the abundance of various cell types present in the tumors (T cells, B cells, cancer-associated fibroblasts etc.) (Figure 11a). We quantified the cell types score for all 13 tumors to evaluate the difference in levels and diversity of immune infiltration between these tumors (Figure 11a). In addition to this, we used a 6-color mIHC panel consisting of EBER (tumor marker for NPC, and marker of EBV infection), CD3, CD8, FOXP3, PD-1, CD39(Figure 11b). This panel would allow us to not only evaluate the total T cell infiltrate within these tumors but also assess levels of immune regulation (Figure 11b). Having scored these tumors using these datasets, we performed network correlation analysis on all the tumor scores as well as peripheral data including peripheral CD8 marker expression frequencies, bulk CD8 cluster frequencies, and plasma serology, as well as clinical parameters for each patient (Figure 11c). Using a false discovery rate of <0.1, we identified correlations between all modalities of data except for the multi-omics analysis collected as a part of this work (Figure 11c).

Within this network correlation analysis, we found confirmation of many of the correlations that we had identified in previous chapters and found several noteworthy correlations between the data from the tumor and blood, and these will be highlighted as follows (Figure 11c). First, we found a high degree of correlation between the mIHC T cell score (CD3+ score) and mIHC CD8 T cell score (CD3+, CD8+ score) and the T cell score created using the gene expression data. It was assuring to see high degrees of concordance between the results for these two different modalities. In addition to this, we found that there was a strong positive correlation between CD71 expression on peripheral CD8 T cells and the mIHC PD-1 score. In addition to these positive correlations, we also identified negative correlations between the level

of nuclear antigen plasma antibody titers and the gene expression CD8 T cell score and the CD39 mIHC score. Lastly, we also found a negative correlation between peripheral bulk CD8 cluster 4 ($CD39^+$, $CD103^+$, $HLA\text{-}DR^+$, $CD71^+$, $CD38^+$, $T_{eff)}$ frequencies and the tumor gene expression neutrophil score. (Figure 11c).

5.1.4 *Summary*

In summation, in this chapter we used data obtained from the tumor to identify connections between the peripheral blood and the tumor. First, we found that the single cell NPC associated cluster 12 ($CD39^+$, $CD103^+$) has the highest degree of TCR clonal overlap with tumor TCRs. Next, we used mIHC and tumor gene expression data to quantify the cell types and immune infiltrate of the NPC tumors and performed network correlation analysis to identify that peripheral expression of CD71 on CD8 T cells is positively correlated with PD-1 levels in the tumor, and that frequencies of total cluster 4 ($CD39^+$, $CD103^+$, $HLA\text{-}DR^+$, $CD71^+$, $CD38^+$, T_{eff}) were negatively correlated with intra-tumoral neutrophils levels. These analyses allowed us to better understand the tumor immune landscape and identify peripheral T cell populations that are connected to the tumor.

Table 7:

Sample ID	Total TCR Reads	Productive TCR clones (%)	Diversity of productive CDR3 region	Productive d50 index	Productive Clonality (1-normalized Shannon Wiener Index)
CIE003	42945	85.8	752	0.926	0.164
CIE005	16491	86.8	375	0.917	0.170
CIE010	29657	89.8	732	0.986	0.354
CIE018	369180	82.1	5861	0.936	0.155
CIE023	160	86.9	48	0.812	0.100
CIE028	258075	82.9	1571	0.902	0.132
CIE030	135674	82.8	2548	0.944	0.160
CIE043	479150	80.8	2874	0.870	0.097
CIE051	184047	78.6	1862	0.921	0.144
CIE059	89062	80.7	261	0.889	0.160

Table 7. Tumor TCR Sequence Data Summary.

Figure 10

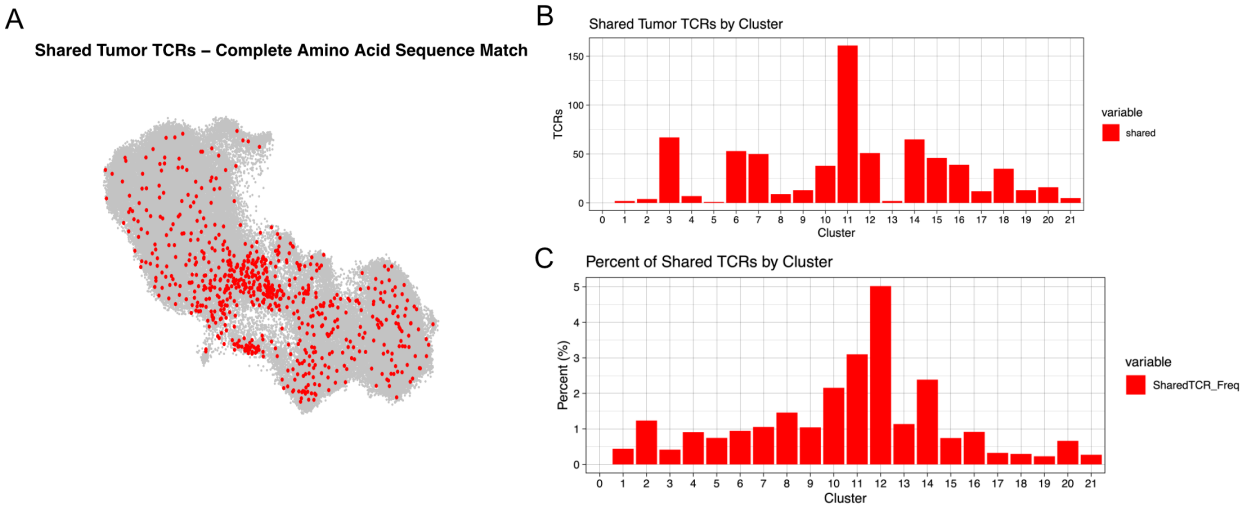

Figure 10. Cluster Enrichment of Clonal Sharing between NPC tumor and peripheralT cells.
a. wnnUMAP plot showing T cells (red) that are were also found in the NPC tumors using bulk TCR sequencing based on exact TCR amino acid sequence matching. **b.** Number of TCRs that are shared with the tumor by clusters. **c.** Percent of TCRs within each cluster that were also mapped to the tumor.

Figure 11:

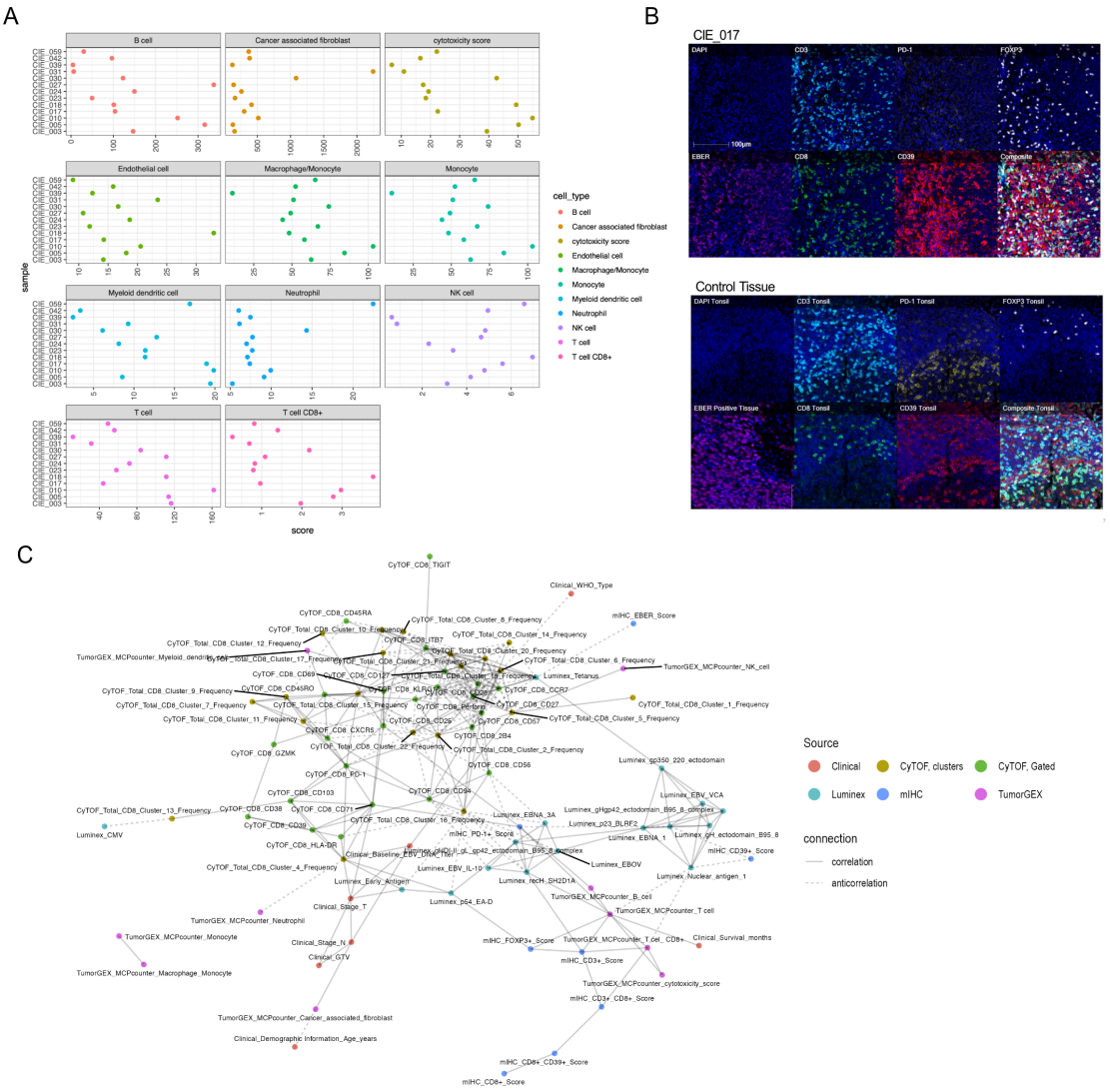

Figure 11. Multiplexed Immunohistochemistry and Tumor Gene expression reveal possible connections between the NPC tumor microenvironment and immune periphery.

a. Dot plots showing cell type scores of bulk tumor gene expression data using MCP-counter from NPC FFPE biopsies (n=13). **b.** Example of mIHC staining of NPC tumor FFPE biopsies (patient CIE_017) and control tissue (tonsil) showing individual and composite staining of EBER (tumor marker), CD3, CD8, PD-1, CD39, FOXP3. **c.** Network correlation analysis showing correlations between all modalities of data. (Spearman correlation, FDR<0.1).

Chapter 6. DISCUSSION AND FUTURE DIRECTIONS

6.1 DISCUSSION

In chapter 2, we found two peripheral CD8 complex phenotypes that were enriched on EBV-specific T cells in NPC patients. One of these complex phenotypes that was identified was cluster 7, which was a T_{eff} cluster defined by high expression of PD-1 and TIGIT. PD-1 and TIGIT are both well-known immune checkpoint (ICB) molecules[81]. Previous research in the context of advanced melanoma has shown that compared to peripheral CMV or Flu-specific T cells, circulating tumor-antigen (NY-ESO-1) specific T cells had higher levels of PD-1 and TIGIT co-expression, and that these cells had improved functionality post-ICB *in vitro*[82]. Since we see similar enrichment of PD-1, TIGIT co-expression exclusively on EBV-specific T cells in NPC patients, this adds credence to the inference that perhaps peripheral PD-1^+, $TIGIT^+$ may also be tumor-specific and furthermore may have reversible levels of exhaustion that may be responsive to dual anti-TIGIT, anti-PD-1 ICB therapy.

The second complex phenotype that we found to be enriched on NPC EBV-specific T cells was the $CD103^+$, $CD39^+$, $CD38^+$, $CD71^+$, and HLA-DR^+ T_{eff} phenotype that defined cluster 4. This phenotype was comprised of markers of T cell activation (i.e., HLA-DR, CD38, CD71), as well as exhaustion (CD39) and trafficking (CD103). In terms of EBV-specific T cells that have high expression of activation markers, we hypothesize that it is possible that these cells are activated in response to recent priming post- antigen presentation in secondary lymphoid organs. It was also notable that these EBV-specific T cells also had elevated expression of CD103, a well-studied marker of tissue residency, cell adhesion, and trafficking, and CD39, an ectonucleotidase that is an emerging marker of T cell exhaustion. In terms of CD39 expression, previous work from our group in the context of non-virally driven lung and colon cancer has

identified that bystander T cells (including virus-specific T cells such as EBV) in the tumor microenvironment lack CD39 expression[48]. Other work in the context of lymphoepithelioma-like carcinoma (LELC), an EBV-driven lung cancer considered to be very similar to NPC, showed that EBV-specific TILs within this tumor do express CD39[83]. Taken together, this suggests that in the context of EBV-driven cancers, EBV-specific T cells do not have a bystander-like role in the tumor microenvironment (TME). In terms of expression of CD39 on peripheral tumor-specific T cells, other groups have also identified that in the context of urothelial carcinoma that peripheral tumor-specific, neo-antigen reactive CD8 T cells had higher expression of CD39 in comparison to virus antigen-reactive T cells[84]. When considering this along with our finding that peripheral EBV-specific T cells in NPC patients express CD39, this suggests that these cells may be more phenotypically similar to circulating tumor-specific T cells. When considering the co-expression of CD39 and CD103 on EBV-specific T cells, previous groups have identified that within the TME, this phenotype is highly enriched for tumor reactive TILs[85]. Recently published work from our group also found increased co-expression of CD103 and CD39 on peripheral Merkel-cell polyomavirus-specific T cells in the context of virally-driven Merkel cell carcinoma, which sets a precedent regarding the presence of this peripheral phenotype on virus-specific T cells in a virally driven cancer. [86] CD103, or Integrin αE, is a well-studied marker of tissue resident memory (Trm) T cells, which are known to express high levels of both CD103 and CD69[87]. CD103 mediates T cell trafficking and homing (particularly to epithelial tissues) through the binding of E-cadherin, which is highly expressed on epithelial cells[87]. In terms of expression of CD103 on peripheral T cells, recent work by Klicznik et al. has shown that in a human skin xenograft mouse model, skin Trm CD4 T cells that express CD69 and CD103 can exit the skin via downregulation of CD69 and these peripheral $CD103^+$ T cells are able migrate

to other tissue sites[88]. Considering that the EBV-specific $CD39^+$, $CD103^+$ cells we identified in NPC patients do not express CD69, we hypothesize that these exhausted peripheral $CD103^+$ EBV-specific T cells may have exited the tumor post-chronic antigen stimulation through downregulation of CD69[88]. Overall, these results suggest that the unique NPC-associated EBV-specific peripheral phenotypic profiles identified in this chapter may be linked to the tumor as a part of the cancer immunity cycle[43].

In the third chapter we found that mass cytometry cluster 4 ($CD103^+$, $CD39^+$, $CD38^+$, $CD71^+$, and HLA-DR^+ $T_{eff)}$ phenotypes on EBV-specific T cells and on bulk CD8 T cells were positively correlated with EBV-DNA titer and tumor burden (stage and GTV) and therefore represented an NPC-associate phenotype. In the fourth chapter, we further parse this NPC-associated T cell population into four subpopulations based on their surface marker and transcriptional profiles, and we used the defining markers of these subsets to see which of these NPC associated subclusters were associated with clinical outcomes. We found that the frequency of $CD103^+$, $CD39^+$ EBV-specific T cells is more strongly correlated to EBV-DNA plasma titers, GTV, and stage (N) as compared to total $CD103^+$, $CD39^+$ total CD8 T cell frequencies (agnostic of antigen specificity). This suggests that the $CD103^+$, $CD39^+$ phenotype is closely linked to disease burden only when expressed on EBV-specific T cells in the context of NPC. In terms of $CD103^+$, $CD39^+$ expression in Merkel cell carcinoma, recent work from our group found that $CD103^+$, $CD39^+$ CD8 T cells was related to increased tumor burden and poor response to ICB therapy[86]. In addition, many studies has shown that $CD103^+$ on TILs suggests favorable prognosis. [89] This may suggest that retention of $CD103^+$ cells in the tumor is favorable, however circulating $CD103^+$ cells connote poor patient outcomes.

In addition to clinical outcomes, in chapter 4, we showed that $CD103^+$, $CD39^+$ CD8 T cells in NPC patients express additional markers of T cell migration such as CCR6 and CXCR3, and we also saw that these cells expressed tissue residency genes such as *CD69* and *ZNF683,* suggesting migration from a tissue origin. We also saw that these cells were less clonally expanded, have less clonal sharing with the $CD71^+$ and $CD38^+$, HLAR-DR^+ cells, and have the highest number of EBV-specific T cells suggesting that these cells are not as clonally related to the activated/proliferative cells. Finally in chapter 5, we found that these cells also had the highest degree of clonal overlap with the tumor. Overall, these findings suggest that $CD103^+$, $CD39^+$ CD8 T cells in the NPC patient periphery have many connections with the tumor and EBV antigen-specificity and may possibly originate from the tumor.

On the other hand, we see that $CD71^+$, and $CD38^+$, HLAR-DR^+ cells are correlated with GTV and/or stage both when expressed on EBV-specific T cells and when expressed by total CD8 T cells. This suggests that while these phenotypes are also enriched among EBV-specific T cells in NPC patients, that correlations with tumor burden and this phenotype are not completely driven by EBV antigen-specificity. This suggests that perhaps cells within this phenotype that are not EBV-specific may also have correlations to the tumor or may represent T cells with other tumor antigen specificities. In chapter 5, we found that these two cell populations have a high degree of clonal expansion and clonal overlap suggesting that they originate from similar priming events.

We believe that the findings enumerated here have improved our understanding of the cancer-immunity cycle in the context of nasopharyngeal carcinoma (Figure 12). In the last section, I will discuss how our results might fit into this paradigm and share our hypotheses regarding the roles of the immune subsets that we profiled. We hypothesize that EBV-specific T

cells in the tumor become exhausted due to chronic EBV antigen stimulation and the immunosuppressive tumor microenvironment. We believe that this leads to increased expression of CD39, and since we see a positive correlation between the size and volume of the tumor and the frequency of $CD39^+$, $CD103^+$ EBV-specific T cell in the periphery, we conjecture that these cells egress from the tumor. Concurrently we believe that as tumor burden increases, we see higher levels of cancer antigens and EBV antigens slough off from the tumor, since EBV-DNA titer is positively correlated with stage (T). NPC cancer antigens are released by increasing tumor size and are presented on APCs in the periphery, where CD8 T cells are primed and activated and therefore express HLA-DR, CD38 and CD71. This is supported by the fact that $CD71^+$ and HLA-DR^+, $CD38^+$ bulk CD8 T cell frequencies are positively correlated with tumor size agnostic of antigen specificity. This is additionally corroborated by the fact that CD71 expression on peripheral bulk CD8 T cells is also positively correlated with PD-1 expression in the tumor, suggesting that tumor T cell exhaustion is related to the presence of proliferative peripheral CD8 T cells. Finally, we suggest the possibility that these activated/proliferative T cells then traffic back to the tumor where they either mediate tumor cell death or experience exhaustion due to external immune regulatory stimuli in the tumor microenvironment. Based on our current knowledge it is unclear what the tumor reactivity of these cells are. Finally, it's possible that the cycle begins again when more cancer antigens are released due to activated/proliferative CD8 T cell mediated tumor cell death. While our findings fit this proposed framework, much work is needed to validate this putative paradigm.

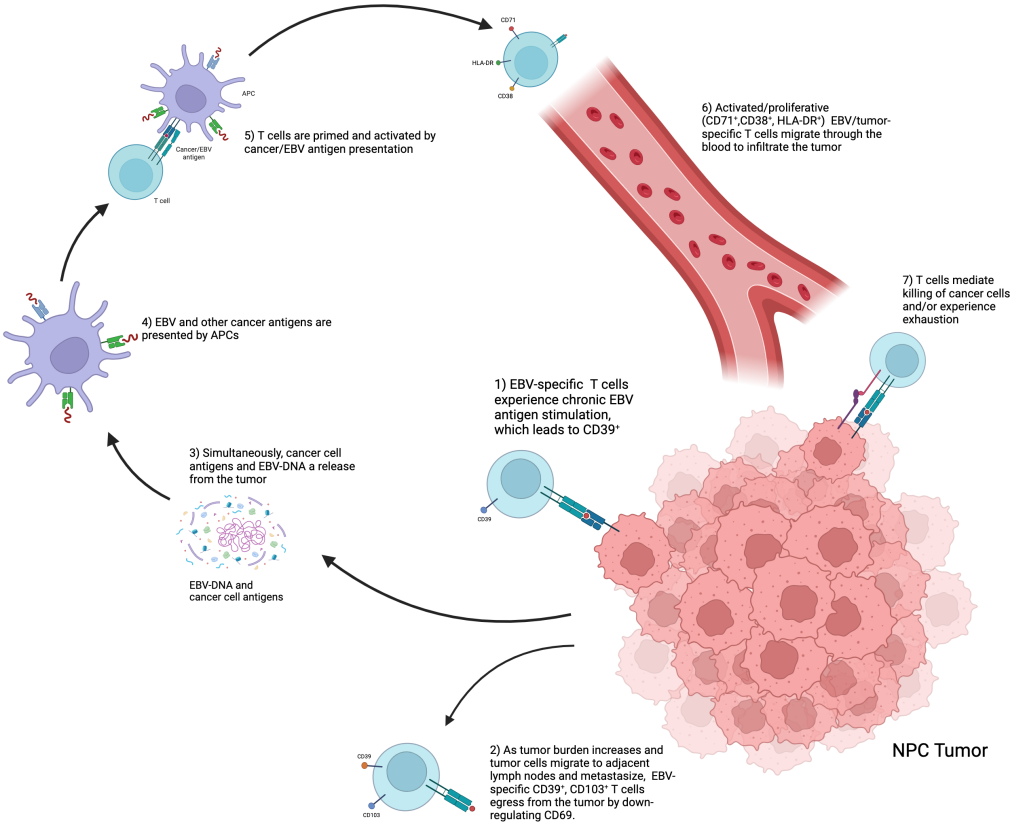

Figure 12: NPC Cancer immunity paradigm based on synthesized results.

6.2 Future Directions

One important aspect of this research that require follow-up is to assess the functionality of cells with these EBV-specific T cell phenotypes. Currently, it is unknown which of these cells have the functional capacity to mediate cytotoxicity against tumor cells. Based on our hypothesis that peripheral $CD103^+$, $CD39^+$ cells have extravasated from the tumor post-chronic antigen stimulation and that CD71 and $HLA\text{-}DR^+$, $CD38^+$ represent peripheral T cells that have undergone recent priming, we could conjecture that the latter population would have greater tumor reactivity than the former. However further studies are needed to validate this hypothesis. *In vitro* functionality assays with NPC cell lines and intracellular cytokine staining (ICS) might be useful in delineating which of these cell types have the greatest capacity for tumor reactivity.

The second direction to take this work would be to assess the role of these NPC-associated phenotypes in the context of immunotherapy and particularly immune checkpoint blockade therapy. In fact, we have currently initiated a study to profile peripheral CD8 T cells using mass cytometry and single-cell multi-omics in collaboration with Dr. Darren Lim (National Cancer Centre, Singapore), who has been conducting a trial to assess the efficacy of anti-PD-1, anti-CTLA4 immune checkpoint blockade (ICB) therapy in the context of nasopharyngeal carcinoma[42]. So far, the results of his work show that this treatment is safe, has better efficacy than single-arm anti-PD-1 therapy, and that patients with low EBV-DNA titers respond better than those with high EBV-DNA titers. This might suggest that the presence of the NPC-associated T cell phenotypic populations we identified in this work might connote poorer outcomes since there is a strong positive correlation between the frequency of these phenotypic populations and EBV-DNA titer. However, in this trial, EBV-DNA titers were measured prior to the study initiation but after the patients had received treatments, which is not directly comparable to our work where EBV-DNA titers were measures prior to any treatment and at diagnosis[42]. Therefore, further investigation of these phenotypes is needed to understand their role in ICB therapy.

6.3 IMPACTS

Characterization of peripheral EBV-specific T cell phenotypes in NPC patients using multiple approaches not only improved our understanding of the NPC immune landscape, but also identified potentially clinically relevant EBV-specific T cell phenotypes that could be used as correlate of disease burden for NPC. If further studies corroborate the finding here that suggest that bulk CD8 frequencies of the phenotypes that can be easily gated using clinical flow

cytometry, there could be translational value in using these markers for clinical immune monitoring. In addition to NPC, these findings could have application for other EBV-driven cancers, and even disease beyond cancers. For example, there has even been interest in characterizing T_{rm} T cells in the context of multiple sclerosis in the brain [90]. Overall, we hope that this work has contributed to the field of EBV virology and cancer immunology, and that continued exploration of these findings will have a positive impact on the lives of patients with cancer and EBV-related diseases.

BIBLIOGRAPHY

1. Hislop, A.D., Taylor, G.S., Sauce, D. & Rickinson, A.B. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. *Annu Rev Immunol* **25**, 587-617 (2007).
2. Young, L.S., Yap, L.F. & Murray, P.G. Epstein-Barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer* **16**, 789-802 (2016).
3. Damania, B., Kenney, S.C. & Raab-Traub, N. Epstein-Barr virus: Biology and clinical disease. *Cell* **185**, 3652-3670 (2022).
4. Jangra, S., Yuen, K.S., Botelho, M.G. & Jin, D.Y. Epstein-Barr Virus and Innate Immunity: Friends or Foes? *Microorganisms* **7**(2019).
5. Hislop, A.D. & Taylor, G.S. T-Cell Responses to EBV. *Curr Top Microbiol Immunol* **391**, 325-353 (2015).
6. Newell, E.W.*, et al.* Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat Biotechnol* **31**, 623-629 (2013).
7. Chen, Y.P.*, et al.* Nasopharyngeal carcinoma. *Lancet* **394**, 64-80 (2019).
8. Young, L.S. & Dawson, C.W. Epstein-Barr virus and nasopharyngeal carcinoma. *Chin J Cancer* **33**, 581-590 (2014).
9. Bjornevik, K.*, et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **375**, 296-301 (2022).
10. Bray, F.*, et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018).
11. Patel, V.J., Chen, N.-W. & Resto, V.A. Racial and Ethnic Disparities in Nasopharyngeal Cancer Survival in the United States: A SEER Study. *Otolaryngology–Head and Neck Surgery* **156**, 122-131 (2017).
12. Mahdavifar, N., Ghoncheh, M., Mohammadian-Hafshejani, A., Khosravi, B. & Salehiniya, H. Epidemiology and Inequality in the Incidence and Mortality of Nasopharynx Cancer in Asia. *Osong Public Health Res Perspect* **7**, 360-372 (2016).
13. Xu, M.*, et al.* Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nat Genet* **51**, 1131-1136 (2019).
14. Xu, M.*, et al.* Genome sequencing analysis identifies Epstein–Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nature Genetics* **51**, 1131-1136 (2019).
15. Li, X., Fasano, R., Wang, E., Yao, K.T. & Marincola, F.M. HLA associations with nasopharyngeal carcinoma. *Curr Mol Med* **9**, 751-765 (2009).
16. Wu, C.C.*, et al.* Perspective: Contribution of Epstein-Barr virus (EBV) Reactivation to the Carcinogenicity of Nasopharyngeal Cancer Cells. *Cancers (Basel)* **10**(2018).
17. Thompson, M.P. & Kurzrock, R. Epstein-Barr virus and cancer. *Clin Cancer Res* **10**, 803-821 (2004).
18. Song, H.*, et al.* Interpretation of EBV infection in pan-cancer genome considering viral life cycle: LiEB (Life cycle of Epstein-Barr virus). *Sci Rep* **9**, 3465 (2019).
19. Li, H.*, et al.* Therapies based on targeting Epstein-Barr virus lytic replication for EBV-associated malignancies. *Cancer Sci* **109**, 2101-2108 (2018).
20. Hu, L.*, et al.* Comprehensive profiling of EBV gene expression in nasopharyngeal carcinoma through paired-end transcriptome sequencing. *Front Med* **10**, 61-75 (2016).

21. Fang, C.Y.*, et al.* The synergistic effect of chemical carcinogens enhances Epstein-Barr virus reactivation and tumor progression of nasopharyngeal carcinoma cells. *PLoS One* **7**, e44810 (2012).
22. Chiu, S.H.*, et al.* Epstein-Barr virus BALF3 mediates genomic instability and progressive malignancy in nasopharyngeal carcinoma. *Oncotarget* **5**, 8583-8601 (2014).
23. Chan, K.C.A.*, et al.* Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer. *N Engl J Med* **377**, 513-522 (2017).
24. Chan, K.C.*, et al.* Molecular characterization of circulating EBV DNA in the plasma of nasopharyngeal carcinoma and lymphoma patients. *Cancer Res* **63**, 2028-2032 (2003).
25. Chan, K.C.A.*, et al.* Analysis of Plasma Epstein–Barr Virus DNA to Screen for Nasopharyngeal Cancer. *New England Journal of Medicine* **377**, 513-522 (2017).
26. Tan, L.P.*, et al.* Systematic comparison of plasma EBV DNA, anti-EBV antibodies and miRNA levels for early detection and prognosis of nasopharyngeal carcinoma. *Int J Cancer* **146**, 2336-2347 (2020).
27. Simon, J.*, et al.* Validation of an Epstein-Barr Virus Antibody Risk Stratification Signature for Nasopharyngeal Carcinoma by Use of Multiplex Serology. *J Clin Microbiol* **58**(2020).
28. Twu, C.W.*, et al.* Comparison of the prognostic impact of serum anti-EBV antibody and plasma EBV DNA assays in nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys* **67**, 130-137 (2007).
29. Fan, H.*, et al.* Laboratory markers of tumor burden in nasopharyngeal carcinoma: A comparison of viral load and serologic tests for Epstein-Barr virus. *International Journal of Cancer* **112**, 1036-1041 (2004).
30. Song, L.*, et al.* Identification of anti-Epstein-Barr virus (EBV) antibody signature in EBV-associated gastric carcinoma. *Gastric Cancer* **24**, 858-867 (2021).
31. Liu, Z.*, et al.* Characterization of the humoral immune response to the EBV proteome in extranodal NK/T-cell lymphoma. *Sci Rep* **11**, 23664 (2021).
32. Hsu, C.*, et al.* Safety and Antitumor Activity of Pembrolizumab in Patients With Programmed Death-Ligand 1-Positive Nasopharyngeal Carcinoma: Results of the KEYNOTE-028 Study. *J Clin Oncol* **35**, 4050-4056 (2017).
33. Ma, B.B.Y.*, et al.* Antitumor Activity of Nivolumab in Recurrent and Metastatic Nasopharyngeal Carcinoma: An International, Multicenter Study of the Mayo Clinic Phase 2 Consortium (NCI-9742). *J Clin Oncol* **36**, 1412-1418 (2018).
34. Fang, W.*, et al.* Camrelizumab (SHR-1210) alone or in combination with gemcitabine plus cisplatin for nasopharyngeal carcinoma: results from two single-arm, phase 1 trials. *Lancet Oncol* **19**, 1338-1350 (2018).
35. Huang, J.*, et al.* Epstein-Barr virus-specific adoptive immunotherapy for recurrent, metastatic nasopharyngeal carcinoma. *Cancer* **123**, 2642-2650 (2017).
36. Lutzky, V.P.*, et al.* Optimization of LMP-specific CTL expansion for potential adoptive immunotherapy in NPC patients. *Immunol Cell Biol* **87**, 481-488 (2009).
37. Comoli, P.*, et al.* Cell therapy of stage IV nasopharyngeal carcinoma with autologous Epstein-Barr virus-targeted cytotoxic T lymphocytes. *J Clin Oncol* **23**, 8942-8949 (2005).
38. Chia, W.K.*, et al.* Adoptive T-cell transfer and chemotherapy in the first-line treatment of metastatic and/or locally recurrent nasopharyngeal carcinoma. *Mol Ther* **22**, 132-139 (2014).

39. Cho, H.I., *et al.* A novel Epstein-Barr virus-latent membrane protein-1-specific T-cell receptor for TCR gene therapy. *Br J Cancer* **118**, 534-545 (2018).
40. Mai, H.Q., *et al.* Toripalimab or placebo plus chemotherapy as first-line treatment in advanced nasopharyngeal carcinoma: a multicenter randomized phase 3 trial. *Nat Med* **27**, 1536-1543 (2021).
41. Wang, F.H., *et al.* Efficacy, Safety, and Correlative Biomarkers of Toripalimab in Previously Treated Recurrent or Metastatic Nasopharyngeal Carcinoma: A Phase II Clinical Trial (POLARIS-02). *J Clin Oncol* **39**, 704-712 (2021).
42. Lim, D.W., *et al.* Clinical efficacy and biomarker analysis of dual PD-1/CTLA-4 blockade in recurrent/metastatic EBV-associated nasopharyngeal carcinoma. *Nat Commun* **14**, 2781 (2023).
43. Chen, D.S. & Mellman, I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* **39**, 1-10 (2013).
44. Newell, E.W. & Becht, E. High-Dimensional Profiling of Tumor-Specific Immune Responses: Asking T Cells about What They "See" in Cancer. *Cancer Immunol Res* **6**, 2-9 (2018).
45. Schreiber, R.D., Old, L.J. & Smyth, M.J. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565-1570 (2011).
46. Gros, A., *et al.* Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nature medicine* **22**, 433-438 (2016).
47. Rizvi, N.A., *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-128 (2015).
48. Simoni, Y., *et al.* Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575-579 (2018).
49. Rosato, P.C., *et al.* Virus-specific memory T cells populate tumors and can be repurposed for tumor immunotherapy. *Nat Commun* **10**, 567 (2019).
50. Thomas, R., *et al.* NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives. *Front Immunol* **9**, 947 (2018).
51. Simoni, Y., Fehlings, M. & Newell, E.W. Multiplex MHC Class I Tetramer Combined with Intranuclear Staining by Mass Cytometry. *Methods Mol Biol* **1989**, 147-158 (2019).
52. Becher, B., *et al.* High-dimensional analysis of the murine myeloid cell system. *Nat Immunol* **15**, 1181-1189 (2014).
53. Finck, R., *et al.* Normalization of mass cytometry data with bead standards. *Cytometry A* **83**, 483-494 (2013).
54. Chevrier, S., *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst* **6**, 612-620 e615 (2018).
55. Nowicka, M., *et al.* CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Res* **6**, 748 (2017).
56. Zheng, Y., Jun, S.-H., Tian, Y., Florian, M. & Gottardo, R. Robust Normalization and Integration of Single-cell Protein Expression across CITE-seq Datasets. *bioRxiv*, 2022.2004.2029.489989 (2022).
57. Levine, J.H., *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197 (2015).
58. Becht, E., *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* (2018).

59. Leland McInnes, J.H., James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *Arxiv* **arXiv:1802.03426**(2018).
60. Goswami, S.*, et al.* Immune profiling of human tumors identifies CD73 as a combinatorial target in glioblastoma. *Nat Med* **26**, 39-46 (2020).
61. Weber, L.M., Nowicka, M., Soneson, C. & Robinson, M.D. diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. *Commun Biol* **2**, 183 (2019).
62. Hao, Y.*, et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304 (2024).
63. McGinnis, C.S., Murrow, L.M. & Gartner, Z.J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e324 (2019).
64. Korsunsky, I.*, et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).
65. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
66. Kumar, B.V.*, et al.* Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep* **20**, 2921-2934 (2017).
67. Borcherding, N., Bormann, N.L. & Kraus, G. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Res* **9**, 47 (2020).
68. Samuel Marsh, M.S., & Paul Hoffman. samuel-marsh/scCustomize: Version 2.1.2 (v2.1.2). *Zenodo* (2024).
69. Becht, E.*, et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology* **17**, 218 (2016).
70. Yeong, J.*, et al.* Higher densities of Foxp3(+) regulatory T cells are associated with better prognosis in triple-negative breast cancer. *Breast Cancer Res. Treat.* **163**, 21-35 (2017).
71. Lim, J.C.T.*, et al.* An automated staining protocol for seven-colour immunofluorescence of human tissue sections for diagnostic and prognostic use. *Pathology* **50**, 333-341 (2018).
72. Yeong, J.*, et al.* Prognostic value of CD8 + PD-1+ immune infiltrates and PDCD1 gene expression in triple negative breast cancer. *J Immunother Cancer* **7**, 34 (2019).
73. Lam, J.H.*, et al.* Expression of CD38 on Macrophages Predicts Improved Prognosis in Hepatocellular Carcinoma. *Frontiers in immunology* **10**, 2093-2093 (2019).
74. Fiore, C.*, et al.* Utility of multispectral imaging in automated quantitative scoring of immunohistochemistry. *Journal of Clinical Pathology* **65**, 496-502 (2012).
75. Abel, E.J.*, et al.* Analysis and validation of tissue biomarkers for renal cell carcinoma using automated high-throughput evaluation of protein expression. *Hum. Pathol.* **45**, 1092-1099 (2014).
76. Feng, Z.*, et al.* Multiparametric immune profiling in HPV– oral squamous cell cancer. *JCI Insight* **2**(2017).
77. Giraldo, N.A.*, et al.* Multidimensional, quantitative assessment of PD-1/PD-L1 expression in patients with Merkel cell carcinoma and association with response to pembrolizumab. *J Immunother Cancer* **6**, 99 (2018).
78. Mascaux, C.*, et al.* Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature* (2019).

79. Tan, W.C.C.*, et al.* Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Commun (Lond)* **40**, 135-153 (2020).
80. Liu, W.*, et al.* The diagnostic value of EBV-DNA and EBV-related antibodies detection for nasopharyngeal carcinoma: a meta-analysis. *Cancer Cell International* **21**, 164 (2021).
81. Morad, G., Helmink, B.A., Sharma, P. & Wargo, J.A. Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* **184**, 5309-5337 (2021).
82. Chauvin, J.M.*, et al.* TIGIT and PD-1 impair tumor antigen-specific $CD8^+$ T cells in melanoma patients. *J Clin Invest* **125**, 2046-2058 (2015).
83. Simoni, Y.*, et al.* Partial absence of PD-1 expression by tumor-infiltrating EBV-specific CD8(+) T cells in EBV-driven lymphoepithelioma-like carcinoma. *Clin Transl Immunology* **9**, e1175 (2020).
84. Holm, J.S.*, et al.* Neoantigen-specific CD8 T cell responses in the peripheral blood following PD-L1 blockade might predict therapy outcome in metastatic urothelial carcinoma. *Nature Communications* **13**, 1935 (2022).
85. Duhen, T.*, et al.* Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat Commun* **9**, 2724 (2018).
86. Ryu, H.*, et al.* Merkel cell polyomavirus-specific and CD39(+)CLA(+) CD8 T cells as blood-based predictive biomarkers for PD-1 blockade in Merkel cell carcinoma. *Cell Rep Med*, 101390 (2024).
87. Xu, W., Bergsbaken, T. & Edelblum, K.L. The multifunctional nature of CD103 (αEβ7 integrin) signaling in tissue-resident lymphocytes. *Am J Physiol Cell Physiol* **323**, C1161-c1167 (2022).
88. Klicznik, M.M.*, et al.* Human CD4(+)CD103(+) cutaneous resident memory T cells are found in the circulation of healthy individuals. *Sci Immunol* **4**(2019).
89. Kim, Y., Shin, Y. & Kang, G.H. Prognostic significance of CD103+ immune cells in solid tumor: a systemic review and meta-analysis. *Scientific Reports* **9**, 3808 (2019).
90. Serafini, B., Rosicarelli, B., Veroni, C. & Aloisi, F. Tissue-resident memory T cells in the multiple sclerosis brain and their relationship to Epstein-Barr virus infected B cells. *Journal of Neuroimmunology* **376**, 578036 (2023).

VITA

Nandita Kumar is PhD candidate at the University of Washington. She conducted her thesis dissertation work in Dr. Evan Newell's lab at the Fred Hutchinson Cancer Center in Seattle, WA. She completed her undergraduate degree in Microbiology at Oregon State University in 2016.